TRAFFIC AND TRANSPORTATION STUDIES

PROCEEDINGS OF THE SEVENTH INTERNATIONAL CONFERENCE ON TRAFFIC AND TRANSPORTATION STUDIES

August 3–5, 2010
Kunming, China

SPONSORED BY
Systems Engineering Society of China (SESC), China
American Society of Civil Engineers (ASCE), USA
Beijing Jiaotong University, Beijing (BJU), China
Institute of Transportation Engineers (ITE), USA
State Key Lab of Rail Traffic Control and Safety (SKLRTCS), China
Beijing Transportation Research Center (BTRC), China
Japan Society of Civil Engineers (JSCE), Japan
Hong Kong Society for Transportation Studies (HKSTS), China
Institution of Highways & Transportation (IHT), UK

SUPPORTED BY
National Natural Science Foundation of China (NSFC)

EDITED BY
Baohua Mao
Zongzhong Tian
Haijun Huang
Ziyou Gao





Published by the American Society of Civil Engineers

TRAFFIC AND TRANSPORTATION STUDIES

PROCEEDINGS OF THE SEVENTH INTERNATIONAL CONFERENCE ON TRAFFIC AND TRANSPORTATION STUDIES

August 3–5, 2010
Kunming, China

SPONSORED BY
Systems Engineering Society of China (SESC), China
American Society of Civil Engineers (ASCE), USA
Beijing Jiaotong University, Beijing (BJU), China
Institute of Transportation Engineers (ITE), USA
State Key Lab of Rail Traffic Control and Safety (SKLRTCS), China
Beijing Transportation Research Center (BTRC), China
Japan Society of Civil Engineers (JSCE), Japan
Hong Kong Society for Transportation Studies (HKSTS), China
Institution of Highways & Transportation (IHT), UK

SUPPORTED BY
National Natural Science Foundation of China (NSFC)

EDITED BY
Baohua Mao
Zongzhong Tian
Haijun Huang
Ziyou Gao





Published by the American Society of Civil Engineers

Copyright and Disclaimer

ISBN: 978-0-7844-1123-0

Any statements expressed in these materials are those of the individual authors and do not necessarily represent the views of ASCE, which takes no responsibility for any statement made herein. No reference made in this publication to any specific method, product, process or service constitutes or implies an endorsement, recommendation, or warranty thereof by ASCE. The materials are for general information only and do not represent a standard of ASCE, nor are they intended as a reference in purchase specifications, contracts, regulations, statutes, or any other legal document.

ASCE makes no representation or warranty of any kind, whether express or implied, concerning the accuracy, completeness, suitability, or utility of any information, apparatus, product, or process discussed in this publication, and assumes no liability therefore. This information should not be used without first securing competent advice with respect to its suitability for any general or specific application. Anyone utilizing this information assumes all liability arising from such use, including but not limited to infringement of any patent or patents.

Copyright © 2010 by the American Society of Civil Engineers.
All Rights Reserved.

ASCE and American Society of Civil Engineers—Registered in U.S. Patent and Trademark Office.

Photocopies and reprints.
You can obtain instant permission to photocopy ASCE publications by using ASCE's online permission service (www.pubs.asce.org/authors/RightslinkWelcomePage.html). Requests for 100 copies or more should be submitted to the Reprints Department, Publications Division, ASCE, (address above); email: permissions@asce.org. A reprint order form can be found at www.pubs.asce.org/authors/reprints.html.

American Society of Civil Engineers
ASCE International Headquarters
1801 Alexander Bell Drive
Reston, VA 20191-4400 USA

Call Toll-Free in the U.S.: 1-800-548-2723 (ASCE)
Call from anywhere in the world: 1-703-295-6300
Internet: http://www.pubs.asce.org

Preface

The International Conference on Traffic and Transportation Studies (ICTTS) has entered its seventh meeting since its origination in 1998. ICTTS 2010 was hosted by Beijing Jiaotong University (BJU) and Transportation Systems Engineering Committee of Systems Engineering Society of China (SESC). We are also grateful for the sponsorships by the American Society of Civil Engineers (ASCE), the Institute of Transportation Engineers (ITE), Japan Society of Civil Engineers (JSCE), State Key Lab of Rail Traffic Control and Safety (SKLRTCS), China, Beijing Transportation Research Center (BTRC), China, Hong Kong Society for Transportation Studies (HKSTS), and Institution of Highways & Transportation (IHT), UK. We would also like to express special thanks to the National Natural Science Foundation of China (NSFC) for their financial support.

All the papers in the proceedings were peer reviewed. There were 300 initial manuscript submissions. After careful peer reviews, a total of 141 papers were accepted for presentation and publication in the proceedings, among which 50 papers were by overseas authors representing 14 different countries and regions, and the remaining 91 papers were by authors from Mainland China. As the editors of the proceedings, we would like to thank many individuals who contributed to the paper reviewing process, especially Prof. Ning Wu, Dr. Pengjun Zheng, Dr. Anthony TH Chin, Dr. Haibo Chen, Dr. Tin-kin Ho, Dr. Junyi Zhang, Dr. Huijun Sun, Dr. Meng Xu, Dr. Bin Jia and Dr. Shaokuan Chen.

In addition, we would like to thank Prof. Quanxin Sun, Prof. Jun Liu, Prof. Chunfu Shao and several faculty members of Beijing Jiaotong University who volunteered their time and effort contributing to the overall organization of the conference. Without their help, this conference could not have been as successful as it is today.

Finally, we would like to thank Ms. Donna Dickert and Mr. Jonathan Esslinger of ASCE, who have been extremely helpful during the process of publishing these proceedings.

Editors:
Baohua Mao, Beijing Jiaotong University, China
Zongzhong Tian, University of Nevada - Reno, United States
Haijun Huang, Beihang University, China
Ziyou Gao, Beijing Jiaotong University, China

August 2010

Acknowledgments

HONORARY CHAIRMAN

Prof. NING, Bin, BJU, China
Prof. SINHA, Kumares C., USA
Prof. ZHANG, Guowu, SESC, China

INTERNATIONAL SCIENTIFIC COMMITTEE

Chair:

Prof. WANG, Qingyun, State Bureau of Material Reserve of National Development and Reform Commission of P. R. China

Co-Chairs:

Prof. BELL, M G H, UK
Prof. GAO, Ziyou, China
Prof. HARATA, Noboru, Japan
Prof. LI, Xuewei, BJU, China
Prof. QUAN, Yongsheng, BTRC, Beijing, China

Members:

Prof. BONSALL, Peter, UK
Prof. BOSKE, Leigh B., USA
Dr. BERTINI, Robert L., USA
Dr. CHEN, Haibo, UK
Dr. CHIN, Anthony, Singapore
Prof. De PALMA, A. France
Prof. FERREIRA, Luis, Australia
Prof. FUJIWARA, Akimasa, Japan
Prof. GUO, Jifu, BTRC, China
Prof. HANSEN, Ingo A., Netherlands
Dr. HO, Tin-kin, Hong Kong
Prof. HUANG, Haijun, China
Prof. IIDA, Yasunori, Japan
Prof. KUHNE, Reinhart D., Germany
Prof. LAM, William H.K., Hong Kong
Dr. LI, Guoquan, Japan
Prof. LI, Yinzhen, China

Prof. LUBASHEVSKY, Ihor, Russia
Prof. MARCHAL, Jean, Belgium
Prof. MORIKAWA, Takayuki, Japan
Prof. ORTUZAR, Juan de Dios, Chile
Dr. WONG, S.C., Hong Kong
Dr. YAMAMOTO, Toshiyuki, Japan
Prof. YANG, Hai, Hong Kong
Dr. ZHENG Pengjun, UK

ORGANIZING COMMITTEE

Chair:

Prof. SUN, Quanxin, BJU, China

Executive Chair:

Prof. MAO, Baohua, BJU, China

Co-Chairs:

Dr. TIAN, Zongzhong, ITE, U.S.A.
Dr. ZHANG, Junyi, Japan
Prof. LIU, Jun, BJU, China
Dr. LIU, Jenny, U.S.A.

Members:

Dr. NIE, Lei, BJU, China
Mr. YANG Xiaoye, Yunnan, China
Mr. SU, Xiaojun, CAST, China
Ms. SUN, Huizhang, BJU, China
Ms. XUE, Xinwei, SESC, China
Prof. SHAO, Chunfu, BJU, China
Prof. LIU, Zuoyi, BJU, China
Dr. ZHAO, Shengchuan, DLUT, China
Dr. CHEN, Shaokuan, BJU, China
Ms. GUO Wei, TSEIT, China

Secretary:

Ms. CHENG, Xiaoqing, BJU, China
Ms. YANG, Lingling, TSEIT, China

Table of Contents

Extracting Land Cover/Use from Remotely Sensed Imagery: Potentials for Urban Planning 1

Ming Zhong, Yun Zhang, Ahad Beykaei, Bahram Salehi, Mike Ircha, Yongdae Gweon, and Sheng Gao

Backcasting Assessment of Strategies for Efficiently Sustainable Urban Transport Developments of Developing Cities 13

Xuesong Feng, A. Fujiwara, J. Y. Zhang, X. J. Niu, and Y. Hayashi

Sustainable Mobility and Sustainable Transport for Medium-Size Cities in China 27

Reinhart Kühne, Martin Ruhé, and Xiaoxu Bei

A Sustainable Development Based Land Use Assessment for Freeway Construction 37

Yan Liu, Jianfeng Liu, and Kerong Deng

The Future of Interdisciplinary Policy Research in Africa: A Case Study of the Transport Sector 47

Fred Amonya

Taxi Demand Modeling to Ensure Appropriate Taxi Supply: Do Both Open Access and Model Based Restrictions Fail? 60

James M. Cooper and Wim Faber

Actions Taken on Green Transport: Experiences of Beijing 73

Xiaoming Liu, Jinchuan Chen, Zhili Liu, Haibo Chen, and Yu Ge

Shanghai Poly-Center Urban Planning and the Impact of Transport System 85

Haixiao Pan

Guangzhou Comprehensive Transportation Development Strategy (2010-2020) 97

Xiaoyi Ma

Development in Market Structure and Regulation of Urban Bus Services in Developing Countries and Countries in Transition 110

Brendan Finn and Corinne Mulley

An Available International Freight Transport System with Rail Container and RORO/Ferry Ship 122

Guoquan Li

Ports Spatial Structure Analytical Method and Case Study 132

Shilin Han, Tian Xu, Tao You, and Liping Zhao

A Study on the Development of Urban Environment Measure's Evaluation Model 144

Tetsuo Morita, Hiroshi Sugita, and Shengchuan Zhao

Cost Structure of Urban Bus Operations in Yangon, Myanmar **156**

Hironori Kato, Nozomi Saito, Akihiro Inagi, and Phyo Thet Thet Htun

Transport Infrastructure and Economic Development: An Empirical Study for China **166**

Huang Qian and Harata Noboru

Estimating Greenhouse Gases from Roadway Transportation—Methodology Overview **174**

Delwar Murshed

Adaptation Strategies for Low Carbon Transport by 2050 **186**

Hector G. Lopez-Ruiz

Emission Volume, Air Quality, and Health Effects in Japan: The Cases of Tokyo, Kawasaki, Osaka, Amagasaki, and Kita Kyusyu **201**

Keiko Hirota and Shigeru Kashima

Optimization of Mode Share for Low-Carbon-Oriented Urban Passenger Transportation System **220**

Liang X. Pi, Yamamoto Toshiyuki, and Morikawa Takayuki

A Comparative Study on Chinese and Foreign Road Transport Energy Consumption Statistical Indicators **228**

Qipeng Sun, Baohua Mao, and Shunping Jia

Comparison Analysis of Idle Emissions Measurement Program in Jabodetabek, Indonesia **238**

S. B. Nugroho, A. Fujiwara, and Junyi Zhang

Estimation of CO_2 Mitigation Potential from Freight Transport Based on System Dynamics Model—A Case Study of Beijing, China **256**

Yunjing Wang, Jinling Fei, Xuesong Feng, Hidefumi Imura, and Yoshitsugu Hayashi

An Analytical Method for Intersection Entering Traffic Flow Estimation **269**

Rui Wang, Ryota Horiguchi, Shinji Tanaka, and Masao Kuwahara

Cycle Length Optimization at Isolated Signalized Intersections from the Viewpoint of Emission **275**

Danpeng Ma and Hideki Nakamura

Influence of the Number of Through Lane Groups on Saturation Flow Rates at Signalized Intersections **285**

Wanchao Zhang, Yang Wang, and Xiaokuan Yang

The Impact Analysis of Intersection Sight Distance on Vehicle Speed **295**

Miaomiao Liu, Guangquan Lu, and Yang Li

Intersection Safety Evaluation Model **305**

Shengxue Zhu, Jian Lu, and Gu Wang

Measuring the Saturation Flow, Lost Time, and Passenger Equivalent Values of Signalized Intersection 312

Henk J. van Zuylen and Jie Li

One Approximate and Simple Method for Signal Timing of Overlap Phasing 330

Jingyuan Wang and Wencai Liu

Arterial Signal Timing and Coordination: Sensitivity Analyses and Partition Techniques 338

Wei (David) Fan and Zong Z. Tian

Green Wave Control of Traffic Signal of Inducible Pre-Formation at Section of Road 351

Shuqiao Chen and Jingyi Sun

Optimizing Signalized Roundabout by Computer Simulation (Case Study: Pelem Gurih Roundabout, Yogyakarta) 357

Ahmad Munawar and Sigit Haryanto

Dynamic Traffic Flow Equilibrium and Coordinated Signal Control for Ramps and Adjacent Intersections 365

Lixia Bao, Xiang Gao, Weiquan Zhu, and Hong Chen

Design of Signal Timing Optimization Based on the Unsymmetrical Traffic Flow of Signalized Intersections 374

Zhongyi Zuo, Guangchuan Yang, and Chunfu Shao

A Study on the Improvement of T-Intersection Signal Timing 386

Rui Li and Wenquan Li

Effects of the Spatial Distance between Two Adjacent Bus Stops on Traffic Flow 394

Tieqiao Tang, Haijun Huang, Yan Li, and Xin Bai

Traffic Stream Modeling under Heterogeneous Traffic Conditions 401

Ajitha Thankappan, Yamem Tamut, and Lelitha Vanajakshi

A Novel Method for Traffic Bottlenecks Identification in Urban Traffic Network 412

Yu Zhou, Jianming Hu, Yi Zhang, and Huasheng Yang

The Area Traffic Control System Evaluation Method Based on Travel Time Reliability 423

Zhiyong Guo and Wei Wang

Evaluation on Vehicle Restriction Measure in Beijing 433

Shuling Wang, Chunfu Shao, Fang Wang, and Jianping Sun

444

Traffic Flow Characteristics of Urban Expressway in the Period of Ice and Snow of Cold Areas
Enhui Xing and Rui Wang

Performance Evaluation Model for Traffic Networks **451**
Yang Jiang, Jianjun Wu, and Huijun Sun

The Adjustment Factor of Road Section Crosswalk on Urban Road Capacity **459**
Aizeng Li, Xinsheng Song, Binghua Wu, and Lihua Liu

Effects of Speed Difference on Capacity of Freeway **469**
Hao Wang and Wei Wang

Spatial Correlation Analysis of Congested Links in Urban Traffic Networks **477**
Dan Ma, Huijun Sun, and Linghui Han

Improvement Project Choice Based on Efficiency Analysis **486**
Gu Wang, Shengxue Zhu, and Xiucheng Guo

A Distributed Shortest-Path Routing Algorithm for Transportation Systems **494**
Bin Zou, Jianming Hu, Qi Wang, and Gangkai Ke

Multi-Constrained Shortest Path Model and Solution with Improved Ant Colony Algorithm **501**
Yaomin Hu, Yuyin Liu, and Weiming Liu

Congestion Pricing Model Based on Cumulative Prospect Theory **511**
Yuyin Liu, Weiming Liu, Yaomin Hu, and Jie Li

Evaluation of Tools for Incorporating Road User Costs in Pavement Rehabilitation Decision Making on Rural Highways in Saskatchewan **522**
Ningning Zheng, Guoxin Liu, and Yee-Chuang Jin

Comparison of Cordon and Optimal Toll Points Road Pricing Using Genetic Algorithm **535**
Zhi Zuo, Ryo Kanamori, Tomio Miwa, and Takayuki Morikawa

Identifying Key Factors Affecting Students' Travel Modes Using the Multi-Perspectives Diagnosis Approach **545**
Huaguo Zhou, Peter Hsu, and Shaoqiang Chen

Modeling Destination Choice for the Intermediate Stop Trips Included in the Commute Activity Patterns **557**
Min Yang, Wei Wang, Xuewu Chen, and Yingxiang Yang

Analysis on Effective Information and Travel Activity Returning Home by Car under Downpour **566**
Jun Sakamoto and Motohiro Fujita

Behavioral Characteristics of Self-Driving Tourists: A Case Study in Kunming **577**
Baohong He, Jun Chen, and Wei Wang

Route Choice Analysis of Bus-Use Commuters in Yangon, Myanmar **587**
Hironori Kato, Akihiro Inagi, and Phyo Thet Thet Htun

An Inhabitant Travel Time Distribution Model **598**
Junjuan Liu, Fenyi Dong, and Bingjun Li

Traffic Cognitions, Behaviors, and Education of Child Pedestrians: A Case Study in Kunming, China **605**
Yaqin Qin, Jian Xiong, Fengxiang Guo, Huasen Wan, and Xianguang Jia

Comparison of Urban Link Travel Time Estimation Models Based on Probe Vehicle Data **615**
Fangfang Zheng and Henk van Zuylen

Research Review about Moped Accidents and the Necessity of Time Dimensional Management in University Campus Traffic **627**
Carlos Nava Fonseca and Makoto Okumura

Future Forecast of Chinese Trade Amount and International Cargo Flow **638**
Ryuichi Shibasaki and Tomihiro Watanabe

Influencing Trip Departure Decisions: The Role of Institutions **656**
Anthony T. H. Chin

Traffic Modal Splitting Model Based on Game Theory with Incomplete Information **667**
Chunjiao Dong, Chunfu Shao, Yuewen Gao, Yinhong Liu, and Dan Zhao

Inductive-Based Machine Learning and the Application in Assessment of Ship Handling **678**
Xinjie Han, Xianyao Meng, and Song Meng

Travel Mode Choice Modeling Based on Improved Probabilistic Neural Network **685**
Dan Zhao, Chunfu Shao, Juan Li, Chunjiao Dong, and Yinhong Liu

Complex Networks Model for Urban Passenger Transport Systems: A Case Study of Harbin Taxi **696**
Jian Wang, Xiaowei Hu, and Chong Wei

Performance Evaluation of Bus Transit Routes Based on the BCC Model **705**
Yadan Yan, Xiucheng Guo, and Chaoping Liu

Creating Multi-Routes Combination Bus Timetable **712**
Bingqiang Situ, Wenzhou Jin, and Ming Wei

Ant Colony System for Multi-Depot Bus Scheduling Problem with Budget and Pollutant Emission Control **723**
Ming Wei, Wenzhou Jin, Bo Sun, and Xiaoni Hao

A Corridor-Based Earthquake Evacuation Model **733**

Fei Song, Yang Wang, and Xiaokuan Yang

Designing and Implementing a Service Oriented Traveler Information System **741**

Jun Liu, Dongmei Lin, and Chunyan Li

Mobile Sensors as Traffic Probes: Addressing Transportation Modeling and Privacy Protection in an Integrated Framework **750**

Xuegang (Jeff) Ban and Marco Gruteser

Deriving Rules for Trip Purpose Identification from GPS Travel Survey Data and Land Use Data: A Machine Learning Approach **768**

Zhongwei Deng and Minhe Ji

Prediction of Freeway Travel Time in Incident Management Evaluation Based on Genetic Neural Network **778**

Ding He, Fujian Ni, and Shunxin Yang

A Preliminary Test on the Interactions between Land Use and Travel Models **787**

Shengyi Gao, Eric Lehmer, Yang Wang, Michael McCoy, Robert A. Johnston, Dimantha De Silva, John Abraham, and John Dougalas Hunt

Assessment of the Congestion Impacts of Fatal Accidents Using SCOOT Data **800**

Leighton M. O. Cardwell, Haibo Chen, and Margaret C. Bell

Radar Velocity-Measuring System Design and Computation Algorithm Based on ARM Processor **813**

Yi Zeng, Jianmin Xu, and Deng Peng

Critical Speed Prediction at Unsignalized Intersection under Mixed Traffic Condition **823**

Joewono Prasetijo, Molla Mohammad Ali, and Ning Wu

First Order Phase Transition in a New Cellular Automaton Model within the Framework of Kerner's Three-Phase Traffic Theory **838**

Junfang Tian, Bin Jia, Xingang Li, and Fangfei Zhang

Traffic Prediction in Metropolitan Freeways **846**

Carlo Liberto, Roberto Ragona, and Gaetano Valenti

Urban Short-Term Traffic Flow Forecasting Based on the Semi-Variable Cell Transmission Model **861**

Xiaojian Hu, Wei Wang, and Jian Lu

Spatial and Temporal Aggregation Analysis for Space-Mean Speed Estimates Using Autoscope Video Detection Data **872**

Zhongyi Wu, William H. K. Lam, and Mei Lam Tam

The Application of Support Vector Machines (SVM) for Traffic Condition Prediction Using ITS Data **882**

Dazhi Sun, Qixing Wang, and Xinguo Jiang

Applying Locally Linear Embedding on Feature Extraction of Traffic Flow Data **891**
Yenan Chen, Jianming Hu, Yi Zhang, and Di Li

Study on the Evaluation System and Prediction Approach on Traffic Congestion Index of RTMS Data **903**
Zhenjun Wang, Lijuan Que, Jing Liu, and Jizhen Guan

Hardware-in-the-Loop Simulation: Challenges and Solutions **915**
Isaac K. Isukapati, Alixandra Demers, and George F. List

Network-Wide Road Travel Time Estimation with Inconsistent Sensor Data **927**
Shao Hu, William H. K. Lam, Agachai Sumalee, and Anthony Chen

A Robust Kalman Filter Based Data Fusion Method for Traffic Speed Estimation **937**
Depin Peng, Jianping Wu, Chunlu Wang, and Xingquan Zuo

A New Concept of Level of Service for Freeway Weaving, Merge and Diverge Segments, and Its Applications **948**
Ning Wu

A Simulation Study of Environmental Impact of Automated Truck Lanes on Intercity Expressways **959**
Xun Sun, Tomio Miwa, Toshiyuki Yamamoto, and Takayuki Morikawa

Determining Design Speed of Truck Escape Ramp Based on Back Analysis Method **969**
Hongqi Liu, Tao Shen, Fang Wang, and Weihan Zhang

Study of the Effectiveness of Nighttime and Truck Speed Limits **977**
Dazhi Sun

Study on Unsignalized Access Points Using Microscopic Traffic Simulation **990**
Guoqiang Zhang and Jian Lu

Effects of Freeway Exit Ramp Types on Traffic Operations **999**
Yongfeng Ma, Junping Xie, Jian Lu, and Chuan Lu

Capacity Analysis of Type-A Weaving Section by Using Cellular Automata Model **1010**
Xingang Li, Ziyou Gao, and Bin Jia

Length of Acceleration and Deceleration Lane on Urban Expressway **1019**
Xinsheng Song, Aizeng Li, Binghua Wu, and Lihua Liu

Self-Motion Speed Perception on Visual Information in Driving: Modeling and Validation **1029**
Hirofumi Yotsutsuji and Hideyuki Kita

Car Routing and Distribution Model Based on Kirchhoff's Electric Current Law **1046**

Yajing Zheng and Xingchen Zhang

Three Dilemma Zone Strategies for High-Speed Rural Intersections: Comparison of Field Results **1056**

Alixandra Demers, George F. List, and Isaac K. Isukapati

The Method of Survey and Statistics of Passenger Transport Volume in County Highway Networks **1067**

Liqiang Wang, Xiansheng Li, and Yuanyuan Ren

Estimating Dynamic Transport Population for Official Statistics Based on GPS/GSM **1076**

Yinyi Ma, Jan van Dalen, Chris de Blois, and Jo van Nuen

Time Contours: Using IC Toll Information to Describe Public Transport Networks Accessibility **1089**

Lianxiong Gao, Jianping Wu, and Rui Liu

Research and Development of Public Transit Route Inquiry System Based on GIS **1098**

Mingming Chen, Xi Chen, and Huimin Niu

Conceptual Framework and Mathematical Model for Optimizing Train Blocking and Scheduling Synchronization **1106**

Haiying Li, Lingyun Meng, Jianrui Miao, Xi Jiang, and Zhenhuan He

An Energy-Efficient Control Algorithm and Simulation of Urban Rail Transit Systems **1116**

Haidong Liu , Yun Bai, Jing Yang, Fangming Zhou, Baoshan Wang, Peng Du, and Baohua Mao

A Method to Correct the Mileage Error in Railway Track Geometry Data and Its Usage **1130**

Haifeng Li and Yude Xu

Train Departure Mode and Line Capacity in Inter-City Railway **1136**

Peng Du, Baohua Mao, Hui Zhao, Baoshan Wang, and Haidong Liu

Wagon Flow Organization Modeling Based on Distributed Autonomous Decision-Making **1147**

Xuesheng Ji, Xi Jiang, and Wei Wang

Petri Net Modeling of Spatial and Temporal Constraints in Track Allocation Problem at Railway Stations **1158**

Wenzheng Jia, Tinkin Ho, Zhenqi Chen, and Baohua Mao

A Rolling Optimization Approach of Train Scheduling of Passenger Dedicated Lines **1166**

Ying Wang, Jianrui Miao, Fang Lei, and Jun Liu

Improvement of CFG Pile Corrosion Resistance Capability in the Application of Wu-Guang Passenger Transport Special Railway **1175**

Kunlin Ma, Youjun Xie, Guangcheng Long, and Fei Feng

Multistage Linear Prediction Model of Track Quality Index **1183**

Huan Chang, Rengkui Liu, and Wei Wang

Relation between Railway Track Alignment Irregularity and Gross Passing Tonnage **1193**

Ran Guo, Baoming Han, and Futian Wang

A Short-Range Prediction Model for Track Irregularity of a Shorter Section of Track **1203**

Peng Xu, Quanxin Sun, Qing Li, Rengkui Liu, and Futian Wang

A Method for Organizing Long-Distance Passengers on Railway Passenger Dedicated Lines **1217**

Rong Hua, Jun Liu, and Yun Bao

The Operating Conditions of Long-Distance Passenger Trains on Railway Passenger Dedicated Lines **1228**

Yun Bao, Rong Hua, and Jun Liu

Safety Management System in Civil Aviation: Our Perspective **1238**

Leping Yuan, Ruishan Sun, and Dongbin Li

A Slot Allocation Algorithm for the Flights **1244**

Tiangang Zhang, Ronglu Sun, Xiaoyun Hou, and Rentian Yue

GNSS Evolutionary Architecture for Civil Aviation Utilization **1254**

Qian Sun, Jun Zhang, and Yanbo Zhu

Application of Jute Geo-Textile in the Moorum-Sand Mixed Layer as Granular Subbase of Flexible Pavement **1269**

Tapash Kumar Roy

Recent Development of Pavement Engineering Research in Hong Kong **1276**

L. W. Cheung, W. G. Wong, and Gordon L. M. Leung

A Nonlinear Viscoelastic Constitutive Model of Aged Asphalt Mixture **1286**

Xiaoge Tian

Influence Factors and Anti-Rutting Agent of High-Temperature Stability of Asphalt Mixture **1293**

Ping Li, Xiuli Wang, Huishan Li, and Yu Qian

Effect of Activators on Mechanical Properties of Steam-Cured Fly Ash Concrete **1303**

Zhimin He and Junzhe Liu

Effect of Sisal Fiber on Anti-Slide Performance for Asphalt Pavement Surface **1310**

Xiaowen Lei, Xiaoming He, Peng Guan, and Pengfei Deng

Optimization of Steel Fiber Used for Induction Heating in Porous Asphalt Concrete 1320
Quantao Liu, Erik Schlangen, Martin van de Ven, and Marco Poot

Effect of Alternative Materials on Engineering Properties of Alluvial Soil for Construction of Road Subgrade 1331
Tapash Kumar Roy, B. C. Chattopadhyay, and S. K. Roy

Effects of Contaminants on Skid Resistance of Asphalt Pavements 1341
Ping Cao, Xinping Yan, Xiuqin Bai, and Chengqing Yuan

Using Mineral Fibers to Improve Asphalt and Asphalt Mixture Behavior 1352
Lihua Zhao, Jingyun Chen, and Shengwu Wang

Dynamic Analysis of Damage Behavior of Vibration Compaction on Concrete Bridges 1361
Xin Yu and Yuhua Fu

Creep Tests and Viscoelastic Analysis of Rubber Asphalt Concrete 1371
Yuliang Chen

Shrinkage Performance of Cement-Treated Macadam Base Materials 1378
Yan Wang, Xiang Ma, and Zhilin Sun

Vertical Load-Carrying Natural Frequency of Railway Continuous Steel Truss Bridges 1387
Jianying Ren, Mubiao Su, and Qingyuan Zeng

Anti-Cracking Ability of Asphalt Mixture Added with Polyester Fiber 1399
Xin Yu and Lijun Sun

Centrifuge Model Tests on Settlement Controlling of Piled Embankment in High Speed Railway 1407
Changdan Wang, Binglong Wang, Xu Wang, and Shunhua Zhou

Determining the Rational Completion Cable Forces Based on Influence Matrix Method United with Minimum Bending Energy Method 1417
Jun Yang

Light Polymer Cement Mortar Overlay for Bridge Decks 1425
Xuelian Li, Yuliang Chen, and Zhigang Zhou

Numerical Simulation and Experimental Research of Bridge Damage Identification 1432
Alan Jiang

Estimation of Average Vehicle Time Headway for Light Traffic 1439
Baibing Li

Trading Facilitation in the Black Sea Area 1447

Radu Hanzu-Pazara, Paulica Arsenie, Anastasia Varsami, Corina Popescu, Georgiana Sivriu, and Ramona Bejan

Determination of Crash Surrogates Using Tire Skid Marks **1457**

Hongyan Gao and Zong Z. Tian

Extracting Land Cover/Use from Remotely Sensed Imagery: Potentials for Urban Planning

Ming Zhong[1], Yun Zhang[2], Ahad Beykaei[3], Bahram Salehi[4], Mike Ircha[5], Yongdae Gweon[6] and Sheng Gao[7]

[1] Associate Professor, [3] PhD Student, [6] Honorary Research Professor,Department of Civil Engineering, University of New Brunswick, Fredericton, N.B., Canada E3B 5A3; phone 1-506-452-6324; Fax: 1-506-453-3568; Email: ming@unb.ca, beykaei@unb.ca, mike.ircha@mac.com
[2] Prof., Canada Research Chair in Advanced Geomatics Image Processing Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, 15 Dineen Dr., P.O. Box 4400, CANADA E3B 5A3; phone 1-506-453-5140; Fax 1-506-453-4943; YunZhang@UNB.ca
[4,7] PhD Candidate, [8] Master Student, Department of Geodesy and Geomatics Engineering, University of New Brunswick, Fredericton, N.B., Canada

Abstract

Remote sensing has a great potential for enhancing current urban planning processes by offering a holistic view of a study area and providing detailed land cover/use information within. This study intends to evaluate the potentials of remote sensing in urban/transportation planning by studying the accuracy of two image processing software (ENVI and Definiens) on extracting urban land cover/use. Several satellite images from the City of Fredericton, New Brunswick, Canada, including Landsat ETM+, SPOT4, IKONOS, and QuickBird, are used. It is found that medium-resolution images, Landsat ETM+and SPOT4, are only good at extracting large-size homogeneous objects, such as vegetation and water bodies, but less powerful for identifying small urban features, including buildings, streets and parking lots. Later experiments focus on extracting these urban features with very high-resolution imagery from IKONOS and QuickBird. Study results show that both software packages have more or less problems in distinguish parking lots, streets and building roofs because of similar materials used and therefore, very close spectral signatures. The results show the object-oriented hierarchical algorithm applied to QuickBird images offers the highest accuracy for building and street extractions, when compared to other algorithms/images combinations (such as maximum likelihood/nearest neighbor applied to QuickBird or IKONOS). The producer accuracy of the two is 78% and 63% respectively and the corresponding user accuracy is 39% and 56%. The study results clearly indicate that a more advanced

approach using auxiliary data, such as GIS and LIDAR, is necessary to achieve an accuracy level acceptable for any real-world planning applications.
Key Word: remote sensing; image classification; land cover/use; urban planning

Introduction

Remote sensing offers fast and efficient ways to monitor urban development and extract/update land use information. It shows a great potential for enhancing current urban planning processes. Literature review, however, indicates that there is a significant gap between its technology development and practical applications in urban planning because of discipline barriers. It is found in the literature that most of technology development is carried out by geographers and geomatics engineers who are not familiar with urban and transportation processes; on the other hand, most urban & transportation planners use remotely sensed imagery merely as backdrops in their planning exercises because they generally do not know how to extract information from the images. This study is, therefore, intent to address such a gap by studying the effectiveness of two image classification software on extracting urban subzonal land covers and investigating the usefulness/accuracy of such information in urban planning process.

Ideally, the most desirable application of remotely sensed images in urban and transportation planning is to collect information without physical contacting the ground, which would significantly reduce the collecting time and cost. If this technology can be successfully developed, it could provide very detailed land use information within a few days in an urban context, which planners usually expect to obtain using traditional ground survey in several years. However, extracting information from raw imagery data has been proven to be very challenging. By far, most remote sensing researches merely provide land cover classification or at best land use information at very coarse scale which doesn't contribute much useful information in the view of planning process. Nevertheless, it appears that the time of fully exploring the possibility of extracting detailed land use information has come. This study is to test the accuracy of extracted land use information using two pieces of mainstream software and various types of satellite imagery. The intention is to check what we can obtain from current tools and to establish a research agenda upon the findings from such tests.

This paper tests the capability of two popular commercial software, namely ENVI and Definiens, in extracting detailed subzonal land cover/use information with 4 different types of remote sensing imageries, including Landsat ETM+, SPOT4, IKONOS, and QuickBird. The choices of the two pieces of software are based on their popularity and their representation of two distinct image classification methods: pixel-based and object-oriented. The extracted land covers are compared with secondary GIS data that serves as ground truth to determine accuracy or uncertainty level by using such a technology.

Literature Review

Many studies have been devoted to extract urban features from remotely sensed images; however, it is found that the research before year 2000 in this area focused on using Landsat, SPOT, or similar imagery with medium resolution from other satellites to extract land cover/use information (Zhang, 1999, 2000, 2001; Colwell, 1985; Haack, 1999). In general, land cover/use information for an area or a group of objects (e.g., building areas, water surfaces, and treed areas) can be extracted in these previous studies, but details regarding individual objects, such as buildings, parking lots, and streets, required for typical urban and transportation planning applications cannot be obtained. This is mostly like because the very high-resolution imagery was not available at the time.

For the proposed urban applications, satellite and aerial data must meet certain conditions in terms of their temporal, spatial, spectral, and radiometric characteristics (Lo, 1986). Landsat TM data provided researchers improved spatial, spectral, and radiometric resolution for finer urban change detection since 1982 (Gomarasca, 1993; Hill, 1996). It was reported that SPOT HRV was able to provide significantly more urban land use/cover information than Landsat MSS and TM data (Colwell and Poulton, 1985). It should be noted, however, the previous research of extracting land covers utilizing such medium spatial resolution imagery is only acceptable for the application of urban mapping at coarse level for grouping detection (Mundia, 2005; Deng 2005) and therefore, may probably be useful only for preliminary regional analysis.

The relatively new generation of satellite sensor, such as those from QuickBird or IKONOS, can offer a resolution of 1 m or higher. Considering the mixed-pixel problem, it seems that the choice for extracting land uses in urban areas is Very High Resolution (VHR) satellite images. The disadvantage of using such VHR imagery, however, is its limited number of spectral channels. Spectral signature of urban land cover classes such as roads, parking lots and opening build up areas as derived from VHR imagery are often too similar to be separable (Chan et al., 2009). Zhang and Couloigner (2006) argued that the misclassification between roads and other spectral similar object like building and parking lots is a challenging part of road extraction from VHR multispectral imagery. An attempt to overcome this problem, lack of spectral knowledge in VHR data, is to incorporate the spatial information as well as spectral information of the image into the classification procedure. Spatial information may be used in different ways, such as contextual-based, object-based classification approaches or classification with texture (Lu and Weng, 2007).

Despite VHR images provide unprecedented details at a very fine level, there are still many problems in extracting urban features, such as buildings and streets, from them. In this regard, the complexity and heterogeneity characteristics of urban features and the inherent low spectral nature of VHR images are the principal reasons. Current research in this area is intent to improve the extraction accuracy using other techniques and information along with the spectral grey value of VHR images. Chan

et al. (2009) used the Geometric Activity (GA) features which describe the geometric context of a pixel without the necessity of segmentation for detecting man-made objects in urban areas. The results show a substantial increase in classification accuracy for the man-made object classes, such as "roads and buildings with dark roof", after inclusion of GA features as extra band of image information.

Gamba et al. (2007) used VHR images to explore object boundaries in the mapping procedure. They combined neural network (NN) and adaptive Markov random field (MRF) model through a decision fusion process and applied it to hyperspectral and satellite VHR imagery. Their results show the superior performance of the combined model over conventional NN and MRF classifiers.

The similarity of spectral characteristics of urban land cover/use classes requires incorporating geometrical information within the classification process. Bellens et al. (2008) used partial morphological reconstruction to better preserve the shape of objects, which contains the information about the minimum and maximum dimension of urban objects to improve the classification results. Very high spatial resolution satellite images (IKONOS, QuickBird, and OrbView-3), however, is limited to four multi-spectral bands and may have specific limitations for detailed land use extraction. Carleer and Wollf (2006) tested a feature selection method that included the ancillary information of urban features, and showed what features are useful for extracting particular land cover/use classes from VHR satellite images. Many studies (Xiao, 2004; Wei, 2005; Pacifici, 2007; Mohaptra, 2008) have shown the great potential of very high spatial resolution in detecting urban features, but no solid accuracy test has been done.

It appears from the above literature review that a lot of research works in this area have been done by geomatics engineers and geographers, but few from transportation and urban planners. Therefore, there is a discipline barrier between its technology development and practical application side. Moreover, most of these studies extract land covers only, rather than land uses, which make them not useful in urban planning process. Further, a rigorous accuracy test between the extracted land use quantities with ground truth need to be developed.

Data Sources

Two image sets of VHR satellite, Quickbird and IKONOS and two image sets of Medium Resolution satellite, Landsat 7 and SPOT 4, are used in this study. VHR images cover the south part of the urban area of the case study, the city of Fredericton, and medium resolution images cover larger areas including urban and surrounding area of the city of Fredericton. The images with a projection and datum of “NAD_1983_UTM_Zone_19N” were tested to extract land cover and urban features. In order to obtain images with better quality for improved classification results, the technique of image fusion of high spatial resolution of panchromatic images with high spectral resolution of multispectral images is applied for all above

images. The specifications of satellite images used for testing are summarized in Table 1.

Different kinds of Geographic Information System (GIS) data are used as ground truths to evaluate the classification results. The Fredericton building footprint and the two datasets consist of Service New Brunswick (SNB) Digital Property Map (DPM) datasets and Fredericton Street Network are used to evaluate accuracy associate with building and street extraction, which has a geographic and projection coordinate systems of "GCS_North_American_1983_CSRS" and NAD_1983_CSRS_New_ Brunswick_ Stereographic" respectively.

Study Methodology

Two software packages, ENVI 4.5 and Definiens Professional 5, are used to test different classification algorithms to extract individual land cover/use features like vegetation, water body, bare land, building, road, etc that are useful in urban and transportation planning. "ENVI 4.5" is the pixel-based classification software while "Definiens Pro 5" is the object-oriented software. The algorithm of maximum likelihood is tested within ENVI 4.5 and nearest neighbor and hierarchical methods are tested within Definiens Pro 5.

All satellite images listed in Table 1 are classified based on the sophisticated levels defined in Table 2 with ENVI and Definiens software. Basically the first level – Level 1 is to classify urban area into pervious (vegetation, grass, soil, water etc.) and impervious surfaces (man-built areas, such as roads, parking lots, and buildings). Then at the second level – Level 2, the pervious surface is divided into water, soil, and vegetation, and the impervious surface is further separated into area of traffic and building. Then this disaggregation is continued to Level 3 as shown in Table 2.

The following two statistics are calculated for accuracy assessment purpose.

User Accuracy. A user accuracy of classification is used to find the proportion of correct classified area in each map classification category/class to the total classified area of that category. On the other hand this ratio represents how much the area of a given class is classified to other classes. This ratio is defined as Eq. (1).

Table 1. Specification of satellite imagery for testing area.

Sensor		Bands	Resolution	Size (pixel)	Data depth	Collection
Landsat	MS	6	30	668*668	8 bit	2002
ETM+	Pan	1	15	1336*1336		
SPOT4	MS	4	20	1002*1002	8 bit	2007
	Pan	1	10	2004*2004		
IKONOS	MS	4	4	850*850	11 bit	2002
	Pan	1	1	3400*3400		
QuickBird	MS	4	2.8	1214*1214	11 bit	2002
	Pan	1	0.7	4856*4856		

Table 2. Different levels of land cover/use categories used for software testing.

Level 1	Level 2	Level 3
Pervious surface	Water	Water
		Shadow
	Soil	Soil
	Vegetation	Grass
		Tree
Impervious surface	Area of Traffic	Road& Street
		Parking Lot
	Building	Bright-roof
		Dark-roof

$$UA = \frac{A_C}{A_{TC}} \times 100 \quad (1)$$

where A_C is area of correct classification for each category and classifier and A_{TC} is total classified area of each category and classifier.

Producer Accuracy. A producer accuracy of classification is used to find the proportion of correct classified area for each category/class to the total area of ground truth of that category that represents how much the area of other classes is classified to a given class. This ratio is defined as below:

Study Results

First, the three level of classification hierarchy shown in Table 2 are used to explore

$$UA = \frac{A_C}{A_{TG}} \times 100 \quad (2)$$

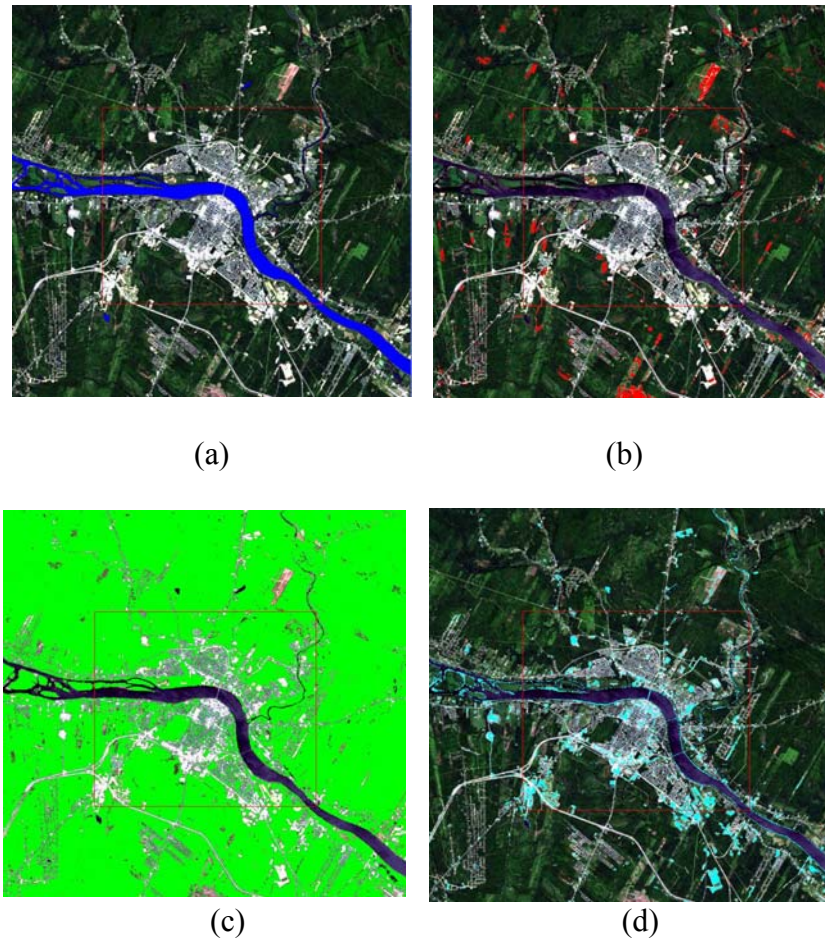
where A_{TG} is total classified area of ground trust for the given category.

The capability of ENVI 4.5 software in extracting urban land covers. It is found that at the first level, the software can well differentiate pervious and impervious surface for all four types of images. However, at the Level two, the software produces a good classification for soil, water and vegetation with Landsat ETM+and SPOT4 images, but results in many misclassifications between buildings and streets, as shown in Figure 1. Such results are expected as these images are featured with high spectral resolution, but low spatial resolution and they are good at extracting large, homogeneous areas with distinct spectral signatures. However, it should be noted that their low spatial resolution (10 to 30 m per pixel) prevent them producing any useful results for discovering man-built features, such as buildings and streets.

Secondly, VHR images from QuickBird and IKONOS are used within ENVI 4.5 to extract the nine categories of urban features at the Level 3 shown in Table 2. It is found, however, at the Level two and especially three, there is a problem of differentiating traffic areas (i.e. streets and parking lots) with buildings. Moreover, a

few building shadows are classified as water. Most of traffic areas are identified, but a lot of commercial building roofs are also miss-classified as parking lots. A lot of streets and parking lots are classified as buildings in this case. It appears that the software, ENVI, only looks at the grey value of each pixel to determine their classes and therefore, it cannot differentiate man-built features using similar materials (e.g., concrete) and therefore, with close spectral signatures.

IKONOS and QuickBird images are used to extract urban features at the Level 3 with the following two algorithms: nearest neighbor and hierarchical classification from Definiens Pro 5 software. The process is done in two steps within the Definiens software: segmentation and classification. The first step – segmentation is toIKONOS and QuickBird images are used to extract urban features at the Level 3 with the following two algorithms: nearest neighbor and hierarchical classification from Definiens Pro 5 software. The process is done in two steps within the Definiens software: segmentation and classification. The first step segmentation is to“segment” or “divide” images into meaningful image units and the second step is to classify these units or objects using the two classification methods, Nearest Neighbor and Hierarchical classification.



(a) (b)

(c) (d)

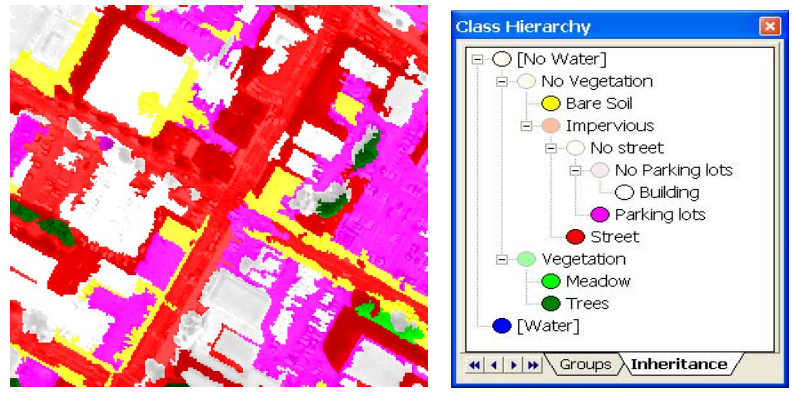
Figure 1. Maximum likelihood classification results of Landsat ETM+for (a) water, (b) soil, (c) vegetation, and (d) building.

The quality of nearest neighbor classification depends on training sample collection and therefore the results are highly dependent on operator's experiences and choices. On the other hand, VHR IKONOS and QuickBird imagery contains too many, diverse and complex objects and the classification results from such a method is not satisfactory. Rule-based hierarchical classification is tried next. It is found that reasonable classification results for buildings, streets and parking lots can be obtained with IKONOS and QuickBird imageries, as shown in Figure 2.

However, it should be noted that there are still misclassifications between the urban features with similar surfaces. For example, the middle section of the road across from top left to lower right is misclassified as parking lot (shown in pink) and the parking lot right below the road at the top left corner is classified as building. In addition, several buildings in the center were classified as "street". For classes of grass and tree, classification result is quite satisfactory, however, some buildings were classified as soil and some streets lost their linear structure.



(a) (b)



(c)

Figure 2. (a) original image of Fredericton downtown, (b) segmentation process with IKONOS and hierarchy method, (c) classification result

Table 3. Accuracy assessment for the Definiens software of Hierarchical Algorithm

Classification Algorithm	Classes	user accuracy (%)	producer accuracy (%)
Hierarchical (IKONOS)	Building	31%	61%
	Street	52%	59%
Hierarchical (QuickBird)	Building	39%	78%
	Street	56%	63%

The statistical evaluations shown in Table 3 indicate that the combination of hierarchical algorithm and QuickBird imagery has the highest accuracy in building extraction, with a user accuracy (UA) of 39% and a producer accuracy (PA) of 78%. The lower value of UA may indicate a better set of rules should be explored for this class because a lot of other land uses were miss-classified as the building. The PA of 78% is quite impressive on the other hand, as it indicates that after considering miss-registration, around 80% of buildings can be successfully extracted at the subzonal level. The experiment also shows a reasonable accuracy for street feature extraction. The UA and PA for the street category is 56% and 63% respectively, as shown in Table 3.

Conclusion and Recommendation

Literature review indicates that there is a significant discipline barrier between technology development of remote sensing and its applications in urban planning. Most of research works have been carried out by researchers from geomatics and geography area and they lack knowledge/insight about how to suit the extracted land cover/use information in urban planning processes. On the other hand, urban planners are rarely equipped with tools and skills for extracting land cover/use information from pixel-based images. In most cases, remotely sensed imagery has been merely used as backdrops. This study is intent to address such a gap by studying the effectiveness of two pieces of popular image classification software on extracting subzonal land cover/use information from the perspective of urban planners.

Four types of satellite images, including Landsat ETM+, SPOT4, IKONOS and QuickBird, are tested in this study. Three mainstream classification methods, namely maximum likelihood, nearest neighbor and hierarchical classification, from ENVI and Definiens are used to classify the four types of images above. It is found that in general, the medium-resolution images, such as Landsat ETM+ and SPOT4, works well in extracting large homogenous features, such as water body and vegetation, due to their wealthier spectral information. On the other hand, they are not useful for identifying man-built “small” urban features, such as building, street, and parking lot, due to their lower spatial resolution and close spectral signatures of the construction

materials used. It should be noted, however, in this regard, that such imagery can be useful in regional analysis by identifying human activities at a large scale.

Latter experiments in this study focus on using VHR images, including IKONOS and QuickBird, to identify detailed urban features, such as buildings, streets, and parking lots. Combinations of two satellite images (IKONOS and QuickBird) and three classification algorithms (maximum likelihood, nearest neighbor, and hierarchical classification) are tested in this study and it is found that hierarchical algorithm has the best result of extracting building and street with PA of 78% and 63% respectively. It is found that miss-registration is an issue for images taken from off-nadir. Visual interpretation shows that there are a lot of misclassifications between the streets and building in all classification results for VHR images. On the other hand, street network GIS data are usually readily available to urban planners. By applying a fuzzy logic analysis based on the street centerline (e.g., any features within 7 meters of each side of the centerline will not be classified as the class other than streets) after classification process the better results could be achieved. It is also found that the timing of satellite imagery is an important issue for urban feature extraction. Study results show that, significant building areas are classified as trees because the areas are covered by leaves. It therefore suggests that the best time for collecting satellite imagery for urban planning applications is in early spring or late fall when vegetation is not in the stage of vigorous growth (Yang and Lu 2002).

It appears from the study results of this paper that extracting urban features solely based on spectral reflection is not successful, although this is the approach that the majority of previous research took. On the other hand, literature review indicates that most remote sensing researches merely provide land cover classification or at best land use information at very coarse scale which does not contribute much useful information in the view of planning process. Therefore, it needs an additional process of converting the extracted land cover information into land uses. However, converting land cover into land use by following traditional image classification process has been proved to be very challenging and unsuccessful.

To further separate man-built features with similar spectral signatures, other complementary data, such as LIDAR and secondary GIS data (building footprint and street centerline), is needed. These data can help further differentiate buildings from parking lots and streets based on their elevation or vicinity to road centerline. It appears that other characteristics of urban features, such as size, shape, height, and location, should be used to develop a set of association rules to successfully convert land covers into land uses.

Acknowledgements

The work described in this paper is mainly supported by a Strategic Project grant from the Natural Science and Engineering Research Council (NSERC), Canada. The authors would also like to thank the City of Fredericton for the data used in this study.

References

Ballens, R., Gautama, S., Martinez, L., Philips, W., Chan, J. C., and Canters, F. (2008). "Improved classification of VHR images of urban areas using directional morphological profiles." *IEEE Transactions on Geoscience and Remote Sensing*, 46(10), 2803-2813.

Carleer, A.P., and Wollf, E. (2006). "Urban land cover multi-level region-based classification of VHR data by selecting relevant features." *International Journal of Remote Sensing*, 27(5-6), 1035-1051.

Chan, J. C., Bellens, R., Canters, F., and Gautama, S. (2009). "An assessment of geometric activity features for per-pixel classification of urban man-made object using very high resolution satellite imagery." *Photogrammetric Engineering & Remote Sensing*, 75(4), 397-411.

Colwell, R. N., and Poulton, C. (1985). "SPOT simulation imagery for urban monitoring: a comparison with Landsat TM and MSS imagery and with high altitude color infrared photograph." *Photogrammetric Engineering and Remote Sensing*, 51, 1093-1101.

Deng, J. S., Wang, K., Li, J., Feng, X.L., and Huang, J.Q. (2005). "Integration of SPOT-5 and ETM+ images to detect land cover change in urban environment." *The proceedings of Geoscience and Remote Sensing Symposium. IGARSS '05, IEEE International***,** Publication Date: 25-29 July 2005, 4, 2284- 2286.

Gamba, P., Dell'Acqua, F., Lisini, G., and Trianni, G. (2007). "Improved VHR urban area mapping exploiting object boundaries." *IEEE Transactions on Geoscience and Remote Sensing*, 45(8), 2676-2682.

Gomarasca, M. A., Brivio, P. A., Pagnoni, F., and Galli, A. (1993). "One century of land use changes in the metropolitan area of Milan (Italy)." *International Journal of Remote Sensing*, 14, 211- 223.

Haack, B., and Bechdol, M. (1999). "Multisensor remote sensing data for land use/cover mapping." *Computer, Environment and Urban Systems*, 23, 53-69.

Hill, J., and Hostert, P. (1996). "Monitoring the growth of a Mediterranean metropolis based on the analysis of spectral mixtures – a case study on Athens (Greece)." In *Progress in Environmental Remote Sensing Research and Applications*, edited by Parlow (Rotterdam: Balkema), 21–30.

Lo, C. P. (1986). *Applied Remote Sensing,* New York: Longman.

Lu, U., and Weng, Q. (2007). "A survey of image classification methods and techniques for improving classification performance." *International Journal of Remote Sensing,* 28(5), 823-870.

Mohaptra, R.P., and Wu, C. (2008). "Subpixel imperviousness: estimation with IKONOS imagery: an artificial neural network approach." Remote Sensing of Impervious Surfaces, 21-37.

Mundia, C. N., and Aniya, M. (2005). "Analysis of land use/cover changes and urban expansion of Nairobi city using remote sensing and GIS," *International Journal of Remote Sensing*, 26(13), 2831-2849.

Pacifici, F., Fabio, D.F., Solimini, D., and Burini, A. (2007). "Urban land cover classification: potential of high and very-high resolution SAR imagery." IEEE International Geoscience and Remote Sensing Symposium, Barcelona, Spain, 1982-1985.

Wei, L., and Prinet, V. (2005). "Building detection from high-resolution satellite image using probability model." Proceedings, 2005 IEEE International Geoscience and Remote Sensing Symposium. 6, 3888-3891.

Xiao, Y., Lim, S.K., Tan, S.T., and Tay, S.C. (2004). "Feature extraction using very high resolution satellite imagery." IEEE International Geoscience and Remote Sensing Symposium, Anchorage, Alaska, 2004-2007.

Yang, C., and Lo, P. (2002). "Using a time series of satellite imagery to detect land use and land cover changes in atlanta, georgia metropolitan area." *International Journal of Remote Sensing,* 23(9), 1775-1798.

Zhang, Q., and Couloigner, I. (2006). "Benefit of angular texture signature for the separation of parking lots and roads on high resolution multispectral imagery. "*Pattern Recognition Letters*, 27, 937-946.

Zhang, Y. (1999). "Optimization of building detection in satellite data by combining multispectral classification and texture filtering." *ISPRS Journal of Photogrammetry & Remote Sensing*, 54(1), 50-60.

Zhang, Y. (2000). "Method for completion and accuracy improvement of multispectrally classified urban rivers." *Photogrammetric Engineering & Remote Sensing*, 66(8), 991-999.

Zhang, Y. (2001). "Detection of urban housing development by fusing multisensor satellite data and performing spatial feature post-classification." *International Journal of Remote Sensing,* 22(17), 3339-3355.

Backcasting Assessment of Strategies for Efficiently Sustainable Urban Transport Developments of Developing Cities

Xuesong Feng[1], A. Fujiwara[2], J.Y. Zhang[3], X.J. Niu[4] and Y. Hayashi[5]

[1] Research Fellow, Dr. Eng., Graduate School of Environmental Studies, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan. Email: charles.x.feng@gmail.com

[2] Professor, [3] Associate Professor, Dr. Eng., Graduate School for International Development and Cooperation, Hiroshima University, 1-5-1 Kagamiyama, Higashi-Hiroshima 739-8529, Japan. Email: afujiw@hiroshima-u.ac.jp, Email: zjy@hiroshima-u.ac.jp

[4] Associate Professor, Dr. Eng., Department of Transportation Management & Engineering, Chinese People's Public Security University, No.1 Muxidinanli, Xicheng District, Beijing 102623, P.R China. Email: gadxnxj@sina.com

[5] Professor, Dr. Eng., Director of International Research Center for Sustainable Transport and Cities, Graduate School of Environmental Studies, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan. Email: yhayashi@genv.nagoya-u.ac.jp

Abstract

In the viewpoints of both sustainability and development, this paper attempts to establish a backcasting analysis framework to assess the urban and transport policies for the developments of developing cities in the future. The proposed backcasting approach is based on a newly developed integrated model; and the stochastic frontier analysis method is adopted in the backcasting process to define the environmental efficiencies of various comprehensive policies' applications. With the urban transport development of Jabodetabek of Indonesia in the future studied using this backcasting framework, it is confirmed that such backcasting approach can

substantially offer a reasonable assessment measure for the sustainable and efficient urban transport developments of developing cities.
Key Words: backcasting; stochastic frontier analysis; sustainability; developing city

Introduction

Sustainability has become a common awareness in all kinds of fields today. Nevertheless, how to deal with the relationship between sustainability and development (especially for a developing city) is turning into a controversial question (MacDonald, 2005). The usual assumption is that it is difficult to get a full achievement of development in an environmentally friendly manner (Robèrt, 2000). Therefore, to comprehensively study the sustainable growth in particular for developing cities with reasonably effective analysis and planning method is becoming more and more important.

As for the sustainability issues, the "Backcasting" approach has been widely applied in recent years based on the "principles of success" (MacDonald, 2005). And its origin goes back to the energy studies in the 1970s; the focus shifted towards exploring sustainable futures in the early 1990s (Quist and Vergragt, 2006). The interest in backcasting for sustainability is still growing. The process of the "Backcasting" study is that envisioning oneself acting in a desirable future where the principles for success have been met, and then planning what have to be done now to move towards that point (Robèrt, 2000), which is about not only how desirable futures can be attained, but also how undesirable futures can be avoided or anticipated (Robinson, 1990). This is quite different from the commonly applied "Forecasting" strategy starting planning procedure from today's situation, then projecting today's problems and trends, and finally proposing the "realistic" solutions onto the future (Robèrt, 2000), which is very difficult to control the outcomes of the proposed solutions. There have been numerous backcasting analyses on various aspects including mobility studies by e.g. Falkenmark (1998), Banister *et al*. (2000), Geurs and van Wee (2000), and Marchau and van der Heijden (2003). Attempts have also been made to combine backcasting with other approaches, such as works of Höjer (1998), Marchau and van der Heijden (2003), Robèrt (2003; 2005), Dortmans (2005), MacDonald (2005) and Robèrt and Jonsson (2006).

However, these previous backcasting studies are principally on the sustainability issues of developed countries. In view of various limitations to the urban transport studies for developing cities, e.g., extreme scarce of survey data, embarrassed financial situation and inadequate planning (JICA and BAPPENAS, 2004; Boyce and Xiong, 2007), the existing backcasting approaches are difficult to be suitably applied for the backcasting analysis on both sustainability and efficiency of the urban

transport development of developing cities in the future. As a result, this study has developed a new backcasting analysis framework to assess the urban and transport strategic policies for a developing city towards its future traffic emission targets with regard to sustainable and efficient urban transport development. The urban and transport policies in the proposed backcasting framework are evaluated with synthesized forecast techniques based on a newly developed integrated travel demand model to consider the behavioral interdependency mechanisms between transport networks, spatial distributions of population related factors and trip-making behavior. Moreover, the Stochastic Frontier Analysis (SFA) (Kumbhakar and Lovell, 2000) is also adopted into the work to define the Environmental Efficiency (EE) of each application of urban and transport policies. In this research, Jabodetabek of Indonesia, one of the representative developing megacities, is selected as the study area. The analysis results empirically confirm that this backcasting approach can present a practical and reasonable assessment measure to find an environmentally sustainable and efficient "path" for the urban transport development of a developing city.

The following parts of this paper are organized as follows. In Section 2, the study area, Jabodetabek metropolitan area of Indonesia, is briefly introduced. Then, Section 3 explains the proposed backcasting approach. Next, in Section 4, according to the traffic emission target, the comprehensive planning policies for Jabodetabek in the future are comparatively analyzed using this backcasting approach in the viewpoint of environmentally sustainable and efficient urban transport progress. Finally, Section 5 summarizes the findings and points out the future research issues.

Study Area

As a representative developing megacity with a population of about 21 million in 2002, Jabodetabek is strategically the most important metropolitan district of Indonesia. To attract more domestic and foreign investment is one of the vital issues towards this region's further economic growth and development. However, its inefficient transport system and the traffic congestion in the urbanized area are indubitable headaches and make this metropolitan area less attractive. First of all, the road network improvement in Jabodetabek cannot keep pace with the rapid urban development. A road network hierarchy has not been established; and the existing road network is neither well equipped. Due to such deficient road network and about 10 times' increase of the volume of commuting trips from surrounding areas to the central part of Jabodetabek between 1985 and 2002 (JICA and BAPPENAS, 2004), the traffic congestion in the central area of Jabodetabek has been becoming more and more serious and the traffic accidents have been increasing as well. Moreover, with respect to its urban railway system, it is characterized with low transport capacity,

delayed schedules, ill-equipped train cars and insecurity. From 2000 to 2002, 174 accidents were reported including serious train collisions and crash accidents (JICA and BAPPENAS, 2004). Furthermore, the bus service in Jabodetabek is also quite poor in many aspects, e.g., delayed schedules, long waiting time and so on. To the worse, it seems difficult to foresee investments in large-scale transport infrastructure improvement projects by the public sectors, due to their difficult financial situation (JICA and BAPPENAS, 2004). Because the inconsistency between rapid urban growth and comparatively lagged development of transport system is typically shown in Jabodetabek, and also comparatively adequate data of Person Trip (PT) survey and transport networks with comparatively high quality can be obtained, Jabodetabek is selected as the study area for this research. These data for Jabodetabek in 2002 is provided by the Japanese International Cooperation Agency.

The Proposed Backcasting Approach

Framework of Backcasting Analysis The proposed backcasting analysis framework for Jabodetabek is presented in Figure 1. In fact, this backcasting approach is combined with forecasting methods and consists of four elements: 1) description of "Actuality", 2) decision of "Target", 3) study of applicable "Policies" possibly able to achieve the "Target" by utilizing the SFA (Kumbhakar and Lovell, 2000), and 4) the final selection of the "Policies" with the biggest value of the EE studied according to the SFA to determine both sustainable and efficient "path". This newly proposed backcasting analysis approach can be widely applied to generally analyze various issues concerning the strategically comprehensive sustainable urban transport developments especially in the context of developing cities. In order to make the procedure of the proposed backcasting approach easy to be clearly understood, the details shown in Figure 1 for the analyses in this study are explained by taking the case of urban transport development in Jabodetabek from 2002 for instance. Firstly, the urban transport "Actuality" of Jabodetabek in 2002 is mapped according to the above-mentioned data concerning the PT survey and the transport networks in Jabodetabek in 2002. Because the traffic emission factors of CO_2 cannot be accurately decided for the Jabodetabek case due to data limitation and the traffic emissions of NO_x, SO_x, HC and CO show the same pattern as the emission trend of CO which is the main representative traffic emission pollutants in this metropolitan region, the control of total traffic emission amount of CO is decided to be the "Target" for the environmentally sustainable development of Jabodetabek in the future.

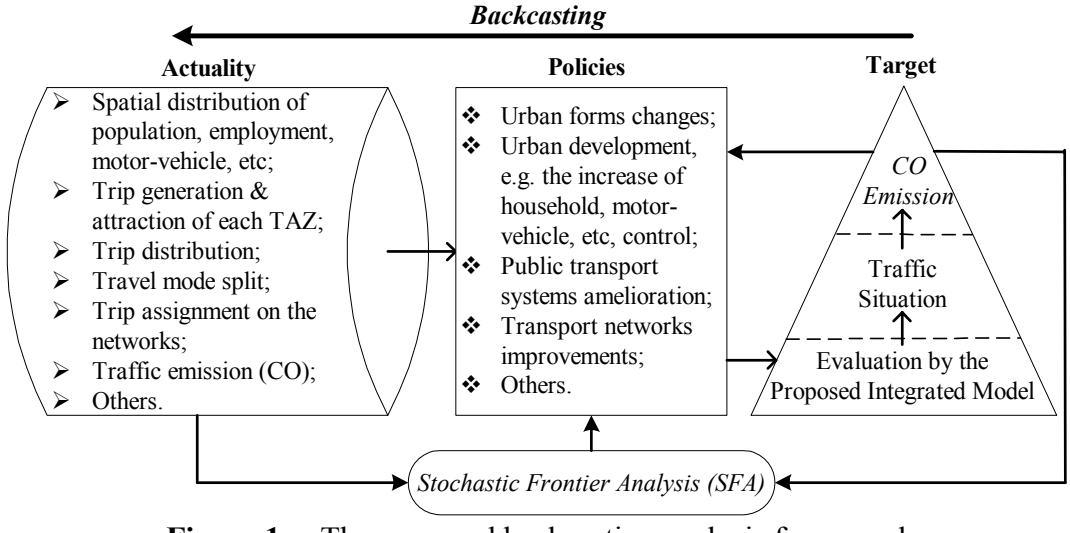


Figure 1. The proposed backcasting analysis framework.

For the urban transport studies of developing cities, because the elementary trip survey data are still relatively scarce and the available data, mostly with very poor qualities, are often not shared among planning agencies (Boyce and Xiong, 2007), the advanced modeling approaches, mostly requiring more detailed trip and/or activity data, are very difficult to be applied. Therefore, generally speaking, the sequentially estimated conventional four-step modeling procedure (McNally, 2000a), though more and more criticized due to its serious lack of a valid representation of underlying travel behavior (McNally, 2000b), is still most widely applied because of its practicability. To properly analyze the urban transport of a developing city in the viewpoint of trip-making behavior, one of the suitable solutions is the feedback method which can actually trace its root to Evans' iterative method put forward in 1973 (Levinson and Kumar, 1994). The research findings on the feedback mechanisms and the subsequently stimulated studies of the combined equilibrium models have provided the insights on how to solve the sequential procedure with feedback (Boyce *et al*., 1994). Referring to these previous studies, a new integrated travel demand prediction model has been developed by Feng *et al*. (2009) to represent the behavioral interference between transport networks, distributions of population-related factors and travel behavior. This newly developed integrated model is applied in the proposed "Backcasting" study framework shown in Figure 1, to comparatively evaluate the traffic emission amounts of *CO* because of the applications of various comprehensive urban and transport "Policies". The adopted values of the traffic emission factors for *CO* are according to the work of Purwanto (2001). Accordingly, it is able to get the primary alternative policies which will achieve the traffic emission target. These alternatives will be further analyzed using the EE index to select the most sustainable and efficient development strategy for the

comparatively full achievement of the urban and transport growth of Jabodetabek in the future.

Environmental Efficiency Definition. For the system efficiency analysis, SFA and Data Envelopment Analysis (DEA) are the two dominating approaches. The DEA method can calculate the efficiency even in case of very small sample size, but the calculation results are very sensitive to the set of input variables. In contrast, SFA technique has an advantage of allowing for random shocks and measurement errors, also with the assumption of specific statistical distribution of both random shocks and error terms. Since the significant statistical characteristics of SFA approach, it has been selected to calculate the EE values for each of the comprehensive policies for the sustainable urban transport development of Jabodetabek in the future, considering the imperfect quality of research data for Jabodetabek.

From an *output-oriented* perspective (Kumbhakar and Lovell, 2000), the environmental frontier model for the traffic emission of *CO* is first shown by equation (1).

$$\ln y_{CO} = \beta_0 + \sum_{\mathrm{k}} \beta_{\mathrm{k}} \ln x_{kCO} + v_{CO} - u_{CO} \tag{1}$$

Where,

y_{CO} : output variable, i.e. the amount of traffic emission of CO,

x_{kCO} : the k^{th} input variable, e.g. the road length and car number,

v_{CO} : the two-sided "noise" component for ,

u_{CO} : the nonnegative technical inefficiency component, and

β_0, β_k : Coefficients to be estimated.

Moreover, $v_{CO} \sim$ iid $\mathbf{N}(0, \sigma_v^2)$; $u_{CO} \sim$ iid exponential; v_{CO} and u_{CO} are distributed independently of each other, and of the regressors. Then, the EE of the traffic emission of CO can be expressed as equation (2) in the production frontier viewpoint for the traffic emission of *CO*, with the consideration of environmental sustainability in relativity.

$$EE_{CO} = \frac{E(u_{CO} \mid \varepsilon_{CO})}{\beta_0 + \sum_{\mathrm{k}} \beta_{\mathrm{k}} \ln x_{kCO}} \tag{2}$$

Where, ε_{CO} is equal to $(v_{CO} - u_{CO})$, and $E(u_{CO} \mid \varepsilon_{CO})$ represents the mean of the conditional distribution of u_{CO} given ε_{CO}, which is illuminated in equation (3).

$$E(u_{CO} \mid \varepsilon_{CO}) = -\varepsilon_{CO} - (\sigma_v^2 / \sigma_u) + \sigma_v \left[\frac{\phi((\varepsilon_{CO} + (\sigma_v^2 / \sigma_u))/\sigma_v)}{\Phi((-\varepsilon_{CO} - (\sigma_v^2 / \sigma_u))/\sigma_v)} \right] \quad (3)$$

Where, σ_v and σ_u are the standard deviations of v_{CO} and u_{CO} respectively; $\phi(\cdot)$ and $\Phi(\cdot)$ are, in a respective manner, the standard normal density and cumulative distribution functions.

Because of the data inadequacy, the environmental frontier model, i.e. the equation (1) applied for Jabodetabek, is estimated with the PT data and the traffic emission amounts of *CO* for each TAZ in 2002. (Based on the actuality of Jabodetabek in 2002, this metropolitan area has been divided into 336 TAZs.) Then, the EE for each TAZ are able to be calculated according to the road length and the number of households in these TAZs in 2002. The mean EE for all the TAZs will be further calculated to represent the overall EE of the whole urban area for various policies' applications to the urban transport development of Jabodetabek region. The validation results of equation (1) for Jabodetabek case are presented in Table 1. According to the values of the T-Ratios shown in Table 1, it is able to easily conclude that the estimations to the coefficients of the variables adopted here in equation (1) applied for the case of Jabodetabek have got both logical and statistical significances.

Backcasting Analysis for the Sustainable Development of Jabodetabek

With the newly developed integrated travel demand model and the afore-explained environmental frontier model estimated, now it is enabled to analyze the sustainable urban transport development of Jabodetabek in the future, following the proposed backcasting analysis framework. In the first place, it is necessary to set the detailed sustainable target of the *CO* traffic emission amount for Jabodetabek in the future and offer the "Policies" to be applied and possibly allowing the target to be achieved.

Setting Target and Providing Planning Policies In consideration of the rapid development of Jabodetabek, motorization in this large developing metropolitan area

Table 1. Estimation to the environmental frontier model

Variables	Coefficients	*z*-statistics
Constant	6.94 (β_0)	6.05
Amount of Households	0.48 (β_1)	4.00
Road Lengths	0.95 (β_2)	13.54
Model Log Likelihood	-442.21	

will be further accelerated without doubt in the future (JICA and BAPPENAS, 2004). This will reasonably increase the traffic emission in Jabodetabek. Maybe in the future, new vehicle technologies (e.g. electric vehicles, hybrid cars, etc) and distribution of renewable fuels (produced from a mix of bio-, solar-, wind- and hydro power) will work to make the traffic emission comparatively decreased in some degree, but it is unpredictable for the coming date of such new technologies and renewable fuels stably applied. To the worse, the embarrassing financial status facing up to the Jabodetabek governments (JICA and BAPPENAS, 2004) is one of the biggest obstacles for the swiftly effective applications of these new inventions in the future. Moreover, as explained before, investments in large-scale transport infrastructure development projects by the public sector are very difficult to be foreseen also due to the strait financial situation (JICA and BAPPENAS, 2004). In view of these realities, the total traffic emission volume in Jabodetabek in the future is going to be increased predicatively mainly due to the further accelerated motorization.

As a result, about the sustainable urban transport development target for Jabodetabek here, controlling the increase speed of traffic emission amount is another choice. Because there are not convincing authority criteria on the traffic emission control for Jabodetabek in the future able to be followed, the target here is generally set as the case that the average annual increase ratio of the total traffic emission amount of *CO* in Jabodetabek per day is no higher than 2.50% from 2002 to 2050, considering the comprehensive urban and transport development of Jabodetabek in the future. Certainly other factors, e.g. traffic congestion release degree and minimization of the car-based mobility, are also able to be utilized together to compose comprehensive sustainable targets, yet in order to make the proposed backcasting framework easily understandable and clearly operable, only the control of the increase ratio of the total traffic emission amount of *CO* is defined here as the simple sustainable development target for the case of Jabodetabek. With this "Target" decided, a total of eight comprehensive planning "Policies" is offered in assumption in Table 2. In order to

Table 2. Comprehensive urban and transport policies

Policies	Urban Form	Number of Households
P1	Uncontrolled	Uncontrolled
P2	Compact-City Development (CCD)	Uncontrolled
P3	Transit-Oriented Development (TOD)	Uncontrolled
P4	Poly-Center Development (PCD)	Uncontrolled
P5	Uncontrolled	Controlled
P6	Compact-City Development (CCD)	Controlled
P7	Transit-Oriented Development (TOD)	Controlled
P8	Poly-Center Development (PCD)	Controlled

make a simple and clear explanation and also improve the policy maneuverability and feasibility, all these policy packages here shown in Table 2 only include the manipulations to *Urban Form* and *Number of Households*, which is also because of the planning data limitation.

From 2002 to 2020, the road network in Jabodetabek will be developed according to the Jabodetabek governments' road development plan for 2020 shown in Figure 2, in which, the blue and red lines represent the new roads (some ring roads and connector roads to the city center) to be constructed. Because the investments in large-scale transport infrastructure development in Jabodetabek are very difficult to be predicted considering the uptight financial situation confronted with the governments (JICA and BAPPENAS, 2004), it is simply assumed that the road network in Jabodetabek from 2020 to 2050 will remain the same as the one in 2020. Also due to the restriction of the survey data about the public transport systems' planning, the railway and bus networks are unchanged in hypothesis from 2002 to 2050.

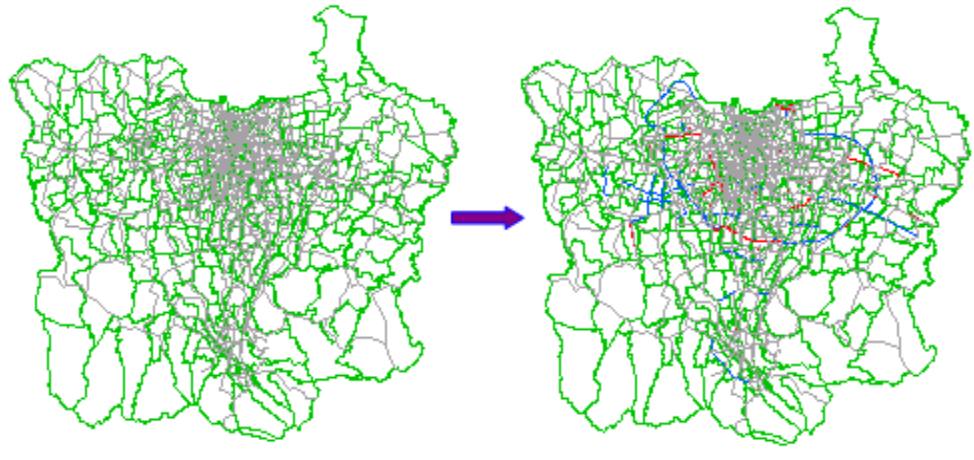


Figure 2. Road network development plan from 2002 to 2020.

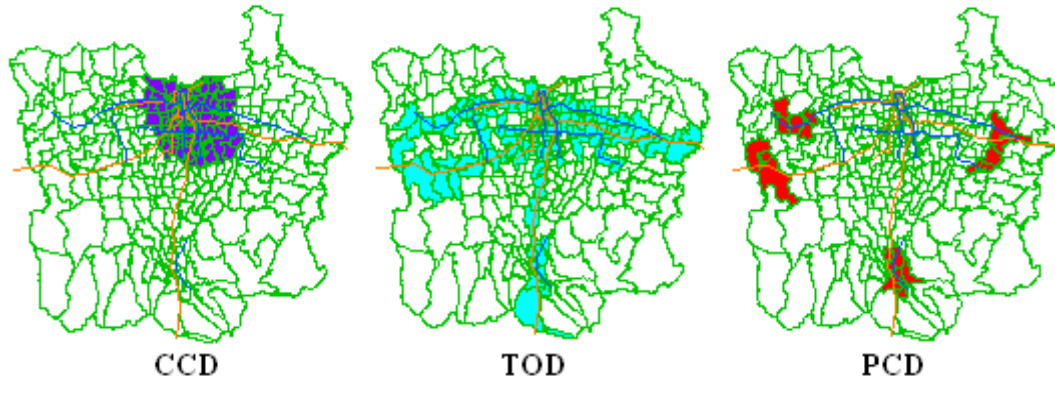


Figure 3. Controlled urban forms for Jabodetabek.

Moreover, as shown in Table 2, with regard to the "Uncontrolled" *Urban Form*, it means that, in the Spatial Autocorrelation step, for the SAR model, the values of the spatial explanatory variables will grow on average for each of the 336 TAZs, except for the mean accessibility which is determined according to the transport conditions and improvements of the transport networks. Consequently, the total amounts of the spatially lagged dependent variables (also the independent variables of the MLR model applied in the trip generation and attraction step) will be distributed on average at these 336 TAZs in a spatially auto-correlative manner. For the urban forms "Compact-City Development (CCD)", "Transit-Oriented Development (TOD)" and "Poly-Center Development (PCD)" shown in Figure 3, the increases of total amounts of the spatially lagged dependent variables are the same with the corresponding cases of the "Uncontrolled" *Urban Form*. Nevertheless, the growth of the spatial explanatory variables will mainly happen at the TAZs respectively in the central urban area, along the bus and railway lines and in the four selected sub-centers at the suburban areas. As a result, the total amounts of the spatially lagged dependent variables will, in an oriented manner, lean to distribute at these corresponding TAZs, with spatially lagged autocorrelation.

Furthermore, because of its important effect and comparatively better data quality, the *Number of Households* in each of the TAZs in Jabodetabek has been selected as one of the important factors in the comprehensive urban and transport development policies. The "Uncontrolled" *Number of Households* shown in Table 2 indicates that the growth ratio of the total households at all of the TAZs is set to be 1.00% annually, which is according to the development experience of Jabodetabek; however as for the "Controlled" case, this growth ratio is assumed to be decreased to 0.20% per year in view of the effort of the governments of Jabodetabek.

Backcasting Analysis. According to the afore-decided target for the traffic emission control of *CO* for Jabodetabek in the future, now it is able to comparatively evaluate the comprehensive urban and transport policies shown in Table 2, based on the traffic emission amounts of *CO* along with applying these policies simulated by the new integrated model and the EE values of the policies' applications analyzed by the afore-explained SFA method. First of all, calculated according to the PT data for Jabodetabek in 2002 and the emission factors supposed by Purwanto (2001), 39,929,361.54 (gram) is the total traffic emission volume of *CO* per day in Jabodetabek in 2002. Then, the blue thick curve shown in Figure 4 is obtained as another representation of the afore-decided "Target" for the control of the total traffic emission of *CO* in Jabodetabek. All the selected alternative strategic "Policies" able to achieve this "Target" should have the total traffic emission of *CO* lower than the one represented by this blue thick curve.

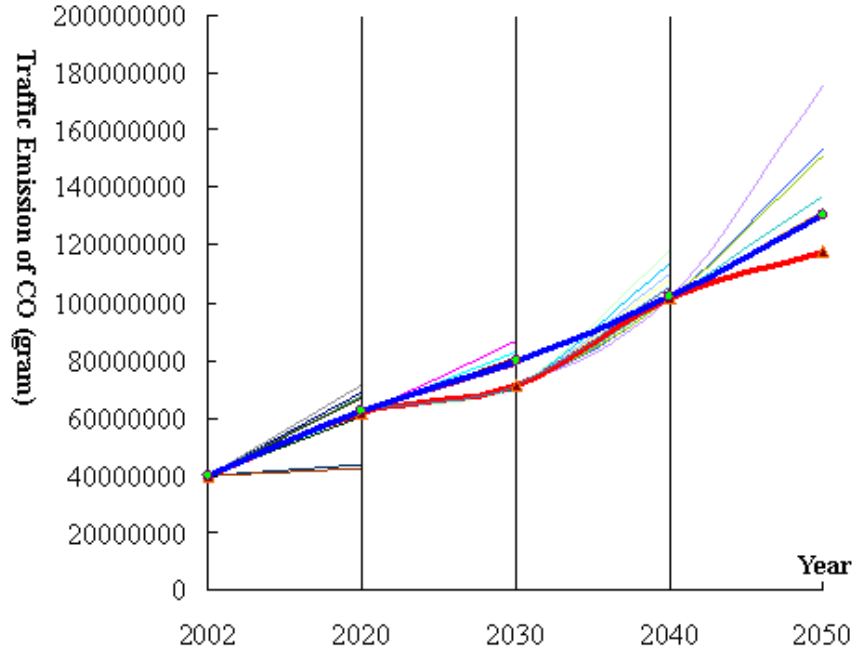


Figure 4. Sustainable urban transport development "path" for Jabodetabek.

From 2002 to 2050, the future planning years for Jabodetabek have been divided into 4 periods, i.e., 1) from 2002 to 2020, 2) from 2020 to 2030, 3) from 2030 to 2040, and 4) from 2040 to 2050. As shown in Figure 4, firstly, for period 1), the *CO* traffic emission volumes of the policies shown in Table 2 are evaluated based on the newly developed integrated model; then, the alternatives (*P2*, *P4*, *P6* and *P8* under the blue thick curve shown in Figure 4) are selected. None of these selected alternatives have the total traffic emission volumes of *CO* in one day more than 62,276,176.82 (gram) from 2002 to 2020 ($62{,}276{,}176.82=39{,}929{,}361.54\times(1+0.025)^{2020-2002}$). Then, it can be further found that compared with the other alternatives, the mean EE of *P2* for all the TAZs in Jabodetabek is the highest. Consequently, with the other tentative alternatives (*P4*, *P6* and *P8*) abandoned, *P2* is decided as the efficiently sustainable "sub-path" for period 1); and the outputs of *P2* become into the inputs of the same kind of study made again to the applications of all the policy-packages shown in Table 2 for period 2). Such comparative analysis will be continued till the period 4). Thereafter, the "sub-paths" for each of these planning periods consist of the efficiently sustainable "path" for the comparatively full urban transport development of Jabodetabek from 2002 to 2050.

Represented by the red thick curve shown in Figure 4, this sustainable and efficient "path" is made up of the "sub-paths" below the blue thick curve in each time period, i.e., *P2*, *P8*, *P7* and *P8*, respectively for each time period from 2002 to 2050. In other words, the sustainable "path" for the comparatively sufficient urban transport development of Jabodetabek from 2002 to 2050 is successively made up of CCD

with the new roads constructed and the households' growth uncontrolled (comprehensive central development), PCD with controlling the households' increase (sub-central development with urban growth control), TOD with the control of the rise of households (connecting development between central and sub-central areas still with the control of urban growth), and again PCD with controlling the household increase (sub-central development improvement and the urban growth control). Simply on the perspective of urban transport planning, such progressive urban transport development (i.e. successively CCD, PCD, TOD and PCD) is also admittedly reasonable and efficient.

Conclusions

In this paper, a new backcasting approach with forecasting techniques synthesized has been established to analyze the future sustainable urban transport development of Jabodetabek of Indonesia, one of the representative developing megacities today. In this backcasting analysis framework, a newly developed integrated travel demand model is applied to study the urban and transport planning policies for the traffic emission from the perspective of environment protection. Moreover, from the viewpoint of the sustaining growth of a developing city, the stochastic frontier analysis has been adopted to further analyze the environmental efficiencies of the primarily selected alternatives basically satisfying the traffic emission control target in the future. Following the proposed backcasting approach, finally a sustainable "path" can be found for the environmentally efficient urban transport evolution of Jabodetabek from 2002 to 2050. This "path" consists of CCD with the new roads constructed and the household growth uncontrolled from 2002 to 2020, PCD with controlling the household increase from 2020 to 2030, TOD with the control of the rise of households from 2030 to 2040, and again PCD with controlling the household increase from 2040 to 2050. According to the application results of this proposed backcasting analysis to the case of Jabodetabek, it is confirmed that such a backcasting approach can substantially offer a practical and reasonable assessment measure to study and instruct the urban transport developments of developing cities and get a comparatively full achievement in an environmentally friendly manner. However, because the abandoned "sub-paths" for the one planning period have nothing to do with the results of the policies' applications in the next time period, this proposed backcasting study approach cannot absolutely ensure that the decided "path" consisting of its sub-paths will make the urban transport development the most effective. This issue needs to be further improved in future research.

References

Banister, D., Dreborg, K. H., Hedberg, L., Hunhammar, S., Steen, P., and Åkerman, J. (2000). "Transport policy scenarios for the EU in 2020: images of the future." *Innovation*, 13(1), 27-45.

Boyce, D., and Xiong, C. (2007). "Forecasting travel for very large cities: challenges and opportunities for China." *Transportmetrica*, 3(1), 1-19.

Boyce, D. Zhang, Y. F., and Lupa, M. R. (1994). "Introducing "feedback" into four-step travel forecasting procedure versus equilibrium solution of combined model." *Transportation Research Record*, 1443, 65-74. Washington, DC: Transportation Research Board.

Dortmans, P. J. (2005). "Forecasting, backcasting, migration landscapes and strategic planning maps." *Futures*, 37(4), 273-285.

Falkenmark, M. (1998). "Dilemma when entering 21st century: rapid change but lack of sense of urgency." *Water Policy*, 1(4), 421-436.

Feng, X. Zhang, J., and Fujiwara, A. (2009). "Adding a new step with special autocorrelation to improve the four-step travel demand model with feedback for a developing city." *IATSS (International Association of Traffic and Safety Sciences) Research*, 33(1), 44-54.

Geurs, K., and van Wee, B. (2000). "Backcasting as a tool to develop a sustainable transport scenario assuming emission reductions of 80-90%." *Innovation*, 13(1), 47-62.

Höjer, M. (1998). "Transport telematics in urban systems─a backcasting delphi study." *Transportation Research Part D*, 3(6), 445-463.

JICA (Japan International Cooperation Agency) and BAPPENAS (National Development Planning Agency, Republic of Indonesia). (2004). *The study on integrated transportation master plan for Jabodetabek (Phase II), final report & summary report*. Jakarta: Pacific Consultants International Almec Corporation.

Kumbhakar, S. C. and Lovell, C. A. K. (2000). *Stochastic frontier analysis*. Cambridge: Cambridge University Press.

Levinson, D., and Kumar, A. (1994). "Integrating feedback into the transportation planning model: structure and application." *Transportation Research Record*, (1413) ,70-77. Washington, DC: Transportation Research Board.

MacDonald, J. P. (2005). "Strategic sustainable development using the ISO 14001 standard." *Journal of Cleaner Production*, 13(6), 631-643.

Marchau, V., and van der Heijden, R. (2003). "Innovative methodologies for exploring the future of automated vehicle guidance." *Journal of Forecasting*, 22(2-3), 257-276.

McNally, M. G. (2000a). "The four-step model." *Handbook of Transport Modelling* .35-52. New York: Pergamon.

McNally, M. G. (2000b). "The activity-based approach." *Handbook of Transport Modelling*. 53-69. New York: Pergamon.

Purwanto, E. (2001). "The cost benefit analysis air pollution control program for vehicle in Indonesia case study." *Master's Thesis*, Institute for Chemical Engineering and Environmental Technology, Graz University of Technology.

Quist, J., and Vergragt, P. (2006). "Past and future of backcasting: the shift to stakeholder participation and a proposal for a methodological framework." *Futures*, 38(9), 1027-1045.

Robèrt, K. H. (2000). "Tools and concepts for sustainable development, how do they relate to a general framework for sustainable development, and to each other?" *Journal of Cleaner Production*, 8(3), 243-254.

Robèrt, M. (2003). "Incentives and policies toward car-sharing and ride-matching ─ a field study based on employees' preferences in Stockholm." *Company incentives and individual preferences towards sustainable travel alternatives* (pp. 47-73). Stockholm: KTH, Infrastructure.

Robèrt, M. (2005). "Backcasting and econometrics for sustainable planning: Information technology and individual preferences of travel." *Journal of Cleaner Production*, 13(8), 841-851.

Robèrt, M., and Jonsson, R. D. (2006). "Assessment of transport policies toward future emission targets: a backcasting approach for Stockholm 2030." *Journal of Environmental Assessment Policy and Management*, 8(4), 451-487.
Robinson, J. (1990). "Futures under glass: a recipe for people who hate to predict." *Futures*, 22(8), 820-843.

Sustainable Mobility and Sustainable Transport for Medium-Size Cities in China

Prof. Reinhart Kühne[1], Martin Ruhé and Xiaoxu Bei

German Aerospace Center (DLR)-Transportation Studies (VS), Rutherfordstrasse 2, 12489 Berlin, Germany; PH: +49-30-67055-200, Fax: +49-30-67055-202, Email: reinhart.kuehne@dlr.de, martin.ruhe@dlr.de, xiauxu.bei@dlr.de

Abstract

The project METRASYS (Mega Region Transport Systems for China) funded by the Federal Ministry of Education and Research of Germany in the program of research for sustainability and co-funded by the Chinese Ministry of Science and Technology follows an interdisciplinary approach integrating spatial planning (new urbanism), transport science, engineering and political science in order to contribute to the mitigation of climate change. An important part of METRASYS is to assess the environmental impacts of the traffic management system and planned urban traffic developments on the basis of a Business-As-Usual (BAU) scenario. Linking technological developments with urban and transport-planning policies is one of the foci of this project activity: The identification of suitable strategies and policies is a prerequisite to ensure permanent and effective implementation of technological approaches.How the results of the analysis as well as alternative measures and instruments for the development of a climate and energy efficient transport system are discussed with local decision makers is reported. A geo information system (GIS) model is presented for spatial visualization to assist decision makers and stakeholders in their comprehension and evaluation of the effects of policy measures. The BAU scenario is also the basis for the analysis of the suitability of transport management as a Clean Development Mechanism (CDM). It will show how to this extent the collected vehicle probe data can be used.The development of the technical system for traffic management and its installation in Hefei will be reported. All information and results of the traffic management system are open for the public. The dissemination activities aim at a broader audience within China and worldwide. It is aimed at transferring the results of the project to different cultural and institutional contexts. Main written products are more comprehensive discussion papers presenting the state of research, new findings from the project and recommendations. An important task

is the invitation of stakeholders of cities similar in size and function to show the technical system with all its options and functionality. Dissemination includes the participation on practical oriented conferences and workshops summarizing the key points of new findings of the project. The METRASYS homepage will provide up to date information such as reports and an online GIS for stake holders, academia, and the public.The paper will give an overview about the specialties in the project work with the Chinese Partners and will point out the findings and goals of the described showcases.
Key Words: China, Sustainable Mobility, Sustainable Transport, Hefei, Mega Cities

Background

Growing urbanisation and the increasing size of metropolitan regions is a challenge as well as an opportunity for the economic development and social balance of societies particular in developing countries like China. The very dynamical evolution of the cities implies special requirements for mobility concepts. These need to be adapted to possible future developments both with respect to their effects and to their implementation. Special challenges of sustainable transport in mega cities today have to address environmental problems and their social implications (e. g. air quality, noise, land use, fragmentation, security), accessibility and ways to manage existing infrastructure capacities through smart engineering. As an example METRASYS is one of the research projects on mega cities supported by the Federal Ministry of Education and Research of Germany in cooperation with Chinese partners like “Research Center for Software Engineering Technology Anhui Province”, “Tongji University” and “Hefei City People’s Government”. The main objective of the project is to reduce the energy consumption and with this the CO2 production of urban traffic in mega cities. A comprehensive traffic control system will be developed, implemented and applied. To check the efficiency of the systems and for calculating and optimizing respectively the emission reduction additionally analysis tools will be set up - based on real data, specific scenarios, measuring and modeling immissions.

As field trial the capital Hefei of the Anhui province in China is chosen. Hefei is on the verge of becoming a mega city and will reach the critical size probably in the next decade. With its present 5.3 million inhabitants the conurbation of Hefei is an excellent example for a huge number of fast growing medium size cities in China in contrast to the small number of real mega cities. At present 167 cities with a number of inhabitants between 1 and 5 million exist in China but only 4 cities with more than 10 million inhabitants. To increase of attraction of these medium size cities is a contribution to relieve the mega cities. The goal is to decentralize the conurbations instead of establishing large mega cities (>15 million inhabitants), this has the advantage that these medium size cities in contrast to the mega cities still have space for controlled and sustainable development. The very dynamic growth of Hefei requests sustainable approaches and technologies. For example currently a new urban district for 700.000 inhabitants is being realized, all plans and developments have to be matched with such dynamic growth. Therefore is the assessment in detail one of

the most important working areas of the METRASYS project. This is a challenging and complex task, but the experience given by the assessment is an important reference for a transfer of the results to other cities in China and elsewhere in countries with similar challenges.

Main Working Areas

Transit Oriented City Planning. An important part of METRASYS is to assess the environmental impacts of the traffic management system and planned urban traffic developments on the basis of a Business-As-Usual (BAU) scenario. The project team was asked to support the design of an environment and climate friendly urban and transport development project for a specific urban quarter in Hefei. This offers the METRASYS project and its partners the great possibility to actively contribute to sustainable and climate friendly development in Hefei. It is planned to scratch a sustainable development concept by using the identified quarter as a case study.

The planning site of around six square kilometers is located in the center of Hefei. It represents the old town Hefei how it was 60 years ago. The closure of the railway site and its subsequent revitalization will increase the favorability of the large areas surrounding the site. Two BRT lines traverse the area from north to south and east to west. Two additional BRT lines start in this area. The planning site is connected to Hefei's private transport network by a super ordinate and several subordinate urban streets. In addition, a subway station is planned on the eastern edge of the site from which Hefei main train station can be reached in a small amount of time. In the future the area is to be expanded to evolve into a new urban quarter predominantly for residential use, shopping as well as the necessary supply facilities (mixed-use). The area should become the deserved (and essential) face of Hefei.

Together with the Transportation Agency and the Urban Planning Agency of Hefei, an idea was developed for this site to devise a concept for the connection of the private and public transport, which was distinguishing and simultaneously able to be rapidly implemented. The goal within the "walk-able city" was to offer an expansive modularly-graded, densely synchronized public transport option (BRT, subway), which significantly and measurably reduces private car traffic with the goal of resource conservation. The idea of the "walk-able city" is to be appropriately devised, within the scope of a transport and urban planning case study, so that it can be translated onto other comparable situations and cities.

The sustainable quarter in central Hefei will serve as an example to assess the applicability of the sustainable development concept and its environmental effects in other parts of Hefei. This can be done for smaller parts of Hefei or based on a classification of Hefei (e.g. areas with high/low potential for a sustainable restructuration/transformation). These variations of the BAU will be analyzed and described within the GIS in order to discuss them with local partners and stakeholders. The main objective of this discussion is the development of recommendations for future planning activities in Hefei.

Traffic Management. In the first step a traffic management concept was developed in cooperation with all essential administrative bodies of the city government like the city planning bureau, the transportation bureau, and the traffic police. The main core of this concept is an open platform for traffic data exchange, were all systems for traffic data collection as well as all transport oriented public (and in future also private) services connected with. This approach gives the possibility to establish standards for traffic parameters and traffic information. This makes the connection of existing modules easy, the implementation of new systems possible, and increases the quality of the data base and therefore the reliability of the information.

The main systems for traffic data collection related to the road traffic are systems based on floating car data were vehicle probes send their positions to the central traffic data base. For travel time measurements up to 10.000 vehicles will be monitored in the final system. Currently a fleet of private taxis, a fleet of police cars and a fleet of dump trucks are connected with the open platform. The data of all these vehicles fused to a description of the current level of service of the traffic flow on each street of the city. Figure two shows a screenshot of the software application for displaying the current level of service of the travel speed. In fife different classes the travel speed is shown (free flow, almost free flow, medium flow, congested, and no data). This software was developed especially for operators in traffic management centers and is installed at the center of the traffic police of Hefei. This application is an enhancement of a module which was developed for the traffic management center of the Ministry of Interior of Baden Württemberg (Germany) during the world soccer championship 2006 (Bonert 2006).



Figure 1. City Center of Hefei with planned public transport 2015 (source: Hefei City Planning Bureau).



Figure 2. Main Screen of the software TrafficViewer (source: LUAX).

For broadcasting the traffic information to travelers and car drivers a communication system in the standard of digital multimedia broadcast (DMB) is in use for the METRASYS project. The station of Radio Anhui operates the system; the special content for travelers and drivers is created by the traffic police in cooperation with ASEC and METRASYS. The information are coded and transmitted by the standard of Traffic Protocol Expert Grub (TPEG). The level of service based on the floating car data is transmitted automatically. Beyond this there is a special software application running on the server of the police traffic management center for creating special traffic information. In the case of an incident or a major accident, the police are able to create interactively special messages in the standard of TPEG. Of course the development of an application to interpret these messages for dynamic navigation onboard of cars is also part of the project METRASYS. A first release is already tested and available.

Using all these modules ASEC was developing a software system for the management of the private taxi fleet. This software shows the current position of each taxi and in the case of occupied taxis the number of passengers. Beyond this the operator can contact the taxi driver by this application. Initially, the primary purpose of this system is to increase the safety of the drivers. The driver can emit an emergency signal, and it is also possible for the operator in the management center to listen into the taxi by mobile communication. Statistical analyses of the taxi fleet can be done with this software too. For example it is possible to show that up to 50 % of all trips are without passengers: the driver is searching for passengers. Additional analyzes regarding the hours of work, operating costs, and income is made possible.

This system gives also the possibility of a taxi call center. This call center will be realized in the future by the taxi company. Such a call center helps to reduce the energy consumption of the transport system, because the taxi driver will get passengers without searching by driving around. Figure 3 shows a screenshot of the software.

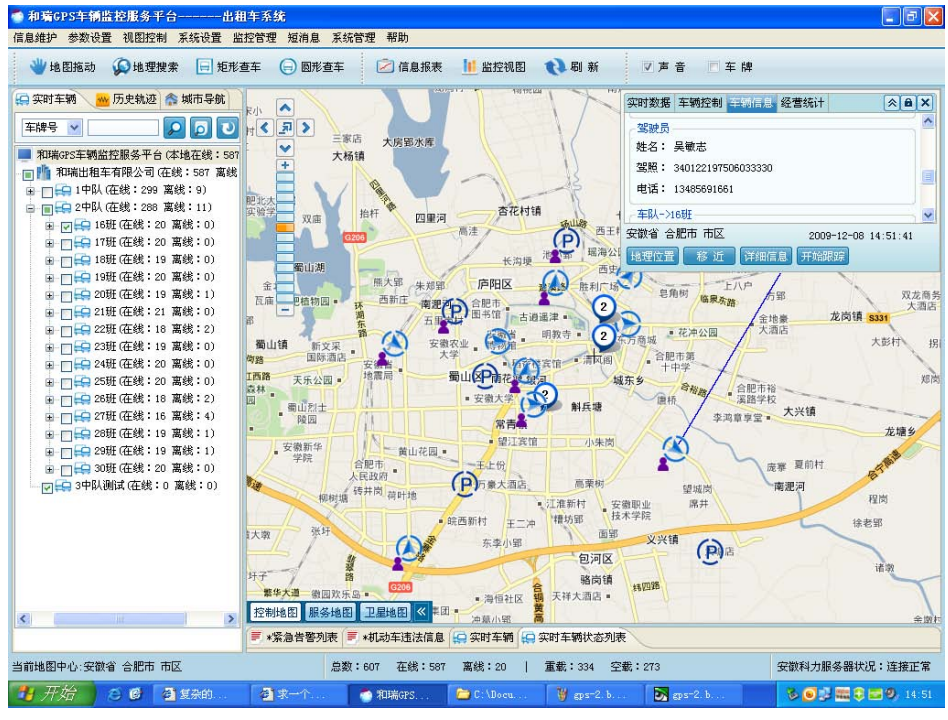


Figure 3. Software for Taxi Dispatching and safety; source: ASEC

These are the technical results of the first eight moth of the METRASYS project. The development of the technical system for traffic management is installed in Hefei.

Evaluation and Immission Monitoring. The METRASYS project provides in a first step an overview of the principles how to evaluate the environmental impact of local transport policies and measures. It is intended as an introduction to the different aspects of environmental evaluation that needs to be taken into account in the METRASYS project. Details of the evaluation techniques will be further elaborated in the respective work packages of the ongoing project.

The METRASYS project is contributing to the sustainable and climate-friendly development of fast growing mega cities. By a transit oriented sustainable city planning and by implementing several approaches of traffic information and traffic management, local transport shall be organized in the most efficient way. This will most likely reduce fuel consumption and generate other environmental benefits such as the reduction of local pollutants and noise. Furthermore, the traffic management in Hefei shall provide the Chinese project partners with comprehensive information about the current traffic situation of the city.

Immission Model. The modeling of immissions of the city of Hefei is a complex combination of measurements, factors and mathematical equations. At the end the immissions for the region of interest will be shown as a dynamic map. In relation to the daytime, the weather, the traffic and the measured immissions the map is dynamically chancing.

For the modeling there are several input data needed. The parameters can be sorted into the following classes:

- traffic data,
- measured immissions,
- measured weather data,
- three dimensional terrain data, and
- land use data.

In addition to that so called emission factors of the common cars are needed and several sub models have to be integrated. In because of the high complexity of the whole system in the end the model needs as a continuous input measured data of the current weather situation, the current immissions measured at several measurements sides spread over the city and the current traffic volume on the streets. To make the system more reliable it should be possible to integrate also data of the main (non traffic related) emitters of the city. Like big energy plants, steel plants or concrete plants.

Traffic Data. In the city of Hefei are exist several fleets of floating car data. The output of such system is the current travel speed on each street segment. With some simple mathematical equations the traffic flow (number of cars in a period of time) will be calculated on the base of travel speeds. If available there should be also a measurement of traffic flow by inductive loops at some single points in the city. The inductive loops give the opportunity to separate the cars into several grubs like trucks, busses or passenger cars. This separation can be done once in the beginning of the modeling process. It is not necessary to do this online and all the time.

The first parameters of the model are the number of cars (separated into several grubs of cars) in a fixed time sequence for all street segments in the region of interest.

Emission Factors. The emission factors are averages of gas emission of cars. The emission factors are developed by measuring the emissions of a few numbers of typical cars, by driving typically. These measurements are already done by XXXX inside the project Quantify. In because of the variation of the fleet of cars over several years, the emission factors should be updated all fife years. Bringing together the numbers of cars with the emission factors gives the emission of the street traffic related to each street segment and to the current time.

Immission Measurements. The transmission, the way from emission to immission, is very complex and strong influenced by the weather for example. Therefore it is also necessary to measure immissions at several measurement sides spread out over the whole region of interest. The internationally common monitored parameters are CO, NO_X, SO_2, O_3 and dust (PM 10). The first four parameters are also the output of the whole model. For the modeling process the parameters should be available online in fixed time sequences and units (like ppm or g). The locations of the measurement sides are needed too. It is sufficient to get this information once in the beginning of the modeling process sown in a city map or similar.

Meteorological Data. The transmission is very strong related to the weather situation. Rain, wind, sun radiation, temperature or air pressure can change the transmission process completely. Therefore it is very important to have current meteorological data.

From different measurement sides continuously needed:

- Precipitation (type and sum),
- Wind speed and direction,
- Sun radiation,
- Air temperature,
- Ground temperature.

These parameters should be available online and without or very small time delay.
From one measurement side and once a day is needed:

- Lake water temperature (it is needed, because Hefei has a lot of lakes),
- The vertical profile of air temperature and air pressure.

Geo Data. Because of the high influence of the temperature and the wind field land use data are needed. These data are necessary once in the beginning by the development of the model. In because of the fast development of Hefei these data should be updated all 6 month.

Evaluation. By providing this information, the traffic management can contribute to the evaluation of environmental impacts of local transport policies and measures.

Firstly, the traffic management system to be installed in Hefei can support the ex ante evaluation of environmental impacts, because past and current traffic development in Hefei can be used to outline the future development. Policy makers can model the potential impact of transport policies and measures and thus use traffic information to inform which transport interventions may become necessary in the near future.

Secondly, the traffic management system can support the ex post evaluation of policies and measures, i.e. policy makers would use the data derived from the traffic management to assess the actual success of (sustainable) transport interventions.

In addition to providing a general overview of how environmental impacts of transport can be evaluated, the METRASYS project focuses on the evaluation of green house gas (GHG) impacts, as this is an important matter of concern for policy makers. In a first step, the concept of sustainable transport on the urban level is explained. This includes a description of environmental impacts that are caused by transport as well as policies and measures that local decision makers can implement in order to facilitate an urban sustainable transport system.

The METRASYS project will in a second step highlight to which extent the traffic management system provides information about the environmental impacts of local traffic in Hefei. In a third step, it has to be identified the key challenges and

principles for evaluating the GHG impacts of local sustainable transport policies and measures.

Conclusions and Outlook

The METRASYS project deals with sustainable mobility for mega cities and runs since ten moths. The city were all systems are used is Hefei the capital of the Anhui province in China. The first systems are already implemented and working. In the beginning of the project, the focus was on the development of the traffic management system. Three different fleets of floating car data are equipped with GPS receivers and communication devices and linked to the main data platform of the METRASYS traffic management system. Based on this data monitoring applications and information platforms were developed and are available now.

In addition to that the work exploring the topic of financing traffic and carbon trade did start also. A lot of analyses were done as well as interviews with stakeholders and official representatives of the local government. As a linking topic in between the team of Metrasys project will do transport oriented city planning for a special area in the city. The results of these tasks will be published in additional articles focusing on these special topics.

Acknowledgment

We owe thanks to all Chinese and German partners of METRASYS as they are: Freie Universität Berlin Arbeitsbereich Stadtforschung (TEAS); Wuppertal Institute for Climate, Environment and, Energy; AS&P – Albert Speer & Partner; Fraunhofer Institute for Computer Architecture und Software Techniques FIRST; Luax Software Engineering; Research Centre for Software Engineering Technology Anhui Province (ASEC); Chinese Academy for Transportation Science, Center for Sustainable Transportation; Tongji University, School of Transportation Engineering; Chinese Academy of Sciences, Anhui Institute of Optics and Fine mechanics; Hefei Municipal Traffic Police Department; Hefei Communication Bureau; Hefei Environmental Monitoring Center; Hefei Institute of Urban Planning and Design and last but not least the Government of Jiading District, Shanghai. Special thanks to Malte Beckmann and Liu Wang of the Wuppertal Institute: the chapter “Evaluation and Immission Monitoring” is mainly based on their work.

References

Bonert, M., Brockfeld, E., Ernst, I., Krajzewicz, D., Ruhé, M. Wagner, P. (2006) “SOCCER Verkehrslageerfassung und –prognose während der Fußball-WM”. *IMA 2006 Informationssysteme für mobile Anwendungen* , Wolfsburg, Germany.

Dalkmann, H. and Brannigan, C. (2007). “Transport and Climate Change”. A Sourcebook for Policy-Makers in Developing Cities: Module 5e. Gesellschaft für Technische Zusammenarbeit – GTZ Eschborn.

Fang, Tingjian.(2007). "Flexible traffic signal control for bus priority", *In Proceedings 14th ITS World Congress*, 2007, Beijing.

Fang, Tingjian.(2007). "Study of modeling diver-vehicle agent and its communication mechanism in microscopic traffic simulation", *In Proceedings 14th ITS World Congress*, 2007, Beijing.

Fang, Tingjian.(2008). "An Automatic Traffic Congestion Detection Method Based on Floating Car Data" . *15th ITS World Congress*, New York City.

Fang, Tingjian.(2009). "An Evaluation Method for Floating Car Data (FCD) System". *In Proceedings 16th ITS World Congress*, Stockholm.

Kühne, R., Bei, X., Ruhé, M. (2009) "Sustainable Mobility & Sustainable Transport for Medium-Sized Cities in China". *International Conference: Megacities: Risk, Vulnerability and Sustainable Development* , 07.-10, Leipzig.

Kühne, R., Ruhe, M., Bei, X., Böhnke, P., Gou, L. (2008) *Traffic Management in Emerging Mega Cities using Vehicle Probes for Data Collection.* 4. International Symposium "Networks for Mobility", Stuttgart.

Kern, K. (1997). "Ansätze einer innovativen und integrativen Mobilitätspolitik in den USA". In: Metz, Lutz und Martin Jänicke (ed.): Sektorale Umweltpolitik – Analysen im Industrieländervergleich, edition sigma, Berlin, 133-164.

Gwilliam, K., Kojima, M., Johnson, T. (2004). "Reducing Air Pollution from Urban Transport". The World Bank: Washington, D.C.

Moore, T., Sanchez, T. (2001). A Guidebook For Evaluating the Indirect Land Use and Growth Impacts of Highway Improvements: Final Report. SPR Project 327. Oregon Department of Transportation: Salem and Federal Highway Administration: Washington, D.C.

OECD - Organisation for Economic Co-operation and Development (2002). Report on the OECD Conference Environmentally Sustainable Transport (EST): Futures, Strategies and Best Practice. Vienna, Austria, 4-6 October 2000.

Schäfer, A. (1998): The global demand for motorized mobility. In: Transportation Research A, Volume 32, Number 6, 455-477.

Ruhe, M. (2008) *Innovationen im Verkehrsmanagement für chinesische Großstädte.* ChinaContact , Seite 42.

A Sustainable Development Based Land Use Assessment for Freeway Construction

Yan Liu[1], Jianfeng Liu[2] and Kerong Deng[1]

[1] Professor, School of Management, Chongqing Jiaotong University, 66#, Xuefu Ave., Nan'an, Chongqing, China, 400074, Email:liuyx@cquc.edu.cn
[2] Engineer, Ba'nan Highway Engineering Company, Chongqing, China, 401320

Abstract

Land use evaluation is the basis for sustainable development of freeway construction. Based on the investigation of the sustainable use of land resources, the inherent nature of sustainability land use of freeway construction was analyzed. An assessment indicator system about freeway land use was established according to the freeway functional requirements, site characteristics and present land use problems in freeway construction. Through a comprehensive consideration of coordinated development in society, economy, resource, and environment, an exponential growth curve land use evaluation model was established which was based on a method named same indefinite-contrary order of set pair analysis (SPA). The model was applied in freeway land use evaluation of China. The results showed that the evaluation model had a convenient and practical feature in calculation, strongly contrasting, and also has a better objective rationality.
Key Words: freeway; sustainable development; construction land use; set pair analysis; same indefinite-contrary order; exponential growth curve model

Introduction

The formal concept of sustainable use of land resources has been brought forward for only about 20 years. Domestic experts have been conducting research on how to economize cultivated land by reducing the occupied land area of freeway construction projects, and have made certain achievements (Yin et al. 2004; Deng 2007; Chen 2008; Luo 2008; Zhao 2008; PRC 1999). During the overseas freeway construction, the coordination between the freeway and surrounding environment is given more attention, with visual harmony highlighted. Based on the diversity of national conditions, overseas researches on freeways and land along them take into

account how the land along freeways will actualize the freeway's service function, neglecting the problem of overusing land resources (AASHTO 1991; Boarnet 1997; Sanchez 2004; Froment and Domon 2006; Beckerman 2008).

Evaluating the sustainability of land use for freeway construction is a dispensable and significant procedure in putting theories from researches on sustainable use of land into practice. The construction of scientific and applicable indicator system for the sustainability evaluation of land use for freeway construction, as well as the detection of its sustainable level in appropriate ways, serves as insurance for the evaluation results being practical and instructional.

Essence of the Sustainability of Land Use for Highway Construction

In addition to considering how to fulfil the need for social and economic development for highway construction land use, sustainability of land use for highway construction, as a brand new outlook on thinking and development, should also try to reduce unfavourable influences on the social and natural environment and coordinate the highway construction with society, economy, resources, and environment. Four aspects are covered in it, namely, sustainable development of economy, sustainable development of society, sustainable development of resources and environment, and sustainable development of highway in itself. The final purpose is to ensure the permanent develop capabilities and lasting state of development so as to meet needs for the development of national economy and promote general progress of society. Therefore, the sustainability of land for highway construction embodies the following points：

Financial Sustainability of Land Use for Highway Construction. The supply capacity of land for highway construction should balance the need of social and economic development for highway construction, which means the sustainable development of highway construction should check with the sustainable development of society and economy.

Resource Sustainability of Land Use for Highway Construction. Sustainability of land for highway construction should take protecting natural resources and ecological environment as a prerequisite, and actualize the sustainable development of highway construction and resources in accordance with the bearing capacity of resources and environment. To achieve that, we must optimize limited wealth of time and space, give priority to cutting down the land resources consumption of highway construction, and improving the overall efficiency of highway traffic and construction land through land saving, intensive utilization, and scientific layout of highway net, thus maximizing the utility value of resources.

Social Sustainability of Land Use for Highway Construction. Profits brought by highway improvement and development should be shared equally among all members in society, with interests of the vulnerable group covered (He 2006)

Table 1. Sustainability land use evaluation indicator system for freeway development.

Indicator	First-class indicator	Secondary indicator	Unit
Sustainability evaluation indicator of freeway construction land use	Economic development indicator	Annual growth rate of GDP	%
		Annual growth rate of industrial added value	%
		Annual growth rate of agricultural added value	%
		Annual growth rate of added value of the tertiary industry	%
		Annual growth rate of fixed assets investment	%
		Annual growth rate of highway construction investment	%
		TPPR / Overall Labor Productivity	RMB Yuan/person
		Per capita disposable income of urban residents	RMB Yuan/person
		Net income of peasants	RMB Yuan/person
		Annual growth rate of highway passenger transportation turnover volume	%
		Annual growth rate of highway freight transportation turnover volume	%
	Social progress indicator	quantity of employment	(10 thousand person)
		urban registered unemployment rate	%
		Quantity of people covered by urban basic endowment insurance	10^8person
		Proportion of education investment in GDP	%
		Length of the freeway per 10 thousand people	km/10 thousand person
		Length of the freeway per hundred square kilometers	km/100 square kilometers
		Per Capita Floor Space of Rural Residents	m^2/person
		Per Capita Floor Space of Urban Residents	m^2/person
	Resource and environment indicator	Annual new rehabilitated soil-water eroded area	million ha/year
		Annual growth rate of cultivated land preservation	%
		Per capita agricultural acreage	mu / person
		land use for freeway & environmental policy and its enforcement	-

Sustainability of Land Use for Highway Construction Also Can Serve for Interpreting Sustainability of Land Use for Freeway Construction. In addition to considering the sustainable development of society and economy, more weight should be attached to thrifty utilization of land resources and environmental protection due to the high standard of freeway alignment, more traffic lanes, and considerable need for land resources.

Sustainability Evaluation of Land Use for Freeway Construction

With the spirit of giving full expression to the essence of sustainability of land use for highway construction and convenience and possibility of data acquisition considered, the sustainability evaluation indicator system of land use for highway construction can be built in three aspects, namely, resources, society, and environment (Smyth and Dumanski 1993; Ahmed and Hossain 1990). Thereinto, resource and environment indicators reflect the exploration and utilization of land resources and the influences on the environment based on natural conditions of land resources; society indicators mirror largely the benefits and influences of freeway construction on society, including freeway construction achievements, freeway service capacities, social benefits of freeways, and so on; economic indicator images investment and outcome of freeway construction and how they take effect in promoting socio-economy growth (Kahandker 1989; Kwon 2001; Kahandker et al. 1994). Specific indicators are shown in Table 1. Thereinto, "land use for freeway & environmental policy and its enforcement" is a qualitative indicator, reflecting the conditions of freeway land use and policy formulation as well as its implementation.

Evaluation Model for Freeway Construction Land Use Based on Set Pair Analysis Theory

Brief Description of Set Pair Analysis Theory. Combining deterministic parsing and uncertainty analysis, set pair analysis is a new systemic analysis method which is applied to describe and solve problems concerning integration. The so-called set pair is actually a pair composed of two sets with certain associations. The core idea of set pair analysis is applying a deterministic system and uncertain system to analyze the connection and transformation of objects from the same, different, and reverse aspects. Hence, set pair analysis refers to make both deterministic and uncertain analysis (i.e., the same, different, and reverse analysis) of the features of the two given sets in the context of certain problems, and set up the same, different, and reverse linkage expressions of the two sets in the context of the given problems:

$$\mu = a + bi + cj = \frac{S}{N} + \frac{F}{N} i + \frac{P}{N} j \tag{1}$$

In the formula, a, b, and c fulfill the normalizing condition

$$a + b + c = 1 \tag{2}$$

N refers to the characteristic total of sets; S refers to the mutual characteristic numbers of the two sets in set pair; P refers to the opposite characteristic numbers of the two sets in set pair; $F=N-S-P$ refers to characteristic numbers neither mutual nor opposite of the two sets in the set pair. Therefore, *a, b,* and *c* is, respectively, called

identity degree, divergence degree, and opposition degree of the two sets in the context of the same problem. They depict the related conditions of the two sets from diverse sides. The i in Eq. 1 refers to the coefficient of divergence uncertainty degree, taking values in the range of [-1, 1] according to various situations; j refers to the coefficient of opposition degree, always taking the value j=-1 in operations. And as a general rule, connection degree μ equals to the right expression, and only under some special conditions does it refer to a numerical value with a replaced name of connection number. The system status ranking according to size relations of *a, b,* and *c* in the connection degree Eq. 1) turns out to be identical, different, and reverse situations ranking. If every kind of ranking is signified by a corresponding real-valued number in the range of [0.1, 1], then each number will be regarded as a situation degree. The identical, different, and reverse situations ranking based on situation degree in Table 2.

Table 2. Hierarchy and order of set pair.

Set pair situation	Size comparison between a, b, and c		Posture degree	Set pair situation	Size comparison between a, b, and c		Posture degree
Same potential	$a>c$	$b=0$	1	Anti-potential	$a<c$	$b>c$	0.3
		$c>b>0$	0.9			$c>b>a$	0.3
		$a>b>c$	0.8			$c>b>0$	0.2
		$b>a$	0.7			$b=0$	0.1
Balance potential	$a=c$	$b=0$	0.6	Uncertain, same potential	$c=0$	$a>b$	0
		$a>b>0$	0.5				
		$b=a$	0.4	Uncertain, uncertain potential	$c=0$	$a\leq b$	
		$b>a$	0.4				

Highway Construction Land Evaluation Model for Sustainability

General principles for determine evaluation criteria. In addition to environmental quality indicators, there is no satisfactory evaluation criteria unanimously acknowledged for the current domestic and international society, economy and environment indicators. In order to make evaluation indicators comparable both lengthwise and horizontally, the conformation of evaluation criteria should follow the principles below[19]: for indicators like "percentage" and "unit mean," set down a three-grade standard suitable for different periods, which comprises grade I (qualified) , grade II (moderate), and grade III (good); For other indicators, set down another three-grade standard of grade I (present value), grade II (short-term target value), and grade III (long-term target value) in combination with development programs.

Situation degree of social and economic development and situation degree of ecological environment dysfunction. Set pair analysis can be applied to decide whether those indicators inside social, economic, and environmental systems

coordinate with each other. Take the indicator systems of society, economy, resources, and environment of freeway construction land and their corresponding evaluation criteria as two separate sets to form a set pair. Suppose N evaluation indicators are selected as the characteristic total of the set pair. As to society and economy indicator systems, take grade *i* (*i*=1,2,3 namely, grade I, grade II, grade III) as the evaluation criterion with evaluation criteria worked out. Then in the *N* indicators, if the number of indicators fit in with grade I, grade II, and grade III criterion is represented, respectively, by N_1, N_2, and N_3 (note: $N_1 + N_2 + N_3 = N$) identity degree a, divergence degree b, and opposition degree c in the connection degree expression of social and economic development are shown in Table 3 corresponding to various criterion of different grade. For environment indicator system, a, b, c in the association expression of ecological environment functions take their value, respectively, as $a = N_3 / N, b = N_2 / N, c = N_1 / N$.

Table 3. Calculation of a, b, and c.

Criteria level	*a*	*b*	*c*
I	N_1 / N	$(N_2 + N_3) / N$	0
II	N_2 / N	N_3 / N	N_1 / N
III	N_3 / N	0	$(N_1 + N_2) / N$

According to size relations of *a, b,* and *c* in the connection degree expressions of society, economy, resource, and environment indicators of freeway construction land, the situation degrees ranked on the basis of the identical, different, and reverse situations in the table are named as situation degree of social development, economic development, and environment dysfunction d_s, d_e and k (Li 2007).

Indicator model for coordinated development of society, economy, and environment. Social and economic development and environmental changes all consist of three phases of birth, development, and maturity. Their mutual relationships can be depicted by growth curve. Thus coordinated development indicators of society, economy, and environment can be defined as (Li et al. 2000):

$$I = \frac{1}{1 + ke^{-d_s \cdot d_e}} \tag{3}$$

In the expression, d_s, d_e and k represents, respectively, the situation degree of social development, economic development, and ecological environment dysfunction under the given criterion. The minimum value of d_s d_e and k is 0.1, and the maximum value 1. With analysis of the value of development indicator *I* when k, d_s, d_e taking certain values, corresponding relations between the value range of *I* and coordinated development as Table 4.

Table 4. Range of coordination development indicator I and the criteria for coordination evaluation.

I	1.00-0.90	0.90-0.75	0.75-0.60	0.60-0.50
Result	good	fair	weak	poor

Table 5. Land use criteria for freeway highway construction and the statistic for 2004-2007.

Indicator		Criteria			China's index value for 2004～2007			
		I	II	III	2004	2005	2006	2007
Economic development indicator	Annual growth rate of GDP (%)	7.00	7.25	7.50	9.50	9.90	10.70	11.40
	Annual growth rate of industrial added value (%)	10.00	12.50	15.00	11.50	11.40	12.50	13.50
	Annual growth rate of agricultural added value (%)	5.00	5.05	5.10	6.30	5.20	5.00	3.70
	Annual growth rate of added value of the tertiary industry (%)	9.00	12.25	15.50	8.30	9.60	10.30	11.40
	Annual growth rate of fixed assets investment (%)	10.00	14.47	18.93	25.80	25.70	24.00	24.80
	Annual growth rate of highway construction investment (%)	9.00	12.00	15.00	26.60	16.60	13.60	4.20
	TPPR / Overall Labor Productivity(RMB Yuan/person)	18824	24534.5	30245	15381	18824	21725	26127
	Pper capita disposable income of urban residents(RMB Yuan/person)	10493	11941.5	13390	9422	10493	11759	13786
	Net income of peasants(RMB Yuan/person)	3255	3702.5	4150	2936	3255	3587	4140
	Annual growth rate of highway passenger transportation turnover volume(%)	8.3	9.2	10.1	13.9	6.3	9.1	13.00
	Annual growth rate of highway freight transportation turnover volume(%)	4.3	5.5	6.7	7.00	9.30	11.00	15.40
Social progress indicator	Quantity of employment(10 thousand person)	80000	85000	90000	75200	75825	76400	76990
	Urban registered unemployment rate(%)	5	4.25	3.5	4.2	4.2	4.1	4
	Quantity of people covered by urban basic endowment insurance(10^8person)	17488	19894	22300	16353	17488	18766	20137
	Proportion of education investment in GDP (%)	0.78	0.97	1.15	1.33	1.08	1.01	0.9
	Length of the freeway per 10 thousand people(km/10 thousand person)	0.19	0.30	0.4	0.26	0.31	0.34	0.41
	Length of the freeway per hundred square kilometers(km/100 square kilometers)	0.26	0.42	0.57	0.36	0.43	0.47	0.56
	Per Capita Floor Space of Rural Residents (m^2/person)	25	30.11	35.21	27.9	29.7	30.7	31.6
	Per Capita Floor Space of Urban Residents (m^2/person)	22	26.52	31.03	25	26.1	27.1	28
Resource and environment indicator	Annual new rehabilitated soil-water eroded area (million ha/year)	500	390.5	280.9	659	265.4	283.7	290.2
	Annual growth rate of cultivated land preservation(%)	-1	-0.65	-0.3	-0.77	-0.31	-0.22	-0.05
	Per capita agricultural acreage(mu/person)	1.4	1.36	1.32	1.41	1.40	1.39	1.38
	Land use for freeway & environmental policy and its enforcement	6	7.5	9	6.00	7.00	8.00	8.5

Sustainable and Coordinated Development Evaluation of China's Freeway Construction Land

Determination of Indicator Value and Evaluation Criteria. Investigate the situations of freeway construction land during 2004～2007 based on China's freeway construction. Relevant statistics materials provide indicator values from 2004 to 2007 in Table 5. The indicator of "land use for freeway & environmental policy and its enforcement" is determined by experts on the basis of the conditions of sensible land use of freeway construction, as well as formulation and implementation of environmental policies.

In this paper, the Eleventh Five-Year Development Plan for highway and waterway transportation, Eleventh Five-Year Plan Scientific Development Plan for highway and waterway transportation, Eleventh Five-Year Plan for national rural economic and social development, and the planning value of national land planning outline from 1997 to 2010 are taken as Grade III criterion, the mean lowest value of various provinces as Grade I criterion, and the mean value of Grade I and Grade III as Grade II criterion. This is shown in Table 5. The resource is National economic and social development statistical communique, the development of highway and waterway transportation industry statistics bulletin, China agricultural development report, China statistical yearbook.

Table 6. *d*s, *d*e, *K*, *I* and the evaluation results in different periods of Freeway highway development.

Year	*d*s	*d*e	*K*	*I*	Result
2004	0.8	0.7	0.5	0.7778	Fair
2005	0.6	0.7	0.4	0.7919	Fair
2006	0.7	0.7	0.4	0.8031	Fair
2007	0.8	0.7	0.4	0.8140	Fair

As seen in Table 6, social, economic, and environmental coordinated development indicators were all greater than 0.75 during 2004 to 2007, and the indicator value of coordinated development increased year by year. The criterion in Table 4 shows that China s freeway construction has embarked on the coordinated way in respect to society, economy, resource, and environment. The Chinese government attaches great importance to land resource preservation while carrying on economic construction all these years, developing and implementing a series of land polices and measures, which have been taking progressive effect.

Result. Based on the above connection degree expression and ranking values of identical, different, and reverse situations, values of d_s, d_e, and K can be obtained, and coordinated development indicator I of China's freeway construction land can be calculated by expression (6) as shown in Table 6.

Conclusions

Setting down an indicator system with the purpose of freeway construction land being sustainable, this paper created a freeway construction land evaluation model for the coordination of society, economy, resources, and environment, and assessed China's freeway construction land. The results indicated that it is quite concise and pellucid to assess the sustainability of freeway construction land by set pair analysis, in which the determination of weight in methods like gray association analysis, analytic hierarchy process, and fuzzy comprehensive evaluation can be avoided, thus producing more objective and reasonable evaluation. Meanwhile, this model is simple to calculate with an intuitive evaluation method, high comparability, and no restriction of the indicator number of the system.

References

American Association of State Highway and Transportation (AASHTO). (1991). "A Guide for Landscape and Environmental Design", *Washington, D.C.*

Ahmed, R., and M. Hossain. (1990). "Developmental impact of rural infrastructure in Bangladesh", *IFPRI Research Report No.83.* Washington, D.C.: International Food Policy Research Institute.

A. J. Smyth., J. Dumanski. (1993). "An international framework for evaluating sustainable land management", *FESLM.*

Bernardo, Beckerman. (2008). "Correlation of nitrogen dioxide with other traffic pollutants near a major expressway", *Atmospheric Environment,42(2)*.

Chen, C. N. (2008). "A scan of freeway economical use of land", *China land,* Beijing, 32(8).

Deng, X. (2007), Reflections on the building of the expressway, *Hunan Communication Science and Technology*, Changsha, 33(4), 183-184.

He, G. G. (2006). "A discussion of sustainable land use of collective infrastructure construction from the system theory point", *Land and Resources Science and Technology Management*, 23(3), 50-53.

Jos é , Froment., G é rald, Domon. (2006). "Viewer appreciation of highway landscapes: The contribution of ecologically managed embankments in Quebec." *Landscape and Urban Planning*, Canada

Kahandker, S. (1989). "Improving rural Wages in India, Policy, Planning, and Research Working." *Working paper 276*. Washington, D.C.: World Bank.

Khandker, S., V. Lavy., and D. Filmer. (1994). "Schooling and Cognitive Achievements of Children in Morocco: Can the Government Improve Outcomes?" *World Bank Discussion Papers*, No.264. Washington, D. C.: World Bank.

Kwon, E. (2001). "Infrastructure, growth, and poverty reduction in Indonesia: A cross-sectional analysis", *Asian Development Bank*.

Luo, Y. G. (2008). "Manage and temporary use of land highway." *China land, Beijing*,8 , 30-31.

Marlon G. Boarnet. (1997). "Highways and Economic Productivity: Interpreting Recent Evidence." *Journal of Planning Literature*,11(14).

Ministry of construction of the peoples republic of China, Ministry of Land and Resources of the peoples republic of China. (1999). *Land indicators for highway construction projects*, Beijing.

Thomas W. Sanchez. (2004). "Land Use and Growth Impacts from Highway Capacity Increases." *American Society of Civil Engineers(ASCE).*

Yin, L. L., Zhang, Q., and Chen, G. X. (2004). "A discussion about construction land use type of freeway." *Guangdong Highway Communications*, Guangzhou, 82(1), 48-51.

Zhao, K. Q. (2000). "Set pair analysis and its application." *Zhejiang Science and Technology Press,* Hangzhou.

Zhao, X. S. (2009). "Research on highway construction projects to save land." *A dissertation for Master Degree*,China University of Geosciences, Beijing.

Zuoyong Li, Shilun Shen, Xinmin Deng(2000), Assessment Model of Community, Economy and Environment Coordinative Development Index, *Shanghai Environmental Sciences*, Shanghai 19(5), 201-204.

Zuoyong Li(2007). "Theory and application of sustainable development evaluation model." *Science Press*, Beijing.

The Future of Interdisciplinary Policy Research in Africa
A Case Study of the Transport Sector

Fred Amonya[1]

[1] Centre for Transport Studies, Imperial College London, London SW7 2AZ, United Kingdom; PH. +44 (0) 20 7594 6043, Email: f.amonya@imperial.ac.uk; Lead Consultant (Infrastructure Policy), Lyciar Ltd, www.lyciar.com, Email: Fred.Amonya@lyciar.com

Abstract

The interaction between transport and social welfare is one of the most complex problems in Africa. It is a challenge of intertwined disciplinary issues of science – both social and natural sciences. It begs interdisciplinarity. This paper explores the future of interdisciplinarity in research on the continent through two lenses. First, through a rendering of drivers and agencies of research on the continent. Secondly, by presenting an analysis of the process of policymaking in the transport sector. Development of research in 16 countries of East, Central, and West Africa is traced back to pre-independence. An emerging model of research on the continent is presented. The model identifies weak linkages between private firms and public research institutions as the hindrance to the growth of interdisciplinarity on the continent. As regards transport policymaking, six regional policy initiatives were considered. This strand of the enquiry identifies the absence of local research entities in the policy process as the main obstacle to interdisciplinarity; the local research entities would in turn require strong private sector participation for vibrant interdisciplinary research. The paper concludes that interdisciplinarity will come to bear. It could be through a natural process of evolution, or through discrete action by donors, local political actors and universities.

Background

Science and research are two words that feature prominently in this paper. The context in which these two words are used must therefore be made clear at the onset. The meaning of the term science has attracted a lot of debate over the ages. This paper does not explore this debate. Instead, it presents a definition that facilitates a

wider discussion on the history and future of research. Science is in this context is knowledge grounded in evidence and acquired through logical reasoning, and hence reliable as a predictive tool or premise. A systematic effort at broadening knowledge is termed research. Broad and complex problems demand an interdisciplinary research. A definition of 'interdisciplinarity' presented by Repko (2008) is adopted:

Interdisciplinarity is a process of answering a question, solving a problem, or addressing a topic that is too broad to be dealt with adequately by a single discipline or profession. *It* draws on *various* disciplinary perspectives and integrates their insights through *the* construction of a more comprehensive perspective [italics added].

The paper makes a limited defence of this statement. For a broader and deeper defence, Repko's work is a good reference. This statement of definitions sets the stage to explore the use of research as a guide in policymaking, with a focus on transport policy in Africa – particularly, Sub-Saharan Africa (SSA).

The paper is guided by questions in Box 1. The paper primarily questions the nature and direction of research within the continent. It achieves this goal by tracing changes in the structures (drivers and agencies) of research in the region. It then explores future interactions between research and policymaking in Africa. Textual analysis is used as the investigatory tool, yielding empirical results presented as frames. Note that the paper adopts an integrated approach, not a discrete presentation of the problem, methodology, results and analysis.

The paper is organised as shown in Figure 1, which is consistent with guiding questions in Box 1. Section 2 discusses welfare challenges of SSA, identifying how these challenges relate to transport in the region. The transport-welfare challenge is assessed as non-linear and complex, and one that begs interdisciplinarity. Not only are welfare problems complex, the regional organisations that attend to these challenges are prone to organisational tension. This is illustrated.

The paper then focuses on science (through research) as a tool for understanding the evolving challenges of the region, and for diagnosing solutions to these challenges. Section 3 traces the path of research in Africa. This is done against a background of the evolution of the structures of research in the West. This section identifies obstacles to interdisciplinarity through the lens of research structures. Section 4 examines the uptake of research products [science] in policymaking. It uses the institutional rational choice framework (IRC), as a lens for examining the uptake of science. Section 5 ponders the future of research on the continent. The paper concludes in Section 6.

Transport and Social Welfare in Sub-Saharan Africa

Transport infrastructure binds humanity. Sub-Saharan Africa is home to 800 million people, dispersed over 24, 240 square kilometres (World Bank, 2009). These data are

Box 1: Guiding questions.

Issues of welfare in Africa are diverse and complex. Transport is a key strand in this complex mesh. The challenge begs interdisciplinary research.
Questions:

1. What drives research in Africa, and do these forces favour interdisciplinarity?
2. What framework encapsulates transport policymaking at regional and continental level in Africa, and does the process favour

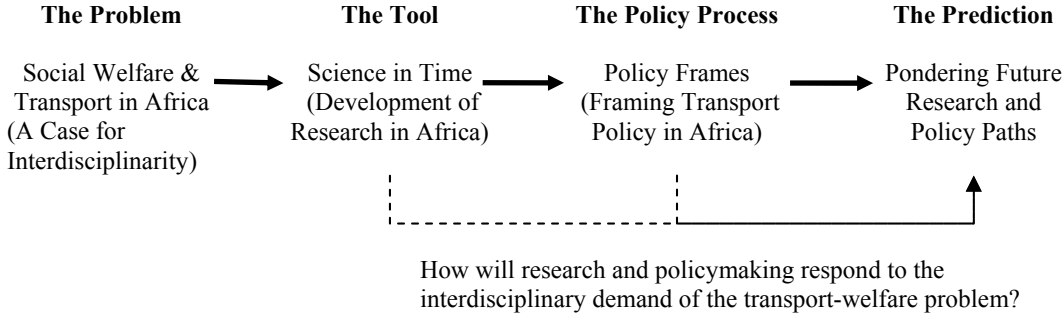


Figure 1. The structure of the paper.

brought to perspective by Table 1. In respect to the productivity of road network (i.e. paved roads to GDP ratio), SSA compares well with China. However, in terms ofaccessibility, ["Paved Roads per 1000 Persons" is a good proxy for accessibility], SSA lags way behind the other three regions. This is a major problem in the region. Most people do not have access to reliable transport infrastructure. Not only is this an economic problem, it is a severe sociological problem. Given a high rate of urbanisation, family interaction has been severely hampered. The urbanites struggle to maintain health family relationships. This means the growth economist attending to transport policy must have a strong appreciation of sociology, and vice versa, for a truly engaged discussion to be realised.

The HIV epidemic in Africa has presented a new challenge to the transport sector. In light of the fact that transport nodes (e.g. overnight stoppages & boarder crossings) are an important platform of transmission of the virus, nearly all transport policy initiatives consider HIV (& public health issues in general). The sociologist policy actor must therefore have good grounding in public health, and so must the doctor understand basic sociology [note: In Africa medical doctors often feed transport policy forum. Fortunately, medical doctors trained in public health have a strong appreciation of sociologist, a credit to interdisciplinarity].

The challenge of transport policy is not just one of growth economics, sociology and public health. Nearly all regional transport policy initiatives are driven by trade objectives. The continent does not have north-south link at either the west or the east. Neither is there a continuous east-west link (ref. Figure. 2). This poor interconnectivity is a hindrance to inland trade and competitiveness of the continent.

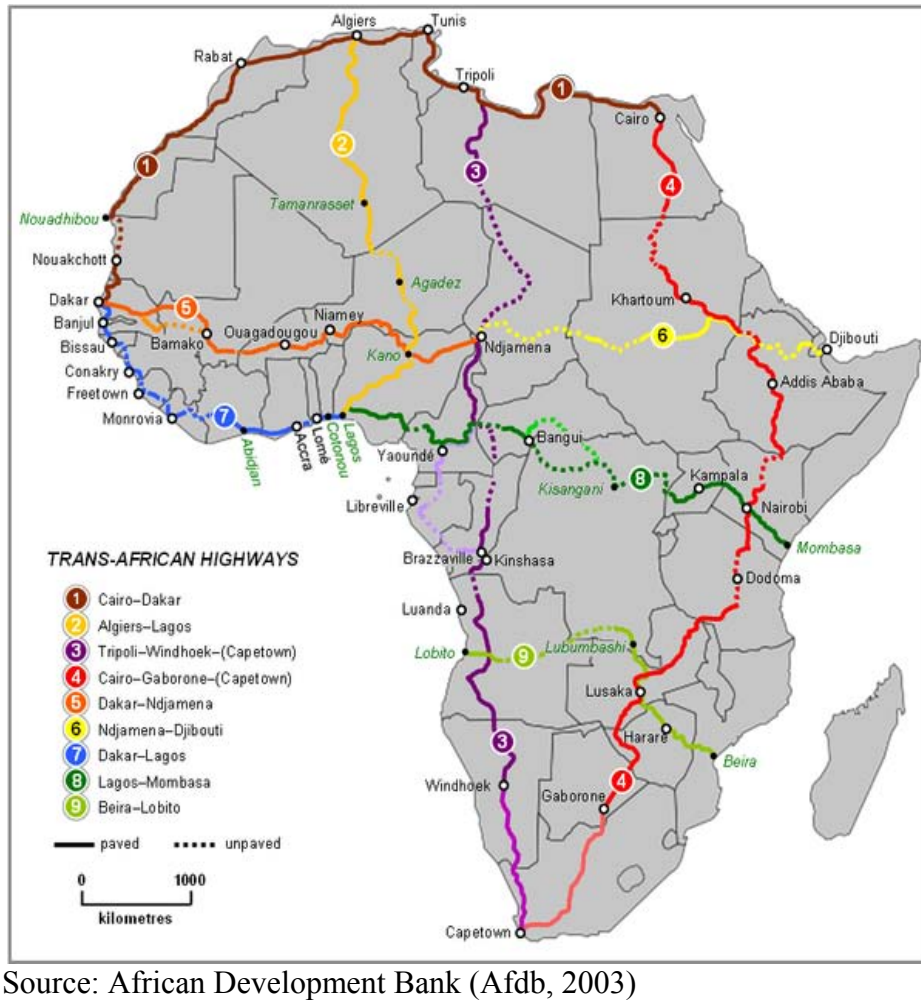
Table 1. Road statistics for selected regions.

	Sub-Saharan Africa	China	India	UK
Land Area (1000 km^2)	24260	9598	3287	244
Population (million)	800	1,318	1,125	61
Population Density (persons per km^2)	33	137	342	250
Gross Domestic Product (billion US$)	847	3,206	1,177	2,772
Gross National Income Per Capita (US$)	951	2,370	950	40,660
Paved Roads (1000 km)	225	898	1420	388
Paved Road Density (km/1000 km2)	0.01	0.09	0.43	1.59
Paved Roads (km) per 1000 Persons	0.28	0.68	1.26	6.36
Paved Roads : GDP Ratio (km per million US$)	0.27	0.28	1.21	0.14
Aggregate Traffic (million vehicle-km)	350,000*	850,000	1,400,000	1,100,000
Per Capita Traffic (vehicle-km per person)	472	651	1279	18333
Land Area (1000 km^2)	24260	9598	3287	244

Note:* Assumes Traffic: GDP ratio (mobility productivity) of 600 million vehicle-km per billion US dollars.
Sources: WDI (World Bank, 2009) and World Road Statistics (IRF, 2007). GNI is calculated using the Atlas method.

A World Bank study (Buys et al., 2006) shows that investing US$1 billion on the improvement of part of the Trans-African Highways (TAH) [only the Sub-Saharan section] could expand trade by US$ 250 billion. Such potential benefits imply that the mechanics of trade and transport engross regional transport debates. The actors to the policy platform must therefore have a conceptual and linguistic appreciation of trade economics to actively participate in the debates.

A new strand is being added to this transport policy complexion. It is public-private partnerships (PPP). All the six regional transport projects studied in the lead to this paper had a strong PPP dimensions. Each of the requests for proposal (RFPs) specified a "PPP Specialist" on the team. Conspicuously, none of these RFPs did exactly specify the disciplinary background of this role. This raises a few questions: Who are these "PPP Specialists"? How and where are they trained? Unfortunately, a study of past and potential holders of this role [based on the six projects] reveals that none of them had interdisciplinary academic training. This means none of them was formally trained to frame complex problems that PPP presents. This poses a danger. It means the quality of knowledge that they brought to the policy platform was most



Source: African Development Bank (Afdb, 2003)

Figure 2. The proposed trans-african highway.

likely very narrow. It also means that their ability to communicate their thoughts to policy actors was limited. It will be interesting to observe whether this new challenge inspires interdisciplinarity in African universities and research centres – as they aspire to influence transport policy.

The transport policy arena faces not just intertwining welfare issues, the regional organisations entrusted with policy formulation are numerous and complex. This partly explains why despite the indubitable high economic returns of investing in transboundary infrastructure, the TAH remains a dream after almost four decades. The remainder of this section will shed light on the structural (institutional) and agency (organisational) issues associated with transport policymaking at the regional and continental level in Africa. This will set the stage for a discussing the internalisation (or uptake) of science in policymaking in Section 4. Note that the term 'institutions' refers to humanly devised constraints that shape human interaction, which includes rules, norms and practices. Organisations, on the other hand, are structures developed to take advantage of opportunities created by institutions. These definitions are borrowed from Douglas North (1990).

A key actor in infrastructure policymaking in Africa is the regional economic communities (RECs). These organisations aspire to be free trade areas (FTAs) and customs unions. They have infrastructure development as a key item on their policy agenda. A list of the RECs is shown in Table 2, and mapped out in Figure 3. These two abstractions show, first, that the RECs are a complex mosaic of organisations, which are prone to tension as countries fall in and out of different zones of allegiance. Efforts at integrating the RECs through development programmes such as the North-South Corridor (Figure 3: highlighted in red), lessen but do not eliminate the potential for tension.

Secondly, these countries have different colonial legacies, and hence align to different international development partners. Moreover, they have different national languages (mainly English and French), which makes consensus building particularly onerous. Note, unlike the West, bilingualism is not well developed in Africa.

This section has shed light on the interaction of transport and social welfare in Africa. It is a complex mesh, which does not become simpler with the inclusion of environmental concerns. These non-linear complex issues cannot be solved by simple cause-effect consideration. They require interdisciplinary research. The following section attends to the development of science on the continent, and sets a platform for questioning the appropriateness of the existing paths in the development of research on the continent. This rendering is done against a background of evolutions of research in the West.

Science in Time: Tracing and Framing Progress of Research in Africa

The state of transport policy research in Africa can best be understood against a background of research evolutions in the West. Motivations for scientific research have changed over time. This journey is traced in Figure 4; for further study, Nowotny et al. (2001) trace this path succinctly. Before the World Wars, science was considered the tool for imperialist conquests. Government research institutions sought the latest outputs of university research in a bid to lift the country's imperialist status. During the Wars [1914-1918; 1939–1945], research was an instrument of war. It allowed good roads to be built; and superior weapons to be manufactured. After the World Wars, research became mainly an instrument of competition between the Cold War camps (USA and the Soviet Union, and their respective allies).

Trends in the West influenced the growth of research in Africa. Figure 5 lays out results of a textual analysis of ten countries in East, Central and West Africa. Before independence, colonial powers offered scholarships to African students to study and research in Western universities. These students would return home and work as government bureaucrats. It must be stressed that this pattern was most prevalent in the run up to independence.

Table 2. Regional economic communities.

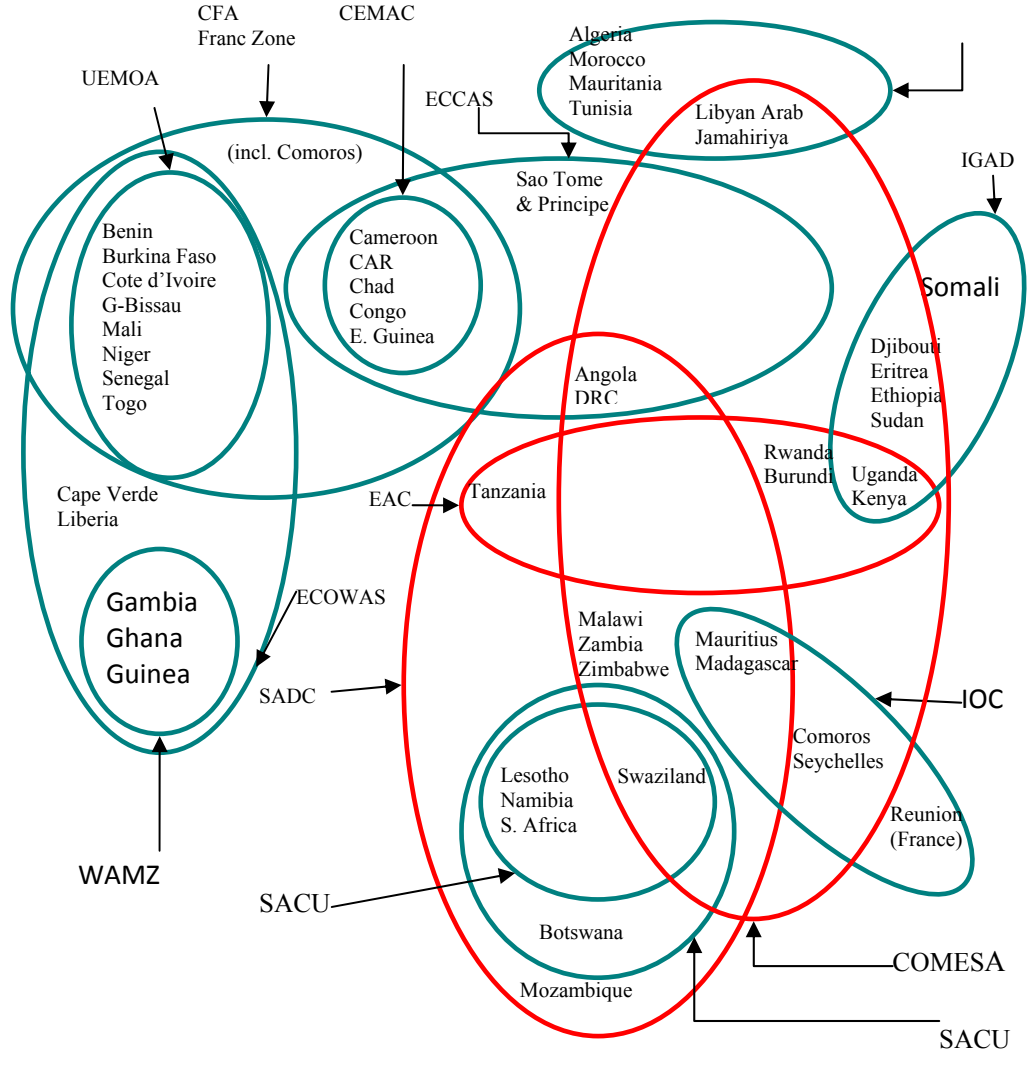
REC	Status
Arab Maghreb Union (UMA)	Free Trade Area
Common Market for Eastern and Southern Africa (COMESA)	Free Trade Area
Community of Sahel-Saharan States (CENSAD)	Free Trade Area
Economic Community of Central African States (ECCAS)	Free Trade Area
Economic Community of West African States (ECOWAS)	Free Trade Area
Inter-Governmental Authority on Development (IGAD)	Free Trade Area
Southern African Development Community (SADC)	Free Trade Area
Economic and Monetary Community of Central Africa (CEMAC)	Customs Union
East African Community (EAC)	Customs Union
Southern African Customs Union (SACU)	Customs Union
West African Economic and Monetary Union (UEMOA)	Customs Union

However, a missing link persists in this emerging model. Hardly any SSA university, outside RSA, has long-term scientific research collaboration with the private sector. This is a troubling observation because failure to engage the private sector stifles the development of interdisciplinary research, which is a key tool to solving the complex problem of transport policy [& most public policies] on the continent. Gibbons (1991) posits that a diffusion of doctoral researchers to private organisations underpins the growth of multidisciplinary research. Repko discusses how multidisciplinary research is mutates into interdisciplinarity.

Transport Policymaking in Africa

The previous section traced the path of research in Africa, and argued for interdisciplinarity as the tool for understanding the complex transport-welfare interactions. This section focuses on the making of regional and continental transport policy. The section has three objectives.

First, to frame (or order) the process of regional and continental policymaking. The second objective is to assess whether the process lends itself to efficient use of science. The third is to examine how the tructure of Africa-based research could change in order to have a greater influence on policy.



Source: Adapted from UNCTAD, 2009
Note: The RECs constituting the North-South Corridor are highlighted in red.

Figure 3. The Anatomy of African Regional Economic Communities.

The paper views continental transport policy through the lens of institutional rational choice (IRC). This framework (IRC) is a collection of policy frames that provide for the influence of the institution on otherwise rational policymakers (Ostrom, 2005). It therefore has rational choice theory as one of its key tenets. That is, it appreciates that to an extent, individuals make decisions after comparing the costs and benefits of different courses of action. Further, the IRC framework provide for bounded rationality. This concept, credited to Herbert Simon (1957), posits that rationality of individuals is compromised by the amount and quality of information available to them; their ability to access of this information (cognitive limitations); and the limited time available to them to make decisions.

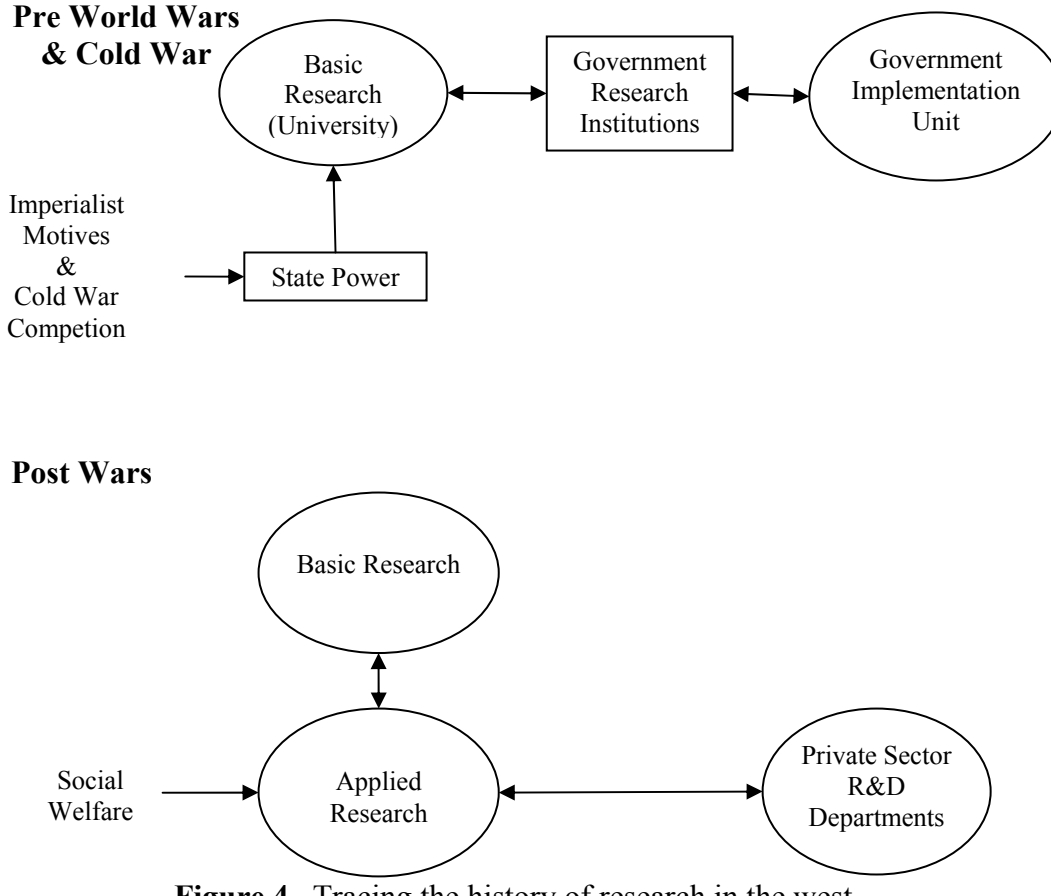


Figure 4. Tracing the history of research in the west.

A textual analysis of six regional policy initiatives in Africa over the past decade yielded the frame shown in Figure 6. The analysis was guided by institutional analysis and development (IAD) framework developed by Ostrom; IAD is a genre of IRC. The resulting framework in Figure 6 shows that the science feed is restricted to donor agencies. Local political actors hardly have scientific research feeding their thinking prior joining the decision arena. Entities like parliamentary research groupsand local autonomous research centres are yet to make a significant contribution on local political views. Consequently, the review and revision loop bypasses local political actors.

The positioning of science agents in Figure 6 limits the influence local actors can have on policy. If the political actors are intellectually well equipped in a particular aspect of transport (for instance, environmental issues), this aspect is likely to dominate policy. This assumes the actor has sufficient political influence to dominate the action arena. A setting where science interacts with politics before the decision arena would result in a better balance of science influences. That is, the various strands of disciplinary issues in the transport-welfare interaction would be betterconsidered. This detachment of science from politics can be degenerative. However, as discussed in the next section, the emerging model of research presented in Figure 5 should eventually close this gap between science and politics.

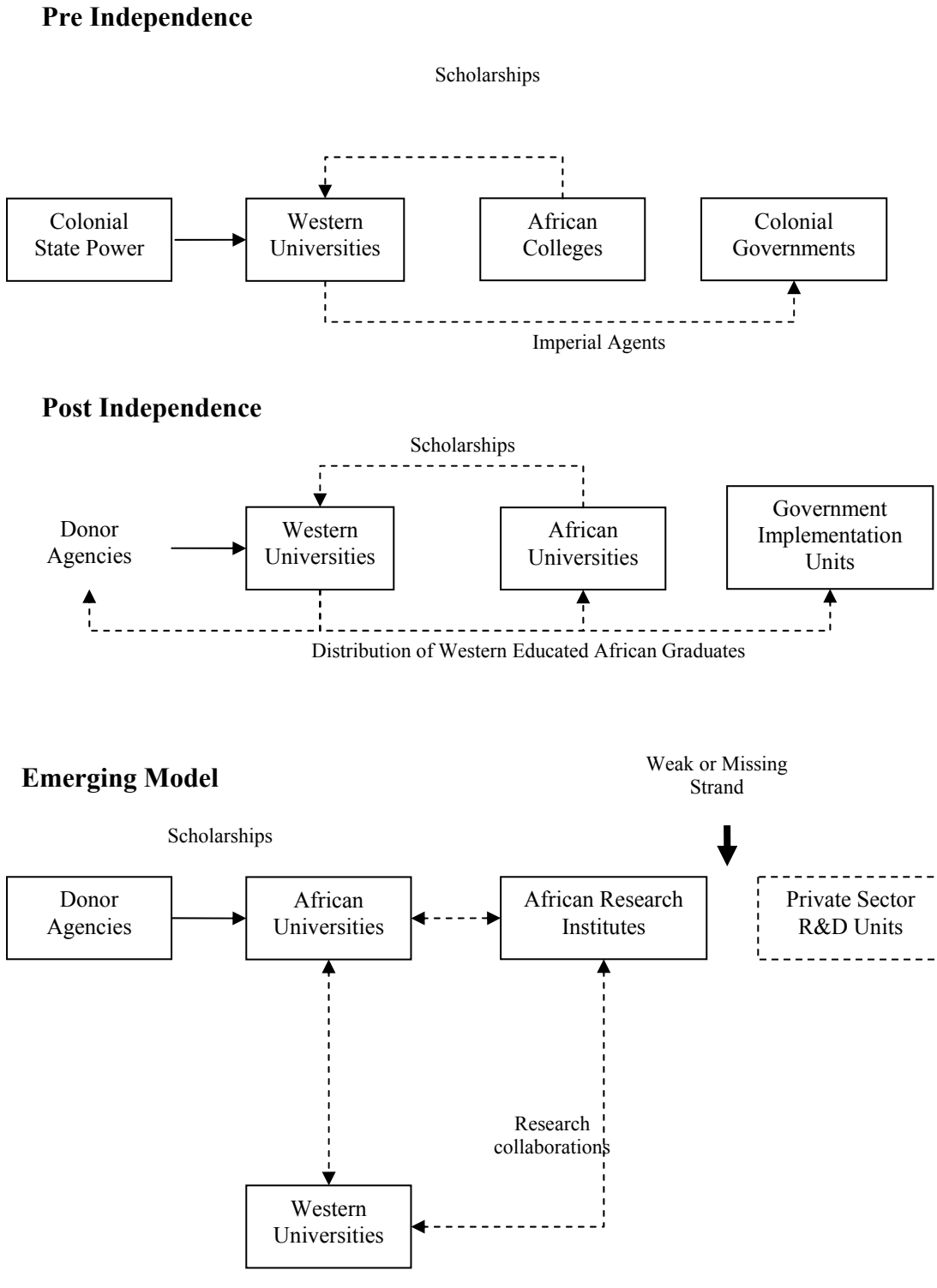
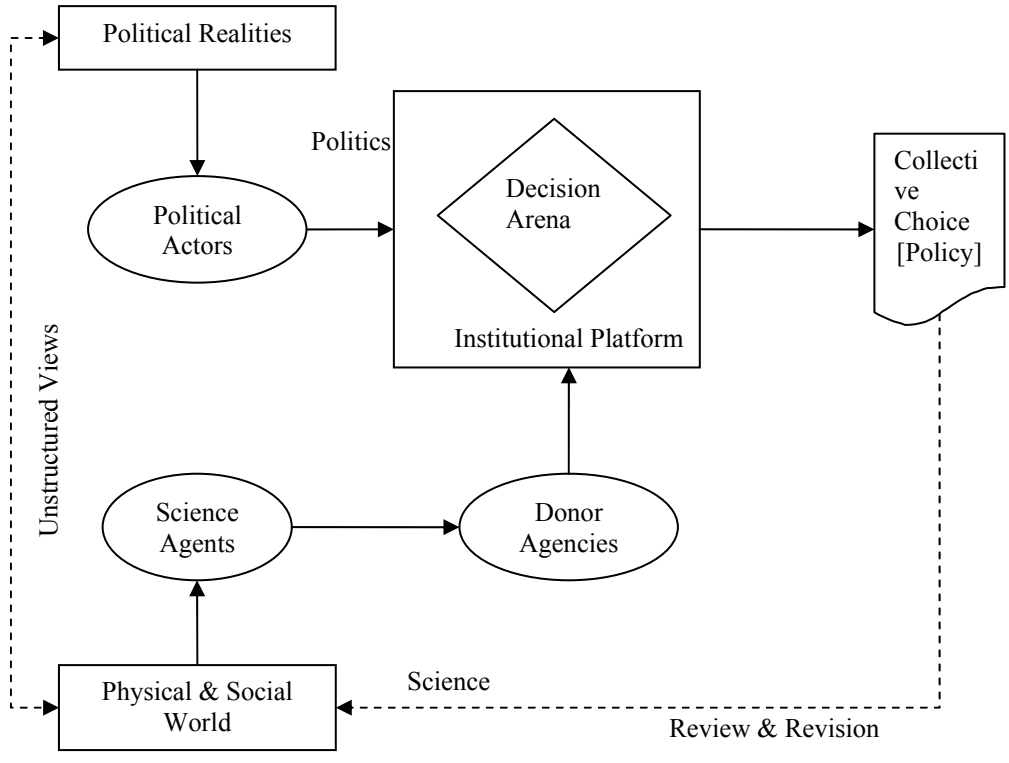


Figure 5. Mapping the emergence of research in Africa.



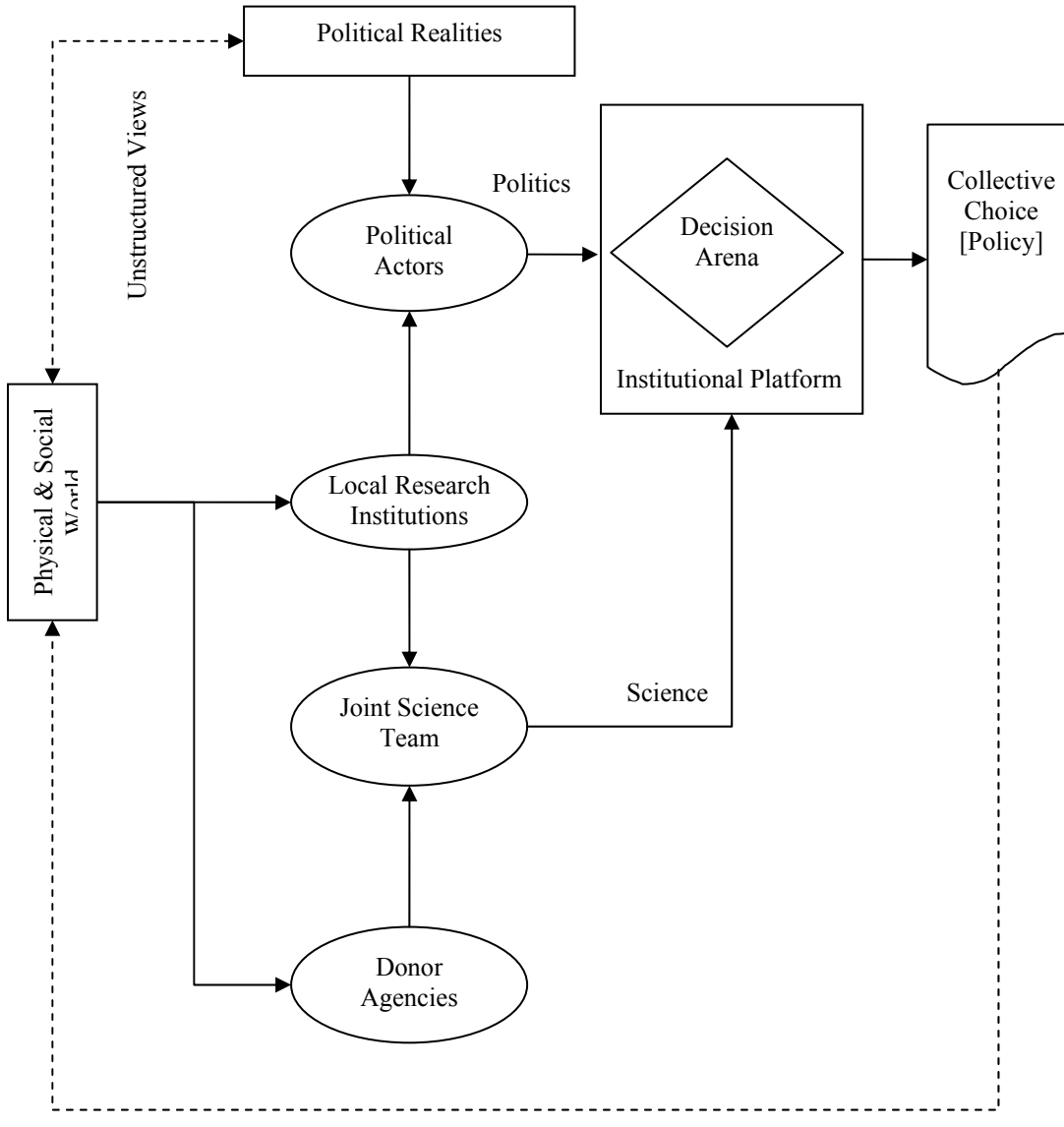
Note: There is a gap between local science agents & political actors

Figure 6. The dominant model transport policymaking in Africa.

Pondering Future Research and Policy Paths

The process of policymaking shown in Figure 6 is likely to change in response to the emerging model of research shown in Figure 5. This predicted change is captured in Figure 7. This framework places local research institutions at the centre of policy guidance. They feed both the political actors and a joint science team, which includes scientists from donor agencies.

Note that the new framework brings science close to institutional and organisational challenges of policymaking. It should allow local scientists from a variety of disciplinary backgrounds to interact in finding solutions to a mesh of welfare problems that relate to transport. These local research centres would begin as multi-disciplinary hubs, and the budding centres in the region are multi-disciplinary. Repko (*ibid*. Section 1) argues that such multi-disciplinary centres should gradually influence local universities to adopt interdisciplinary research. Interdisciplinarity in local universities should produce scientists better equipped to frame the complex challenges that transport policymaking presents.



Note: In the emerging model, policy science is a product of Africa-based research drawing on global collaboration.

Figure 7. Exploring future transport policy process in africa.

Conclusions

This paper has exposed the challenges at the interface between transport and social welfare. These challenges transcend the traditional academic disciplines. They beg interdisciplinarity. The paper has examined the future of interdisciplinarity in the transport sector using two lenses: the history of research on the continent, and the making of transport policy at regional and continental level. A poor linkage between private firms and the research community is identified as the key hindrance to interdisciplinary research. This begs the question: How can a stronger link be developed between the private firm in Africa and budding research institutions on the continent? The paper points to politics. If the joint research centres [of local

researchers and mainly Western donor-based researchers] can be used by local policymakers as the main guide in decision-making, a culture of science is likely to evolve. On the other hand, it could be argued that the quest for efficiency, will drive science in the marketplace, and subsequently infuse respect for science in policymaking. Will politics or the market lead? Further enquiry [research] is required. One thing is certain though: The enquiry must embrace interdisciplinarity. Finally, this can be achieved without compromising the rigour [depth] associated with disciplinary efforts.

References

African Development Bank/United Nations Economic Commission for Africa. (2003). "Review of the Implementation Status of the Trans African Highways and the Missing Links。" Volume 2: Description of Corridors.

Buys P. and Others. (2006). "Road Network Upgrading and Overland Trade Expansion in Sub-Saharan Africa."Washington D.C.: World Bank.

Gibbons, M. and Others. (1994). *The New Production of Knowledge*. London: Sage.

International Road Federation (IRF). (2007). World Road Statistics 2007. Geneva: International Road Federation.

North, C.D. (1990). *Institutions, Institutional Change, and Economics Performance*. Cambridge University Press.

Nowotny H., Scott P., and Gibbons M. (2001). *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*. London: Polity Press.

Repko, A.F. (2008). Interdisciplinary Research: Process and Theory. Sage: London.

Sabatier P. (1991). *Toward Better Theories of the Policy Process.* Political Science and Politics, 2(24), (Jun., 1991),147-156.

Simon H.A. (1957). *Models of Man: Social and Rational*. New York: John Wiley and Sons, Inc., 1957, 279.

UNCTAD, (2009). Economic Development in Africa Report 2009. Strengthening Regional Economic Integration for Africa's Development. Geneva.

World Bank, (2009). World Development Indicators (WDI). Available at http://web.worldbank.org/WBSITE/EXTERNAL/DATASTATISTICS.

Zografos G. and Tsanos C.S, (2007). "Methodological Framework for Analysing Institutional Settings in Transport Policy Making." Transport Research Records: Journal of the Transportation Research Board. No.2036. Washington, D.C.

Taxi Demand Modeling to Ensure Appropriate Taxi Supply. Do Both Open Access and Model Based Restrictions Fail?

James M.Cooper[1] and Wim Faber[2]

[1] Head of Taxi Studies, Transport Research Institute Edinburgh Napier University, Merchiston Campus, Edinburgh, EH10 5DT, UK. PH: +44 (0)131 208 1055, Email: j.cooper@napier.ac.uk
[2] Senior Editor, Taxis and Public Transport, Challans and Faber Business Communications, 312 Av. Rogierlaan, B1030 Brussels. PH:+32 (0) 2 216 3464, Email: wim.faber@skynet.be

Abstract

As a transport mode, the taxi is often given scant attention. The mode is consistent, in appearance and function, and has followed the same patterns of licensing, rather than mainstream transport, from its origin as the first licensed public transport mode to current day arguments between regulation, whether a market operates better under a controlled or de-restricted model. The apparent lack of reporting does not, however, equate to a lack of analysis. Indeed a significant staple of consultancy and authority work exists. This including, in most mature markets, reporting specific to regulation; much of which is focused on the choice between a restriction placed on the numbers of licenses issued and the, apparent, alternative of allowing free market entry.

Underlying the current, often complex, approaches to market form and its analysis, is the simple question, whether there is a "golden number" of taxis that should be provided in any one location; and whether there is a "golden rule" against which this number can be established. Indeed whether either open market access or controlled determined restriction can achieve an optimal solution.
Key Words: taxicabs; regulatory economics; taxi demand modeling; market control

Introduction

To most people the taxicab is an immediately recognizable and readily understood form of transport. The taxi provides, in a variety of guises, a transport service understood and replicated across national boundaries. The most common services

include exclusive transport for individuals, often able to pick up on demand (UK: Hackney Carriages), or by pre-booking (UK: Private Hire); but may also extend to small shared vehicles (UK: Taxibus; AKA: Jitney) and contracted services such as transport provided to schools, hospitals etc. The same range is visible in many locations with Hackney[1] type services virtually universal regardless of country or prevailing transport policy. To many it is this level of commonality that suggests a mature indeed unchanging form of transport. The taxi is indeed the oldest licensed form of public transport with a documented history of supply from the 1630's in both Paris and London. This said, the actual nature of a taxi can differ, as may the controls applied to it, with black vehicles common in London, Yellow Cabs in New York (see: Cairns and Liston-Heyes, 1996; Bakker, 2007). Yet, and despite the commonality of the mode, a controversy exists specific to the control of the vehicle, including the analytical approaches used in their determination (UK: Taxi Demand Models). The basic choice, whether to control or not remains to this day an unsolved dilemma faced by authority, and impacting on the everyday lives of the supplier and user of the mode. Each approach remains vehemently supported and opposed in equal measure, while both display benefits and issues.

This paper identifies arguments as applied to exclusive ride taxis in urban environments, to pose the question whether either open market or model based restriction fail to deliver. The paper identifies that the potential benefits on supply suggested of de-restriction may not have been fully realized. Indeed evidence exists that open access markets can be associated with (negative) impact across a wider range of metrics, in identifiable segments, and most particularly as impacting on service quality and the nature of taxi employment itself. In contrast, the most commonly applied alternative, a defined number of vehicles arising from taxi demand, may provide an appropriate service level, but only where their analysis is determined as part of a wider framework intended to allow flexible market entry rather than as a tool to copper fasten a closed market. The paper further concludes that the most significant failing in a model based approach relates to the application of a standard model on the basis of "one solution fits all" rather than expanding this approach to take cognizance of the full range of local circumstances, including engineering and infrastructure.

Taxi Controls

Restrictions may be placed on the taxi in one of three ways: applied to the Quality of vehicles, the Quantity of vehicles licensed to operate, and Economic controls, predominantly applied to tariff. It is the combination of these three elements (QQE) that underlies the dichotomy between what is visibly a highly consistent service, and one that continues to precipitate highly charged debate between proponents of

[1] The term Hackney Carriage carries legal definition in many countries and is most commonly applied to vehicles able to ply for hire and pick up on street. The term originates from a French term for a cart pulling horse. The term, somewhat confusingly, is applied to Private Hire Vehicles in the Republic of Ireland.

regulation, and those seeking reform. Discussion is regular, if infrequent, and often allied to external events, such as changes to political policy direction (Beesley and Gwilliam, 1977); regulator comments, including those from anti-trust regulation bodies (OFT, 2003); or as a result of consumer or trade pressures. Adler and Posner (2006) go a step further, associating debate with a wider "Deregulatory Movement", seen as a political response to investment need in public utilities including transport. The "movement" in the UK is exampled by privatizations of most utilities, the Transport Act of 1985, which moved bus services to private control, and the 1993 Railways Act introducing franchised operation to most UK rail services. While taxis stood some way apart from the deregulations of the UK, services already being private, away from state ownership, the '85 act also brought the possibility of redefining regulatory controls as applied to the taxi industry in some parts of the UK, with similar reforms in New Zealand and Sweden on the 1980s, the Netherlands and the Republic of Ireland in the period to 2000. It is significant at this point to underline that the legislative structure of the UK differs between the legislature of its countries, England, Wales, Scotland and Northern Ireland, and reserved matters specific to the entire UK. In most aspects taxis remain the responsibility of the devolved administrations, while wider transport policy is split between the UK (Westminster) and devolved governments.

While wider regulation applies in any of the three QQE domains, in recent history, the discussion has tended to be seen as a more pointed argument for and against Quantity restriction (OFT, 2003)[1]. Regulation as applied to Quality tends to relate to vehicle roadworthiness, while Economic controls have been mainly been applied the tariffs (taxi fares). Although the three remain connected (Cooper, 2007), few criticisms are raised, at least publicly, in respect of safe vehicles, or of appropriate limits to the fares charged. Quantity control, in contrast, raises significant and vocal proponents and opponents, both focused on the concept of customer service, and both marshalling economic and market arguments for and against a stated position. Wider interactions also cross the regulatory domains. Linkages and inter-working between domains may, in fact, remain significant and can be highlighted by the use of detailed quality controls, including driver controls (CTR, 2009) on issues of quantity. We will return to the interactions between these elements later.

Allied to the choice between regulated and derestricted quantity controls, a number of analytical models have emerged associated variously with the determination of vehicle numbers, cost and tariff (DfT, 2006), as well as detailed network models (Yang and Wong, 1998; Wong, 2008) specific to patterns of highway use. The use of such models has, in the UK, become associated with the underlying argument

[1] The OFT retains a UK wide remit to provide assessment and advice on fair-trading. Application of any recommendations, however, remain in the domain of the competent authority, whether UK wide, devolved or local government. The OFT report of 2003 (OFT report 676) has no individual standing in law, and has been subject to significant criticism at local level, including its rejection by Scottish Ministers, and similar commentary by the (Westminster) parliamentary advisory committee on transport.

between restriction and deregulation, with the most common, an Index of Significant Unmet Demand (ISUD), being used to support the (political) decision to restrict or deregulate. Given the frequency of its application to quantity restriction in the UK (and for the purpose of this analysis) we will define ISUD as a standard approach), in UK practice, to quantity restriction. It is the application of a standard model, used to define finite taxi numbers based on ISUD values and its apparent alternative, open market access, that forms the basis of our question, do both open access and model based restrictions fail?

Unpacking the Taxi Model

The term 'model' has been used with a variety of meanings in transport literature (Ortuzar, 2004) and may differ in approach and application. A model of taxi services may thus be used to denote either a set of factors typical in applying a range of controls – a regulatory model approach to specific situations; or in relation to determining the impacts of a particular application (Cooper, 2007). The latter, an analytical model, being appropriate to the planning and optimization of services, and the basis of the demand models used in the determination of ISUD (Halcrow, 2002 et seq).

Analytical models may be split further between those that apply to individual areas of regulation (QQE) such as the standard approaches taken toward unmet demand analysis (ibid) and those concentrating on the effects of infrastructural change (Wong et al, 2008). Wong also identifies differences between "aggregate modes", those identifying a single market for taxi supply, and those that address spatially and temporally defined markets, a difference noted by Schaller (2007) and suggesting a split between the markets for cruising taxis (UK: hailed), dispatch taxis (UK: pre-booked), and cabstands (UK: ranking/stance). Temporal differences are also significant, the taxi market actually comprising a multiplicity of market segments varying by location and time, vehicle type and mode of engagement; each directly influenced by regulatory decisions, and each requiring a differing level of supply. Slight variations in approach can exist, allowing us to contrast between the UK majority approach of linking numbers of licenses to service levels, with a French model applied to Paris areas (prefecture) based on a matrix including numbers of bed nights and incoming tourists[1]. While both link changes in demand levels to an adjustment of taxi license numbers[2], the methods by which demand is assessed differ significantly.

The reality of differing taxi markets within a single location may also create something of a complicating factor in the design and use of a single standard model.

---

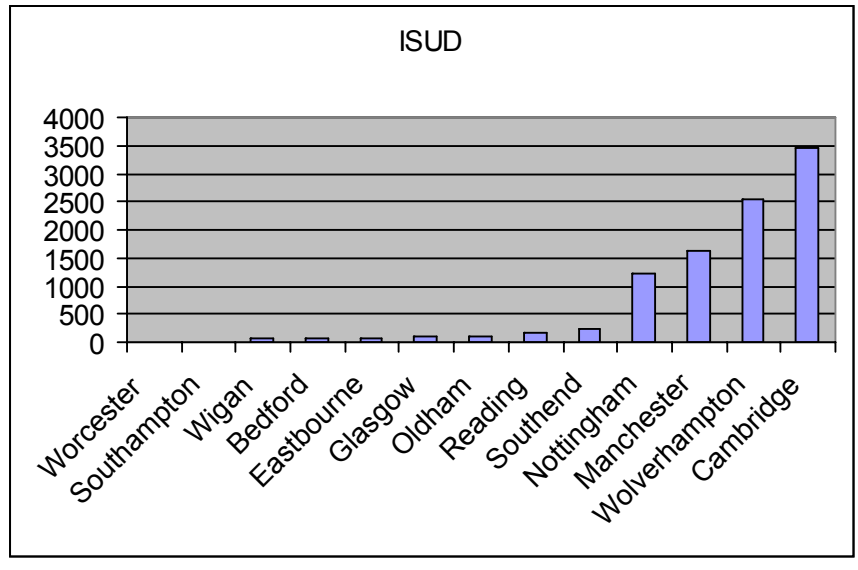
[1] A review of the approach adopted in Paris is available in Darbera (2009).

[2] A further question arises in that while license numbers are typically increased in light of increased levels of demand, no comparable reduction is mandated where reductions exist. Indeed it is questionable whether any such reduction would be possible beyond natural wastage – a policy adopted in Copenhagen, or by the costly expedient of buying back licenses (as occurred in the Northern Territories of Australia).

Traditional approaches, in the UK, have been to address assessment of the different regulatory domains (QQE) as single self standing issues, while approach other differences in market form as though the market was a single unified entity – an aggregate market applying to all times. Thus temporal differences, particularly seasonal and weekend markets and differing city locations may be considered as a unified homogenous market for taxis. Homogeneity within the market allowing for a single solution being applied to a perceived problem, most commonly an increase in the minimum number of licenses to provide for an aggregate ISUD. Such models have become routinely used and are described variously as taxi survey analysis, unmet demand, and taxi supply models. Their application is widespread, formulaic, and seen to satisfy UK legislation including the Civic Government (Scotland) Act, 1982; and the Local Government (Miscellaneous Provisions) Act 1974. Indeed it is the relationship between governing legislation and legal challenge that forms the basis of the strength and widespread application of the standard methodology. Yet this should not be an indicator of the appropriateness of the ISUD approach and may, in reality, pose a significant barrier where application is based on satisfying a legislated requirement (satisficing), rather than market optimization (See: Byron, 2004). This revolves around the apparent function of the ISUD model (in satisfying a legal reqirement) rather than the actual optimization of the fleet.

Application of the standard UK model, undertaken by the research team in 5 Scottish locations in the period 2003–2009 as well as the reported results from 20 similar studies undertaken by commercial consultants on behalf of licensing authorities, suggests a wide ranging variance of ISUD values, see Figure 1, with extremes between lower end values, taken to indicate an absence of unmet demand, and extreme higher end values, said to be indicative of insufficient supply.

The resulting range of observed results include significant differences in the patterns of taxi use, and highlight the difference between locations with extensive night time use and those with a mixed use of taxi services. The observation of peaked demand may, in reality, suggest a significantly different market at night to that during the day, and raise questions of the appropriateness of a one size fits all solution. The difference between day and night use is highlighted in Cooper (2005) and forms the basis of commentary by the UK Department for Transport (DfT, 2006) that peaked demand be included as part of the standard approaches to unmet demand assessment. Schaller's observation (ibid) of differing market segments also feeds into the increased complexity of assessment, and may question the appropriateness of single "license based" solutions, an approach commonly adopted which assumes that increased numbers of licenses will necessarily alleviate observed excess demand. The issue of a seasonal peak is acknowledged and has been partially addressed in the most recent studies including the standard model, but even this fails to account for a full consideration of the differing market segments, or indeed of the range of differences experienced by season, or the appropriateness of a single "one solution fits all".



Source: Previous SUD studies.

Figure 1: Tabulated ISUD values

Changes in license numbers are presented as the most common single solution to unmet demand and tend to follow a logic that more taxis within a fleet will result in a uniformly higher level of service, even in the face of apparent barriers to this.

While it is undoubtedly true that increases in circulating taxis has the effect of increased likelihood of availability, the link between licenses issued and the actual numbers of vehicles available at any one time is not linear, can result in a perverse negative relationship[1]; nor need a direct increase in licenses necessarily be the best method to achieve increased availability. Alternatives may relate to individual rank design, improved designs offering better operating patterns, reduced dwell times and more efficient use of the existing fleet. A significant improvement in cab shift availability can be observed in Glasgow following the introduction of "nite zones" a joint initiative including the deployment of taxi marshals at two of the busiest ranks. Our work over the period 2005 to 2007 demonstrates a significant drop in unmet demand as a result of this targeted initiative, with similar observations in the councils own reporting (Glasgow, 2009).

In other instances a change in numbers of licenses available in a city may have little or no effect at particular points, demonstrated in the example of an Edinburgh based study completed in 2009. The Edinburgh report identified an ISUD value indicative of unmet demand and recommended an increase without full consideration that the worst performing rank, at Edinburgh Waverly train station, was subject to additional permits as a private rank beyond the control of the licensing authority. In this

[1] Observed relationships between actual numbers of licensed vehicles and the number available (Cab Shift Availability) suggests that the additional numbers of shifts available develops at a lower rate than the increase in licenses available.

instance any application of changes to the total numbers of licenses issued within the city would have only a limited effect on cab shift availability at the station and would have no impact on the numbers of vehicles permitted to operate on the station grounds.

Solutions also based purely on license numbers (licensing based solutions) also overlook differences between market segments, passenger willingness (and ability) to pay (Chang and Chu, 2008), and differing levels of price competition (Price Waterhouse, 1993; Cooper, 2007).

Open Market Entry

The direct competitor to the standard model in the UK is most often seen as the removal of licensing restrictions (de-restriction), a practice also referred to a deregulation or delisting. De-restriction effectively allows any operator may to obtain a license without a quantity restriction being applied.

Open market entry has been followed by a number of authorities and was forwarded as a choice by the UK Office of Fair Trading report 676 (OFT, 2003). The benefits of free market entry have been described at a number of points as including increased supply (Buckeye, 1996; Boroski and Mildner, 1998; BSD Consultants, 1999). The economic impacts of regulation are also identified as related to and impacted by entry restriction (Laitila et al, 1995) suggesting deregulation may bring down fares paid by the public.

This said, the numbers of authorities choosing to de-restrict remains lower than anticipated in the UK. Worldwide experiences also suggest a limit to the benefits achieved where deregulation is followed, the most common being the experiences in the USA following the removal of license restrictions in 21 cities in the 1980's and the rapid return to regulation that followed. Price Waterhouse (1993) highlighted failings in the US as including: rate gouging, short-haul refusals, surly and discourteous treatment of passengers. Dempsey (1996) provides a similar review, summarizing that deregulation had resulted in: "a decline in operational efficiency and productivity; an increase in highway congestion, energy consumption and environmental pollution; a decline in driver income; deterioration in service; and little or no improvement in administrative costs." Notably both also identified an increase in the numbers of taxis within a city, but both suggested that the distribution of the fleet was not even. Mundy (2007) equally provides evidence of gaps in supply, using supporting data from Miami suggesting a distinct preference by drivers for specific locations and reducing or eliminating supply in poorer neighborhoods. This being supported in the case of deregulated environments of preferred locations including (Darbera, 2009) a driver mental map of places of likely engagement, or of likely high fares. Such patterns are observable in Dublin with significant queues of taxis at the centrally located O'Connell Street rank, in Galway, Ireland, around the Eyre Square rank, and is particularly noticeable at Dublin airport with mean driver

waiting times of 1-2 hours[1]. Trip refusals are also reported in the instance of short trips requested from incoming passengers at Dublin airport. Wider ranging impacts can be seen worldwide and are reported in Faber (2010).

Utopia Gained?

Despite the relative merits of regulated and de-restricted markets, it continues to puzzle me as to why neither has really gained the traction associated with best practice. De-restriction received legitimatization, in the UK, following the 1985 Transport Act, but with little effective uptake. The numbers of authorities allowing open access increased following the 2003 OFT report, which supported but could not impose deregulation, but it remains an observable fact that the larger numbers of vehicles, mainly in metropolitan cities, remain controlled. Indeed, the DfT (2006) and Scottish Government (2007) Best Practice Guidance give explicit credence to this approach by defining their view of approaches to justify restriction.

The overriding argument supporting both regulation and, ironically, de-restriction, has been the need to protect the public interest. The main metric used to this effect is the avoidance of significant delay, ie: whether numbers of taxis provide for taxis on demand. But it is this assumption, that increased numbers equates appropriate levels of supply, that I would call into question. Open market access has, without question, increased the numbers of licensed vehicles on the street, but has this led to an appropriate level of supply in all instances? Furthermore, the loss of quantity control has led, without exception, to a rapid increase in other forms of regulation, most notably a significant increase in the levels of quality controls applied. The possibility of increased regulation in the remaining QQE domains is acknowledged in the OFT report (ibid) as may be observed in practice in the Republic of Ireland (CTR, 2009) which has recently restricted new license issuance to accessible vehicles alone[2]. The move from regulated to deregulated market, and the further steps of re-regulation or increased regulation suggests (Faber, 2010), a cycle in which restriction remains a likely outcome.

As a counterpoint, it is also necessary to question the current use of demand models as the most appropriate method of determining market control. The ISUD model is in common application in the UK, and is mainly based on a common threshold value, but it remains appropriate to question whether this provides an optimal outcome. Questions remain as to how the ISUD threshold is calibrated, even how this is defined. Moreover, we would seek to question the wisdom that a model run provides a particular solution appropriate to all circumstances. Questions also exist as to the relationships between the authorities and those undertaking assessment, whether as a

[1] Observations were undertaken on behalf of the Irish Taxi Regulator in the period March – December 2009, see Cooper (2010).

[2] The practice of allowing for accessible vehicles to obtain licenses in a restricted market can also be observed in some UK cities.

genuine approach to achieving optimal services, or as a perfunctory "box ticking" exercise to achieve the statutory minimum status of a survey undertaken.

Nor should we ignore the relationships between quantity restrictions and controls applied in the other two domains, quality and economic regulation. In the 2007 thesis "Taxi Licensing, Regulation and Control, an analysis of taxi supply in medium sized UK cities" (Cooper, 2007) we suggested a necessary link between all three areas of control. That fare levels would necessarily impact on driver income, would necessarily impact on choice to operate, at what hours, and in which locations. Quality standards would impact on operating costs, would impact on driver income, etc., and that License numbers also affect the propensity to supply, at what level, and in which vehicle. The market does not operate independently in each of these domains, but as a continuum of all.

Neither the approach of the open market, nor that of current standard practice in the UK can be said to have fully delivered. Open markets consistently fail to achieve on levels of quality and may result in negatives of driver dissatisfaction, even protest and strike. More common approaches to modeling suffer from satisficing rather than optimization. Do both open access and model based restrictions fail? Well yes, probably. Are both approaches broken? Well probably not. In the widest sense, neither open market nor model-based restrictions have failed to deliver a taxi service. Taxis continue to operate in both restricted and open markets. Fares remain surprisingly similar between each, with many restricted cities actually operating a lower tariff than in open access locations. A simple conclusion may be to suggest that local circumstances should be allowed to dictate most appropriate controls, and that model-based assessment be developed to (be able to) take account of local circumstance beyond the extent currently possible. But this does present a relatively lightweight conclusion. Not least that this effectively perpetuates the distinction between model based and open markets as distinct alternatives. The need for model-based approaches to be fully aware of and contribute to the ability for a market to expand is key to its success and, by extension; the open market cannot be without assessment of the type applied to the modeled solution.

What Can Now Be Achieved?

In all of the above discussion, the overarching question continues to be: how to define the optimal taxi service. The function of the taxi remains clear, as does the public understanding of its use, and expectations of its availability, yet we have followed two distinct approaches: restrict to an ISUD threshold or deregulate as those these were the only two options available. We continue to define ISUD without a full understanding of the threshold values being applied, and fail to account for linkages between regulatory domains, where the impacts of one may be felt across each of the others.

In terms of the politics of taxi supply, the modeled approach currently appears to satisfy no more than a regulatory requirement, given the opacity of its results, and

fails to account for the fact that a properly functioning model is not the supporter of continuing restriction, merely a tool to determine optimal supply. Such a tool requires breadth of assessment, and transparency of results, beyond the level currently in place. This said, the analysis in place behind the model, or at least stated as its raison d'etre, does have a distinct and significant place in the definition and application of taxi services, even in locations having chosen to de-restrict. In its current guise the ISUD model provides a frequently applied index measurement and, as such, benefits from consistency in methodological approach. The current application may fail to adequately define the threshold or include wider regulatory impacts, but should not be written off as inappropriate to the assessment of taxi supply.

A reasoned conclusion is not, thus, the removal of the current index, but rather its development as a part of a better framework. Consideration is required of engineering, rank position and design, and better treatment of seasonal demand peaks, and temporal demand.

The aim, of understanding what is a "golden number", or indeed a "golden rule" by which numbers may be determined, does appear appropriate to optimal supply, but more on the basis of actual performance than on a single one approach fits all. This requires the removal of the common assumptions applied regardless of location, which have included assumed stochastic arrival and unlimited capacity. Nor should license numbers be seen as a single solution to the complex problem of supply. Individual rank operation, market segments and latent demand require more considered attention than currently undertaken.

Our current work, following a modeling framework set out in Cooper (2007), builds on current models to demonstrate the potential for both license and non-license based approaches to supply optimization. The current debate between absolutes of regulation and open access may, in fact, both prove inappropriate. Absolute limitation may reasonably be suggested not to have fully delivered, but equally, the alternative of an open market may also fail to deliver optimal services. We would also argue that a properly constituted model based approach does not equate to a closed or monopolistic market but should, in our view, allow for the determination of appropriate numbers of licenses specific to circumstance. The modeled analysis is not therefore the antithesis of a free market, but rather its facilitator.

Optimal Information as the Guardian of Optimal Services

In the writing of this paper we have sought to as wide a range of approaches to taxi controls as possible. It is apparent that the issue of restricted versus open-access markets is deeply felt and the cause of significantly differing views. Few papers have remained truly neutral to this argument, whilst our initial contention - whether both open access and model based restrictions fail – may fairly be seen as indicative that neither is perfect. We would contend that the cycle between regulation and

deregulation is in many ways an inevitable reaction to the deficiencies (in some cases perceived deficiencies) of the alternatives (in some cases the perceived alternatives).

To our view the actual form of regulation does not, and should never, replace the need for an understanding of the market, and assessment of its operation. In this respect optimal information, including the undertaking of transparent and repeatable models, is, without question, the guardian of optimal service delivery.

The current forms of assessment go some way to providing the level of information appropriate to policy decisions, but all stop short of a full review. It remains a truism that properly constituted scientific endeavor must be transparent and repeatable, and we would question whether some of the current models stand up to such scrutiny.

The relationship, in the UK, between the legislated requirements, in Scotland under the Civic Government (Scotland) Act 1982, and the most frequently applied models must also be questioned. In particular the apparent satisficing behavior in which the completion of a model run might be a sufficient outcome rather than the operation of the model as a method of delivering better levels of service delivery in terms of service level, and optimal infrastructure.

The link between local circumstance and actual assessment cannot be ignored, nor can this be achieved while a "one-size fits all" approach is followed. Current assumptions of market homogeneity, even where partially segmented, are likely to result in misleading analysis. Rank location and design remain an essential element in the ability of a taxi market to supply regardless of actual numbers of vehicles, whilst physical characteristics and management at rank will also impact on use. The inclusion of such factors in a reasoned review appears both appropriate and lacking.

In providing a reasoned conclusion, it would appear appropriate to suggest that some failings have existed in both open market access and model based restrictions. This said, we do not agree that these alternatives represent the opposed and polar positions that they first appear to be. It is our strong view that in either approach a full understanding of market operation is an essential pre-requisite to policy development. It is also appropriate to conclude that, where properly undertaken, a modeled restriction does not equate to a closed market but rather allows for the determination of appropriate numbers of licenses specific to circumstance. The modeled analysis is not therefore the antithesis of a free market, but rather its facilitator.

References

Adler, M.D., and Posner, E. A. (2006). *New foundations of Cost-Benefit Analysis*. Harvard University Press, Cambridge, MA.

Bakker, P., (2007) "Deregulation of the taxi industry: experiences in The Netherlands", *(De)Regulation of the Taxi Industry Roundtable* 133.

Beesley, M. E., and K. M. Gwilliam (1977) "Transport Policy in the United Kingdom", in Journal of. *Transport Economics and Policy*, Vol. XI no. 3

Boroski, John W., and Gerard C.S. Mildner. (1998). "Economic Analysis of Taxicab Regulation in Portland", *Oregon. Cascade Policy Institute Policy Perspective* 1007, April.

Buckeye (1996) Buckeye Institute. *Taxicab Regulation in Ohio's largest cities*. Dayton, Ohio.

Byron, M. (ed.). *Satisficing and Maximizing: Moral Theorists on Practical Reason*. Cambridge University Press, 2004.

Cairns, R. D., and Liston-Heyes, C. (1996), "Competition and regulation in the taxi industry", *Journal of Public Economics* 59:1-15, 1996.

Cooper, J. (2005) "Transport Policy in the Night Time Economy". *Report to the Department for Transport*.

Cooper (2007) "Taxi Licensing", *Regulation and Control, an analysis of taxi supply in medium sized UK cities.*

Cooper (2010) "Shared Vehicles at Airports: Role of Taxis, Shared Rides, and Access at Airports--International Comparison", *Presentation to the Transportation Research Board*, Jan 2010.

CTR (2009) "Commission for Taxi Regulation, Republic of Ireland. Economic Review of the Small Public Service Vehicle Industry". *Report by Goodbody Consultants on behalf of the regulator*.

Darbera (2009) Ou Vont les Taxis, *Descartes* et cie., Paris.

Dempsey, P. (1996) "Taxi Industry Regulation, Deragulation, & Reregulation: The Paradox of Market Failure", *Transportation Law Journal* 24(1): 73-120.

DfT (2006) "UK Department for Transport Taxi and Private Hire Vehicle Licensing", *Best Practice Guidance.*

Faber (2010) "How Deregulation Improves Reregulation", *Presentation to the Transportation Research Board.*

Glasgow (2009) "City of Glasgow Council. Internal report regarding ISUD demand.

Halcrow Group (2002)", "Dundee Taxi Study 2002", *report prepared for Dundee City Council* (et seq:).

Laitila, T., A. Marell and K. Westin (1995) "The Characteristics of the Taxi Industry, and the Passengers Perception of Taxi Four Years After the Deregulation", *KFB*, Stockholm.

Mundy R. (2007) "Taxicab Ridership Study, Miami-Dade county", *Tennessee Transport and Logistics Foundation.*

OFT (2003), "Office of Fair Trading", *The regulation of Licensed Taxi and PHV services in the UK. Report* No.676.

Otuzar, J. and Willumsen, L.G. (2004) *Modelling Transport*, Third Edition. John Wiley, Chichester.

Price Waterhouse (1993) "Analysis of Taxicab Deregulation and Re-regulation", *Office of Government Services*, Washington DC

Scotland (2007) *Taxi and Private Hire Car Licensing*: Best Practice for Licensing.

Schaller, B. (2007) "Entry controls in taxi regulation: implications of US and Canadian experience for taxi regultion and deregulation", *In Transport Policy*, No 14., pp 490 - 506

Wong, K.I., SC Wong, H Yang, JH Wu (2008) "Modeling urban taxi services with multiple user classes and vehicle modes", *Transportation Research Part B*, 2008 - Elsevier

Yang, H., Wong, S.C., (1998) "A network model of urban taxi services", *Transportation Research B* 32,

Actions Taken on Green Transport: Experiences of Beijing

Xiaoming Liu[1], Jinchuan Chen[2], Zhili Liu[3], Haibo Chen[4] and Yu Ge[5]

[1]Director, [2] Dr., [5] PhD Candidate, Senior Engineer, Beijing Municipal Commission of Transport, No.317, Guang'anmennei street, Xuanwu District, Beijing, China P.R., 100053, Email: liuxiaoming@bjjtw.gov.cn, chenjc@bjtrc.org.cn, geyu@bjjtw.gov.cn
[3] Dr., Lecturer, School of Traffic and Transportation, MOE Key Laboratory for Urban Transportation, Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China
[4] Dr., Principal Research Fellow, Institute for Transport Studies, University of Leeds, Leeds LS2 9JT, UK, PH: 44-(0)113-343-5355, Email: trahc@leeds.ac.uk

Abstract

Beijing is a city with a history of about 3,000 years. In the past several decades, pursuing to be industrialized used to be a main goal of the city which caused more and more serious pollution problems. Since the late of 1990s, actions have been taken to deal with the air pollution in Beijing in order to enhance living level of its residents. However, countermeasures taken firstly were almost on the adjustment of development routine of industry. Until recent years, especially approaching the 2008 Olympic Games, complicated schemes on green transport was concretely carried out. It was not only aiming for a currently clean air, but also a future sustainable development of transport and urbanization. Among the actions taken on green transport, there are exploration of low emission vehicles, priority to public transport, cycling and walking improvement and intelligent information techniques application. This paper summarizes these action schemes and what's on now in Beijing.
Key Words: green transport; low emission vehicle; public transport priority; cycling and walking

As the capital of China, Beijing city is always a flag to the cities all over it. The experiences and newly-built aims usually attract other cities' attention and probably cause emulations. In the past several decades, when pursuing of industrialization,

Beijing used to be successful. A Chinese top four iron factory was located in the west of Beijing city. A chemical industry with more than 10 thousand workers was located in the southeast of the city. There were many other factories such as jeep corporation, carbonize chalk factory, etc, scattered all over the city. However, industrialization brought with it more and more serious pollution problems besides its dedication to GDP (Zhili Liu et al, 2009). Since the late of 1990s, in order to realize a livable city, the government launched a multi-step plan on cleaning the air. However, firstly takings on is almost on the adjustment of industries. For example, some of the above factories were moved outside and far away from Beijing city. For those still located in Beijing, they should change their power from coal combustion to electronic, natural liquid gas etc. It was until about the year of 2000 when the transport system became one of the most important domains of air pollution campaign. However, it is the recent years, especially approaching the 2008 Olympic Games that the purpose of emission reduction changes from cutting down current air pollution to pursuing a sustainable city development. And as a result of this change, a series of complicated schemes on green transport was concretely carried out. Among the actions taken on green transport, there are plans of new energy vehicles, priority to public transport, cycling and walking improvement and intelligent information techniques application. This paper summarizes these action schemes and what's on now in Beijing.

Background

Though there are 16.4 thousand km^2 of the city's municipal area, the mountains skirting around the west, north-west and the north cover about 10.1 thousand km^2. More than 85% of the total population center on the namely urban area of so-called Beijing city, which locates in the south-eastern plain with about 1.1 km^2 area. Population, including about 17 million registered and about more than 4 million visiting/floating still keeps a steady increase of about 0.5 million per year, which causes more and more daily travel demand on the urban transport system. Moreover, some statistics and estimation show that, motorized vehicle especially private car ownership is increasing great simultaneously and the mode split of cars poses very obvious challenge to the whole transport system. Data about above fact (BTRC, 2001-2009) are shown in Figures 1, 2, and 3.

Unfortunately on the other hand, according to a comprehensive survey results on transport performance in 2005, car travel patterns changed dramatically during only 5 years, as shown in table 1 (BTRC,2007). It means that the car usage increased 362% but the efficiency of transporting passengers decreased more than 5 times.

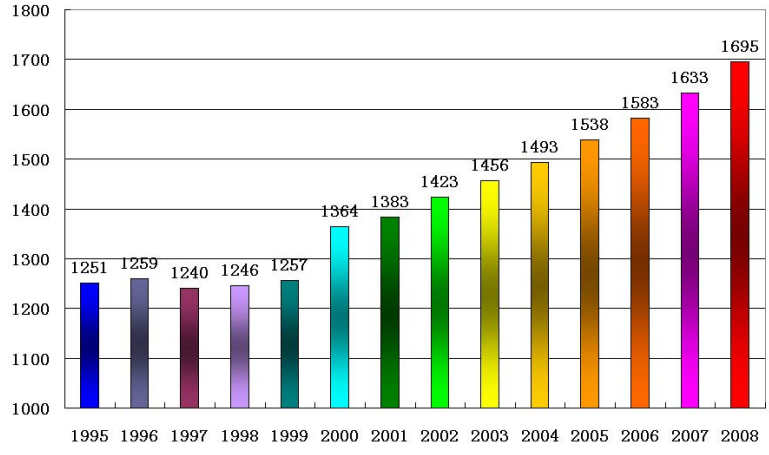


Figure 1. Population registered during the past years

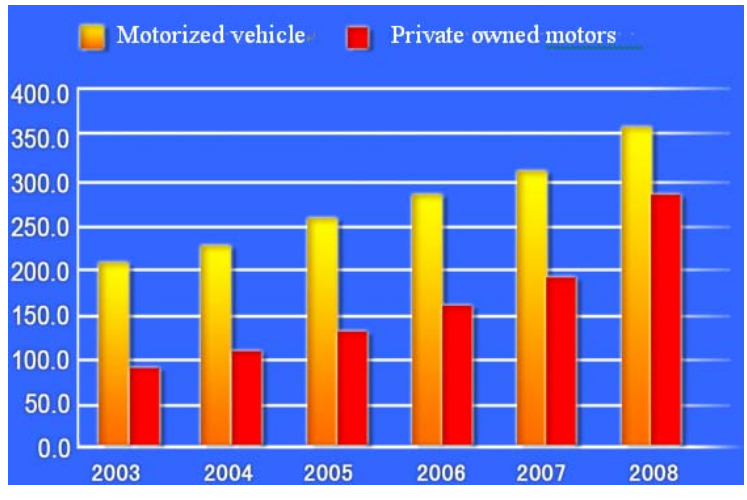


Figure 2. Increase of vehicle ownership.

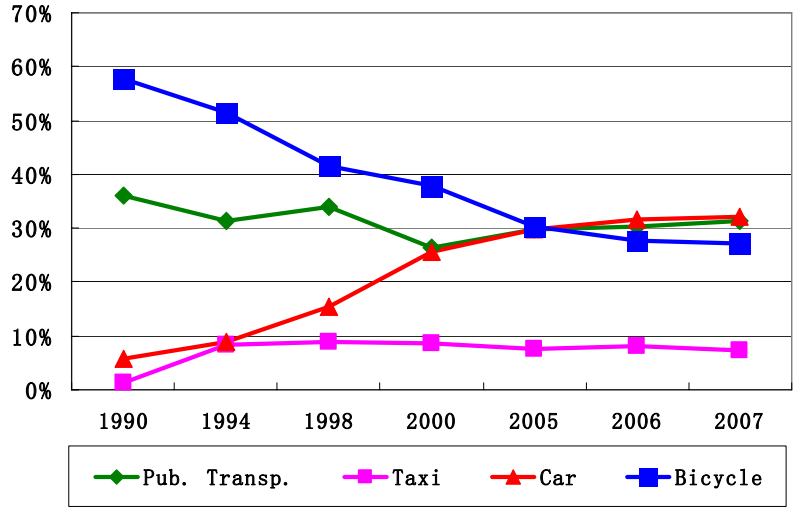


Figure 3. Car mode split is challenging.

Table 1. Survey results about car usage in year of 2000 and 2005.

Year	Travel frequency (trips/car/day)	Travel distance (km/car trip)	Occupancy rate (persons/car trip)
2000	1.26	9.7	1.57
2005	3.16	14.0	1.26

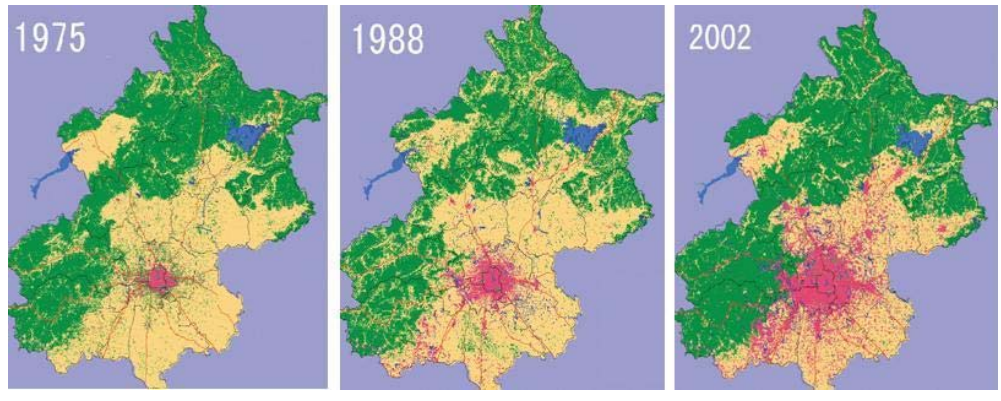


Figure 4. Urban area sprawl during less than a decade.

So, Beijing city is under a period of rapid developing and the problem is combined and complex. For example, the city sprawl (see figure 4) and traffic congestion is mixed together, waiting for solution. However, land-use pattern is changing very slowly, contrasting against the rapidly changing of travel pattern.

With the information above, it can be concluded that it is very challenging for transport to develop sustainably in Beijing.

Transport Develop Strategy

A white paper, *Outline of Beijing Transport Development*, was issued in April, 2005 by the government, addressing to develop rapidly a comprehensive transport system with urban railway and mass rapid bus system as backbone. It says that, the main two tactics are to develop transport in harmony with land-use and to adjust mode split pattern by giving priority to public transport system. (see http://zw.bjjtw.gov.cn/fzgh/zcqgh/200804/t20080421_17783.htm) Based on this white paper, some reform actions were firstly taken on bus service enhancement. We will come back this topic in a follow section.

In Sept. 2009, the government issued the *Action Plan of Cultural, Technical and Green Transport Development(2009-2015)*, which extends the conception of the 2008 Olympic Games aiming to realize a more culture friendly, high-technic and environment friendly development of transport. It addressed burdens during the

next 5 years (See http://zw.bjjtw.gov.cn/xbjjt/xbjjt_xd/ckyw/).

The two strategic documents are mainly on but not confined to the following measures dealing with transport development.

- Transport infrastructure construction. Since it is a booming era of development, there exists a great gap between infrastructure supply and transport demand. Experiences of success in hosting the Olympic Games and in tackling traffic congestion show that, there will be a long run of transport infrastructure construction to meet the transport demand in the future.
- Priority to public transport. The density of the city calls for a public transport based system to run the city effectively, efficiently and economically. The limited natural resources deny a car-led developing pattern of transport system.
- Travel Demand Management Measures. In order to diverse drivers from car to public transport, there must be some pole & carrots measures instead of just only public transport enhancement. This has been verified useful in the Olympic Games.
- Cycling and Walking Enhancement. In order to achieve a low emission system, cycling and walking are of great important role. Moreover, they are very crucial for the public transport system as well.
- Intelligent Transport System Application. Information technique now puts much profit onto the urban traffic and transport administration by adjusting the special-timing allocation of travel demand, assisting trip plan and routine plan, accelerating facility utility etc. There ought to be more functional results of ITS application.
- Exploration of low emission vehicles. This relies on new technique application to explore new energy or new powered vehicles. The fruits of this endeavor had been tested during the 2008 Olympic Games, which we will discuss at follow section.

Priority to Public Transport

Giving priority to public transport system has become a national policy of China. Beijing city gave publicity to the Declaration on priority of public transport development and began a series of actions on this item (Xiaoming Liu, 2009).

The importance of public transport was firstly recognized as one of the aspects essential for sustainable development and identified it as one of public affairs which should be driven by the government instead of the market. According to this knowledge, giving priority to public transport system means giving priority to land-use by, finance support to, road right allocation to and revenue help with public transport. Under this framework, there are some concrete measures.

Develop Urban Railway System. From the year of 2007, several newly-built urban railway lines opened and operated, such as the 4#, 5#, 10-1# underground metro lines and the airport urban railway, Olympic Branch etc. Now, there are nine urban railway lines with 228 km operating mileage in Beijing, which before the year of 2000 was only two lines with of 56 km.

In order to improve capacity of these lines, complex technical measures were taken to lessen the departure intervals of trains. Newly-opened lines take the original departure interval of 3 minutes as a technical standard of operation.

By means of above, there forms a stretched bone frame of railway system in Beijing central city like figure 5 shows. Passengers transport by the urban railway system now is more than 4 million per day, which was only about 1.8 million before the year of 2005.

There are more than 10 lines under construction now and with the pace of 100 km new line every two years, there will be 561 km urban railway system in 2015 in Beijing, namely, 333 extra km lines will be constructed in next 6 years.

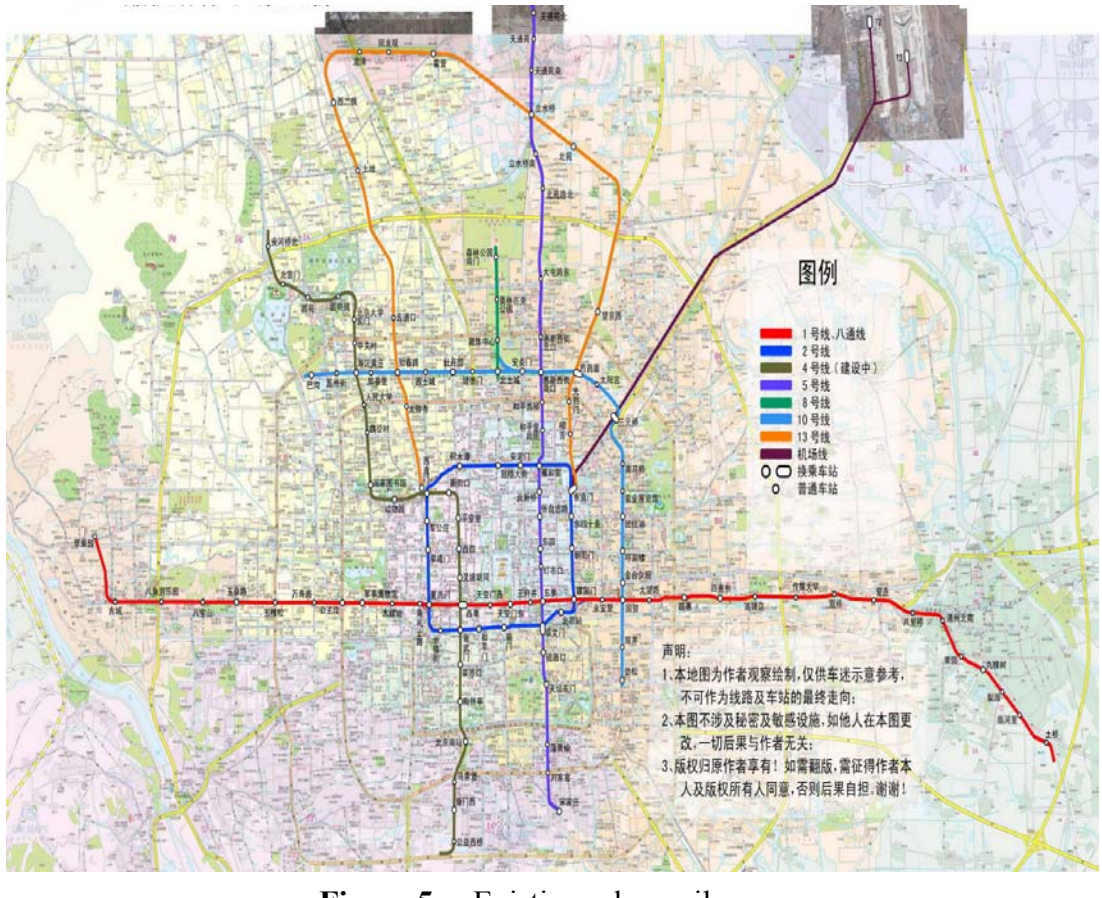


Figure 5. Existing urban railways.

Optimize Bus Routes and Extend Bus Only Lanes. Along with new railway line opening and the land-use development, many bus route lines are adjusted and optimized according to passenger demand and the functional changing of the bus lines. For example, some lines should be changed into feeder lines for the metro system. On the newly-built roads and streets, there should be allocate one lane to bus only use if there are 3 lanes in one direction. By the end of 2008, there have been about 400 km bus-only lanes stretching all over main streets of Beijing.

Tickets Discount. In order to attract passengers to use public transport systems, policies on ticket discount was researched and finally carried out. With the electronic ticket charging system, the intelligent card, passengers can gain a 60% off discount to take bus. No matter where you go, if the bus route number is within 600, one can expend only 0.4 Chinese dollars to take bus once. With regard of urban railways, one can pay 2 Chinese dollars to go as far as he like by train.

Despite of the pressure of this policy posed onto the public finance, ridership increase is sharply. Before the policy, the total ridership of bus system one day is about 9 million and after it, that of bus system became up to 14 million now. From the report of the government, this policy will last for a certain long period.

Construct Transfer System. In order to make the public transport system into a joint network, a cluster of transfer facilities were built and under construction. Among them, there are interchange stations, terminals, park and ride stops and so on. At these transfer facilities, railway-bus, bus-bus, bus-bicylce/walking and railway-bicycle/walking legs can be combined into a trip chain using public transport system.

New Technology

New technology is regarded as the radical measures for green transport because with it low emission or zero emission vehicles may be manufactured leading to no transport emissions in the future (Anthony D May, 2009). This was well recognized by Chinese government also. Several kinds of low emission vehicle or named new energy vehicle had been researched and tested before and during the 2008 Olympic Games. For example, since the 8th-five-year, there took a research on the battery electrical vehicle(BEV), sponsored by the Ministry of Science and Technology of the People's Republic of China(MOST) and the National Development and Reform Commission(NDRC). And the MOST started a special project on BEV application during the 10^{th}-five-year period. Under certain cooperation frame, universities, public transport firms, research institutes and some commercial corporations came together and form the project group to head for breakthrough on application of BEV.

They invent and patent, make technical standards and produces vehicle units and components.

During the Olympics Games(Zhili Liu et al, 2009), there were forty Coach of BEV served more than 145 thousand athletes and median delegates, five Coaches of BEV serve shuttle bus from Central park, Olympic village to No. 6 station, 415 Venue BEV served 592.2 thousand passengers, and 20 Cars of Fuel Cell Vehicles (FCV) served for the MOST and the Beijing Transport Management Bureau. Table 2, 3, 4 presents detail operating parameters.

Nowadays, low emission vehicles are mainly applied to the public transport system. At the beginning of 2009, there were 10 HEV and 50 NEV. In 2009, there planned to purchase more 860 HEV and 50 NEV. On 25th June, 2009, 50 HEV bus started operation on road and till 25th September, 2009, they had operated for 841 thousand km.

Cycling and Walking

Bicycle is always an important part of urban transport system. In the middle of 1980s, bicycle mode split in Beijing was up to 62.7% besides of bus of 28%.

Table 2. Coaches of Fuel Cell Vehicle(FCV) in Olympic Games.

Factory of Coach	Coach number	Km operated
Tsinghua Univ.	3	21thousand
Futian Motor	3	5.554thousand
Dark ?	3	4.365 thousand

Table 3. Coaches of Hybrid Electrical Vehicle (HEV) in Olympic Games.

Factory of Coach	Passengers served	Km operated
FAW	660 thousand	155 thousand
DFM	750 thousand	145 thousand

Table 4. Cars of HEV in Olympic Games.

Factory of Car	Passengers served	Km operated
Chery	-	1735 thousand
Changan	70.8 thousand	429 thousand

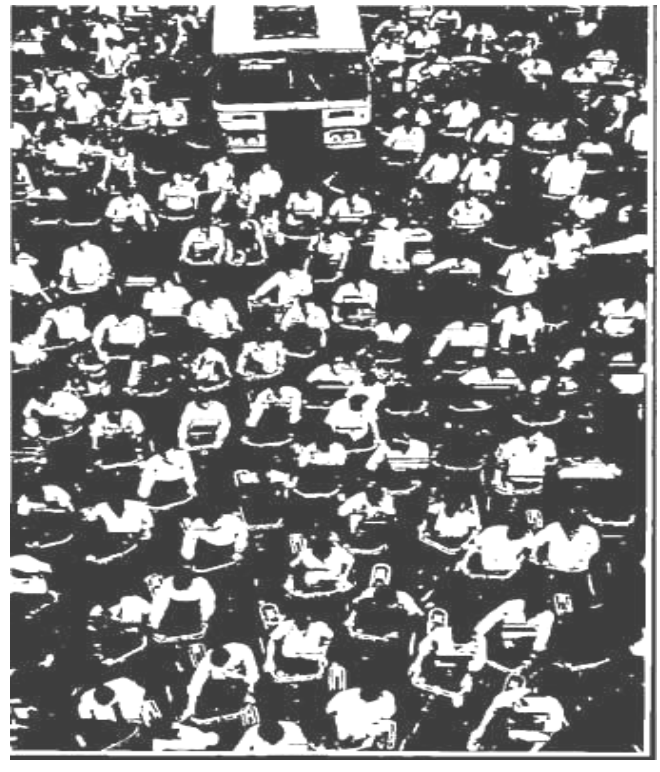


Figure 6. Bicycle traffic in streets of mid 1980s in Beijing.

Therefore, it is no doubt the streets were of sea of bicycles as figure 6 shows(Liren Duan,2008). Even now with the booming of cars, there are still about 10 million bicycles owned in Beijing and everyday there is more than 20% of the mode share taken by cycling.

In fact, cycling facilities are issued by the national technical standard of design and planning. For example, Figure 7 is a typical road section in Beijing(Xiaoming Liu, 2007). Width of cycling lanes is of 1.5 m, 2.5 m, 3.5 m or 4.5 m, which is varied according to the traffic flow demand of bicycle and the function of the street. More often, physical barriers are constructed between cycling lanes and motor lanes as shown below. Unfortunately, for a certain period in the past, in order to meet the broader and broader space needed by motors to drive, physical barriers used to be removed for extra space which leads to dangerous conditions to cyclists because not only in intersection but also along street does bicycle traffic disturbs motor vehicle driving and vice versa.

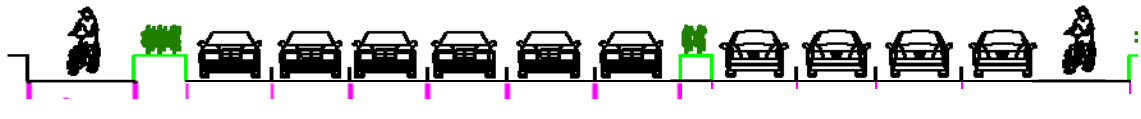


Figure 7. Typical road section in Beijing, lanes for motorway and bicycle.

From the Olympic Games, many actions have been taken to improve the condition, environment and safety of cycling. At each newly built railway stations, there should construct parking space for bicycles in order to attract travelers use bicycle and transfer railway. For example, there constructed more than 9 thousand parking places for bicycle around the stations of 4# underground railway. As a research project, public rental bicycle system is under consideration now by the government in order that there will be a rental stop every about 500 meters away all over the central city in the future.

As to walking system, similar measures are under taking now in case that walking is a radically basic element of trip chain. Width of side walk in the streets is about 4.5 meters. At railway stations, transfer terminals and other public buildings, provision of escalators is very common now. In some special areas such as the Wangfujing Street in the CBD area and the Zhongguancun Area, there has established walking only district or street.

Travel Demand Management Measures

As expressed in the *Outline of Beijing Transport Development*, there should be some pole & carrot measures to relieve traffic congestions. Besides of priority to public transport, there exist some other kinds of measures to manage the demand of travel and traffic. The following are some main items.

Car Driving Banning. It was during the Olympics in 2008 that car banning according license-plate every other day became a concrete and strict policy in order to reduce the emissions and meet the air quality requirement of the Olympic Committee. It led to very obvious improvement of traffic conditions of course. Post the Olympics, with the steady increasing of car ownership and with the hope of citizens to live in a clear city, this policy is extended for a testing period, from the end of Olympics to April, 2009. However, this time, cars were banned driving one day a week according to their license-plate. Based on its straightforward result in traffic improvement, this testing policy is extended from then on with a little bit change. For example, if there are two vehicles in a household, they can change the license plate for one of them. The weekday on which a car is banned is steady during a period of 4 months, which used to be changing every week. More detail information can been found at http://www.bjjtgl.gov.cn/zhuanti/08sxcs/index.html.

Sta*ggered Office Hours.* This measure also made effect during the Olympics in reducing rush hour travels. In the light of that it has been working for many years in overseas cities such as Los Angeles of USA, Munich of Germany etc., Beijing municipal authority determined to apply this measure for a long run and write this

down into its next step action plan for relieving traffic congestions, see http://www.china.com.cn/policy/txt/2009-06/02/content_17870534.htm.

Parking. Regarding parking facility construction, the *Outline of Beijing Transport Development* says there will be different supply rate of parking demand within different areas. That means within some special areas of magnitude attractions such as the CBD, Zhongguancun, historic city area and so on there will be insufficient supply to the parking demand. Of course, at the same time, good public transport service will supply sufficiently within and to/from these areas.

If people park their cars and then ride public transport mode at the P&R parks, the parking fee will be only 2 Chinese dollars for a whole day, which is one-hour fee rate at other parking places. This policy was greatly supported by the government finance.

Motorcycle Banning. This is an old policy in Beijing which dates back to late 1980s when Chinese economy started to grow fast as a result of the open and reform policy and people began to afford a motorcycle. In case of its turbulence effect posing onto street traffic and poor safety security, the Beijing authority banned its license in Beijing for many years. Some cities in China like Guangzhou, Shanghai etc. do not ban motorcycles on streets and there was troubles caused by them. Last year, Guangzhou authority banned motorcycles in its city.

Summery

Beijing is a world famous metropolitan and facing with problems caused by rapid development and growth. Not only future development is its main affair to considerate, but also the protection of historic relics is of important role. Transport drives the city for developing and affects greatly its environments. Facing the challenging situation and burdens, there are a complicated strategy on relief of traffic congestion, improvement of service, enhancement of safety and pursue of sustainability.

Policies are complex, contents varying at different time and different place. The interactions among them are more complex to address. This paper attempts to describe some main measures having been taken in Beijing in order to realize a green developing routine. For more detail information and more complicated scenarios about transport policy, there need another paper to formulate.

Acknowledgement

The authors wish to acknowledge the Beijing Elitist Training Fund for sponsoring

this research through Project 20071A1100700389. We also acknowledge the Open Foundation of Beijing key lab of transportation engineering, BJUT (grant No. BJUTTRC0804) and acknowledge the BJTU Science Foundation of China (grant No.2007RC103).

References

Beijing Transport Research Center (BTRC).(2001-2009). *Beijing Transport Development Yearbook*.

Beijing Transport Research Center (BTRC). (2007). *Report on the 3rd Comprehensive Transport Survey*

Beijing Municipal Authority (BMA). (2005). The Outline of Beijing Transport Development, http://zw.bjjtw.gov.cn/fzgh/zcqgh/200804/t20080421_17783.htm

Beijing Municipal Authority (BMA). (2009). Actions plan on cultural, technical and green transport, http://zw.bjjtw.gov.cn/xbjjt/xbjjt_xd/ckyw/

Zhili Liu et al. (2009). "Green transport practice in Beijing." *Proceedings of the Fifth Advanced Forum on Transportation of China.*

Xiaoming Liu (2009). "Advancing a Scientific Development of Transportation to Construct a Transit-Oriented Beijing." *Urban Transport of China*, 7(6), 1-7.

Anthony D May (2009). "European Experiences in Green Transportation Development." *Urban Transport of China*, 7(6), 17-22.

Xiaoming Liu et al. (2007). "Situation, Challenges and Strategies on Transportation of Beijing." *XXIIIrd World Road Congress*, Paris.

Liren Duan (2008). "Modern conception, ideas and measures of urban traffic administration." *Presentation File on speech to Transport Research Center Workshop*.

Shanghai Poly-Center Urban Planning and the Impact of Transport System

Haixiao Pan[1]

[1]Professor, Department of Urban Planning, Tongji University, Shanghai 200092, China; Email: hxpank@online.sh.cn

Abstract

Shanghai is one of the largest prosperous cities in China with the population over 17million. From the 1940's the major urban planning strategy is trying to de-centralize the overcrowded population in the central part of the city, where the density was even as high as 70000 people/sq.km, now decreased to 40000 people/sq.km. The still very high dense city contributes relative less motorized travel, as over 50% of people travel by foot and bike. But it also creates the problem of congestion in the city center, degenerated quality. And the rapid growth claims more space to accommodate more diversity and dynamic economic activities. The tendency shows Shanghai should transfer its spatial structure from Mono-center to Poly center. But the de-centralization strategy to encourage the people stay in the satellite town has not been success. Recently the municipal government has put great attention to support the secondary city in the suburb of Shanghai, according to the plan the most important three new towns will be with the population of one million. More and more industry has also been moved out to the suburb due to the lower cost for land and convenient for freight transport. In the year 2010, the world expo will be held in shanghai, some major infrastructure is now under-construction, which including the 400km metro system, and several the major passenger interchanges, all those will have a big influence on the spatial structure of Shanghai Region. The transformation of shanghai urban planning policy and the vision of the urban spatial structure will be analyzed in this paper.

Key Words: urban planning; poly-center urban expansion; transportation,

Introduction

Shanghai is one of the world's great cities, with a total area of 6,340.5 sq.km, including 523 sq.km for Pudong New Area.

Due to constant inflow of people from other parts of the country, the size of population in Shanghai keeps growing. The population of long-term residents reached 18 million.

Shanghai has witnessed marked progresses in its social and economic development. With the contrast to the economic development, the infrastructure was far lagged behind urban and industry development for a long time. Even to the year 1990 the road space per capita was 4.55 m^2, in the central area only 2.29 m^2. Long distance travel used to be very difficult, so that people prefer to stay in the city center where the service facility is in a walking distance.

It is noticed that one of the distinct characterizes of Shanghai is the high urban density and quite intensively mixed land-uses in the central build-up area. Mixture of factory within residential area, hospital and institutional activities caused serious conflicts due to air, noise pollution and the freight delivery service between the factories. For long time Shanghai was of shortage of living space. The urban environment degenerated seriously.

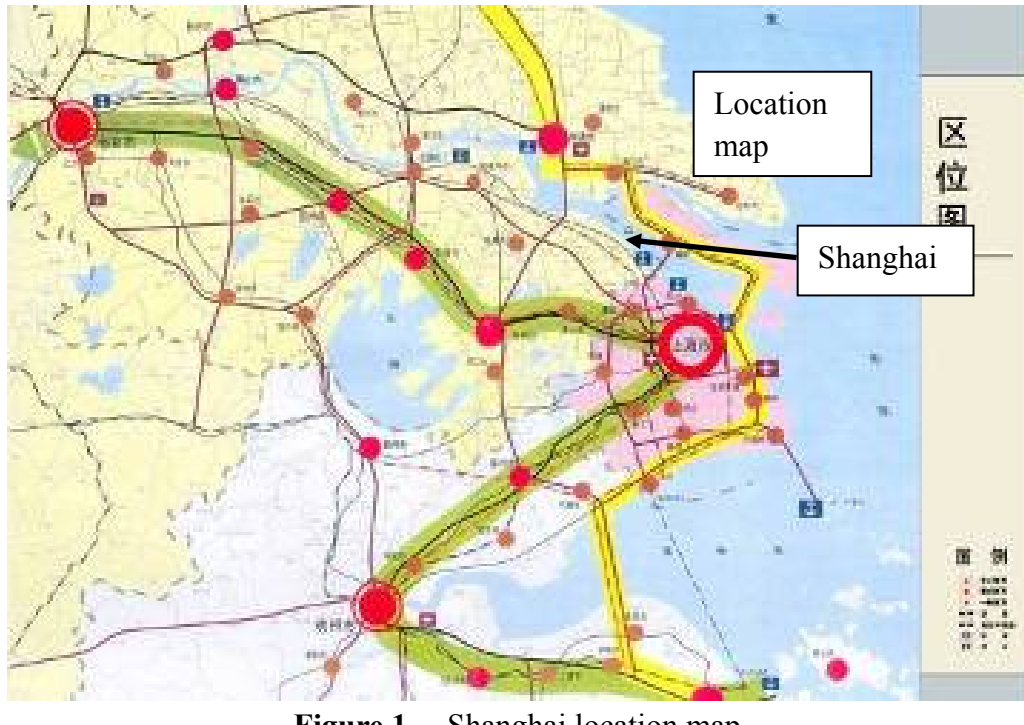


Figure 1. Shanghai location map.

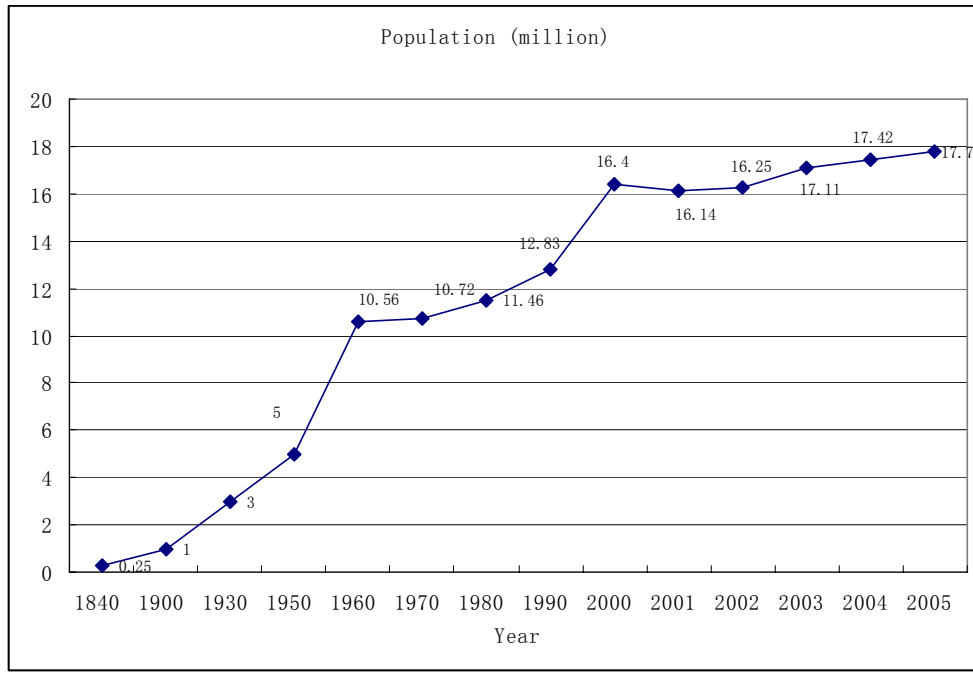


Figure. 2. Population in Shanghai.

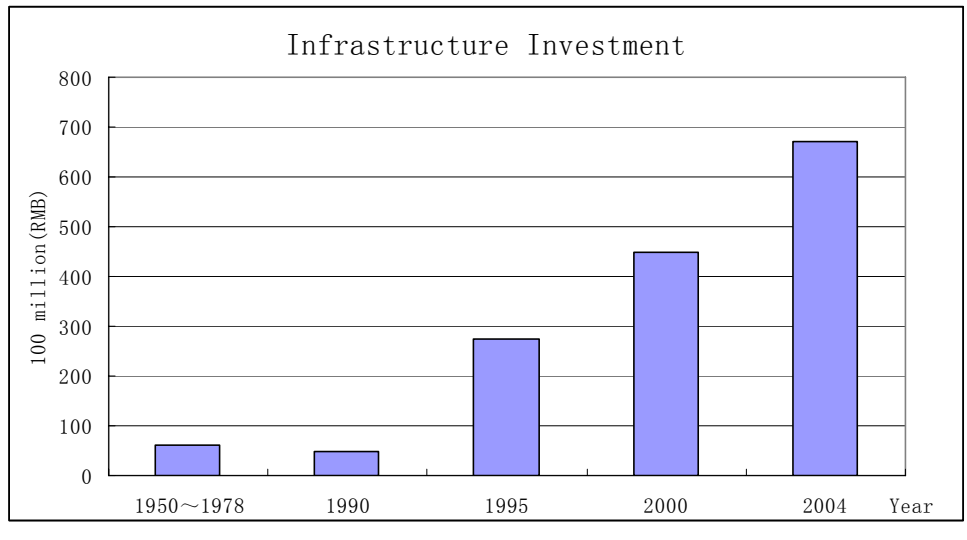


Figure.3. Infrastructure investment in Shanghai.

It is noticed that one of the distinct characterizes of Shanghai is the high urban density and quite intensively mixed land-uses in the central build-up area. Mixture of factory within residential area, hospital and institutional activities caused serious conflicts due to air, noise pollution and the freight delivery service between the factories. For long time Shanghai was of shortage of living space. The urban environment degenerated seriously.



Figure 4. Poor area still waiting for renovation.

From 1990 the population density of the whole metropolitan wide area is increased 484 to 2588 per square km, which is 3.1 times higher than Beijing, and 2.9 times higher than Tianjin.

From the middle of 1990's, the central area of the city has experienced a large scale urban renewal, land leasing, removing the manufacture industry and encourage the development of service industry. During the last five years of 1990's, 470 thousand households have been re-housed. It noticed a great drop of population in the most central districts. But it is still very high comparing with the new districts.

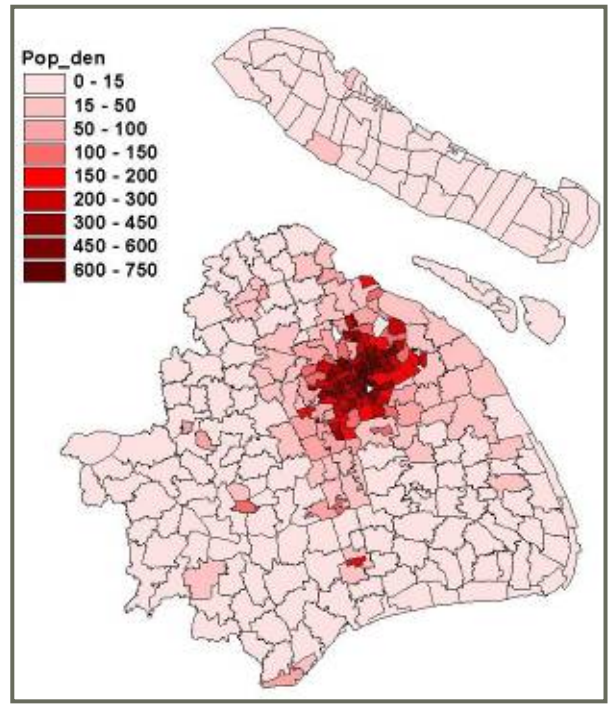


Figure 5. Population density in Shanghai (person/ha, 2000).

Table 1. Central area density in year 2000 and 1990 (Thousand, person/ km^2).

Area	2000			1990	
	Population move out	Population	Density	Population	Density
Huangpu	363.6	574.5	46 296	938.1	75 592
Luwang	146.9	328.9	40 859	475.8	59 105
Jingan	181.3	305.3	40 069	486.6	63 861
New Districts		7418.2	2 041		
County		2059.2	852		
Shanghai Wide		16407.7	2 588	13341.9	2 104

The high dense concrete urban fabric contributes to low automobile-mileage which benefits the low level of air pollution from the car. In the year 1990 43.9% of mechanized trips were made by bike. But the city also suffered the congestion, shortage of apartment for living and the lower performance efficiency, losing its advantage in competition to attract investment. For instant a travel speed survey has been taken in 1990, it showed that 34% of roads with the speed less than 10 km/h.

Comparing with the high density in the central area, land in the suburb is now well utilized. The gross population density drops to 30 person/ha in 15 km away from the city center(World Bank, 1994).

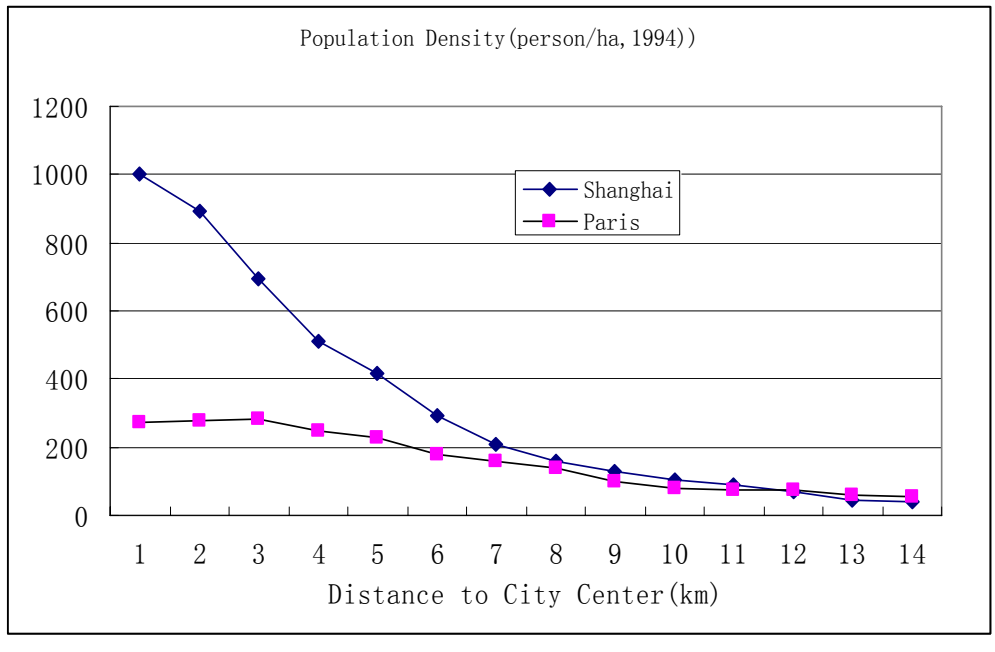


Figure 6. Density with distance.

Urban Planning for De-centralization before 1990

To accommodate the huge demand for dynamic economic activities and the fast increasing in inflow population, the expansion of the city is inevitable. In decades Shanghai has pursued the urban planning policy to transform from a super dense, mono-centric city to a multi-centric metropolis to decentralize its population and economic activities.

An Outline of master plan was made in 1959, in which the urban planning polices targeted at controlling the population growth in the central part of shanghai to not more than 3 million, the population in the periphery should be 1 million and there will be 2 million people staying in the seventeen satellite towns.
The green belt of one to four km wide was planned between the city and near suburb to avoid the city expanded to the suburb. The construction of high-rise apartment was encouraged.

Up to 1980's, Shanghai still faced many city problems, such as confusion in land-use, lack of infrastructure, traffic congestion and lack of residential space. In 1986<Shanghai Master Plan Outline>was approved by the State Council

In 1982, the area of Shanghai central city was 149 square kilometers. This plan put forward that total area of central city in year 2000 should be no more than 300 square kilometers. Furthermore, Shanghai should be a conurbation with convenient transport connection between central city and suburban towns. With the success of the Baoshan iron and steel plant and Jingshanwei petroleum chemical plant, the "two wing" became much more important.

Satellite towns, located 20~40km away from city center, should have comprehensive functions. There should be convenient connections of traffic and communication betweens central city and satellite towns, as well as each satellite towns. There should be high-capacity and high-speed passenger transport, which could reduce the commute time to no more than one hour between central city and satellite towns as well as two wings.

In 1990， the central government announced the open policy for Pudong New Area. <Pudong New Area Master Plan>has been prepared which defined the two development corridors. One is North-South along the east side of Huangpu River; another is east-west from Lujiazui (the financial district) to Zhangjiang (Hi-tech Park). Then several bridges, tunnel and metro across the Huangpu River have been constructed.

Table 2. Population in satellite town statistics of some shanghai satellite towns construction (1990).

Name	Primary industry	Distance to center	Land use(km^2)		Population (10 000)	
			Current	Plan	Current	Plan
Minghai	Machinery	32	20.7	60	10.65	50
Wujing	Chemical	25			2.52	
Jiading	Science and research	33	7.6	21.4	7	25
Anting	Car	40	5	16.7	2.7	15
Songjiang	Machine tool/light industry	40	7.3	20	8.2	25
Jinshanwei	Prtrol chemical	70	10	71	10	45
Baoshan	Iron and steel/port	20	44.8	105	24	50
	Total		95.4	294.1	65.07	21

Because of the poor level of infrastructure and living condition in the central part of Shanghai city, many urban constructions such as the lager scale urban renovation and infrastructure construction had been located in this area, between the year 1995 to 2000, 450thousand households have been relocated, 23million m^2 poor building has been demolished, instead of the planned satellite town. The improvement of urban living quality in central part of city, in return attract more people to stay.

From the Planning history in Shanghai, we can find that Shanghai pursued the multi-center and multi corridor spatial strategy for long time. But in fact during the fast development, the city became a bigger “Pancake”. But the performance of the satellite cities fell below expectations.

Recent Urban Planning Policy

In May 2001 the state council approved the new master plan <shanghai city master plan (1999-2020)>. According to the principle to coordinate the development of the city and its suburb, the spatial structure of Shanghai should be as following. The central city will be the principle part with multi corridor, multi-level and multi-core. From the statement in the plan we can expect the importance of development in suburb, where the new town should be the growth pillar and the service center for its surrounding area(Shanghai Government, 1999).

The growth of population should be concentrated in suburb new town where there are the major transport infrastructure of highway, see-port which will also facilitate the concentration of industry cluster to the industry park. The privilege should be given to the construction of new towns, which should be with relative perfect

function, good industry structure. The population of those new towns should be over 300 thousand, to realize the scale and accumulation effect, to attract people staying. The city wide transport plan also support the spatial structure, where each important town will be connected by Urban Rail (Shanghai Government, 2002).

In the 10th Five Year Plan of Shanghai Government, the planning of Songjinag New Town is part of Shanghai's suburban construction program for "one city and nine town". It covers an area of 40 km^2 located 4okm from Shanghai city center adjacent to the old town of Songjiang and acts as the southwest gateway to Shanghai city.

To collect the concept for New Town development, a competition was launched. Finally a Britain Company won the competition. The key concept elements of the proposal were quite attractive with a high density central area or hub that contains principal public transport, civic service, commercial and retail facilities well connected to the rest of city by a network of transport linkage, the clearing segregation of land use, Central business district, the cultural quarter, industry quarter etc.

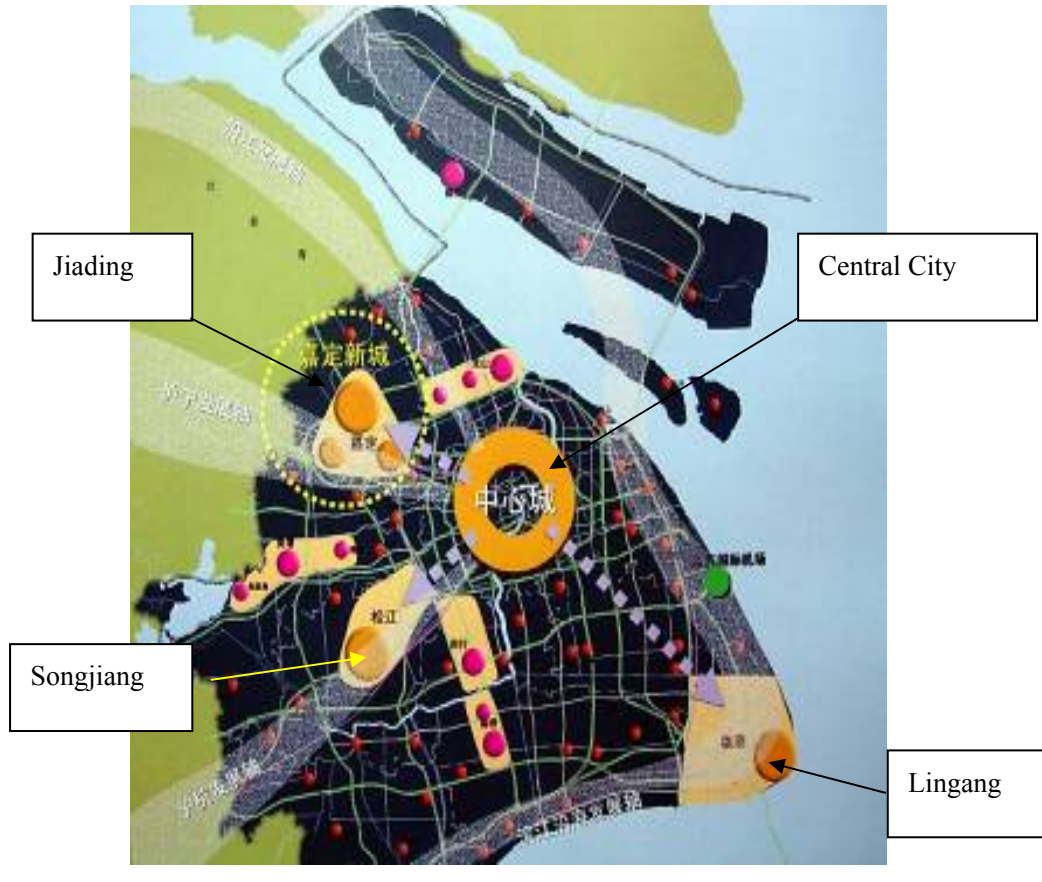


Figure 7. Major development corridor in Shanghai region.



Figure.8. The urban rail plan and sub-city in Shanghai region.

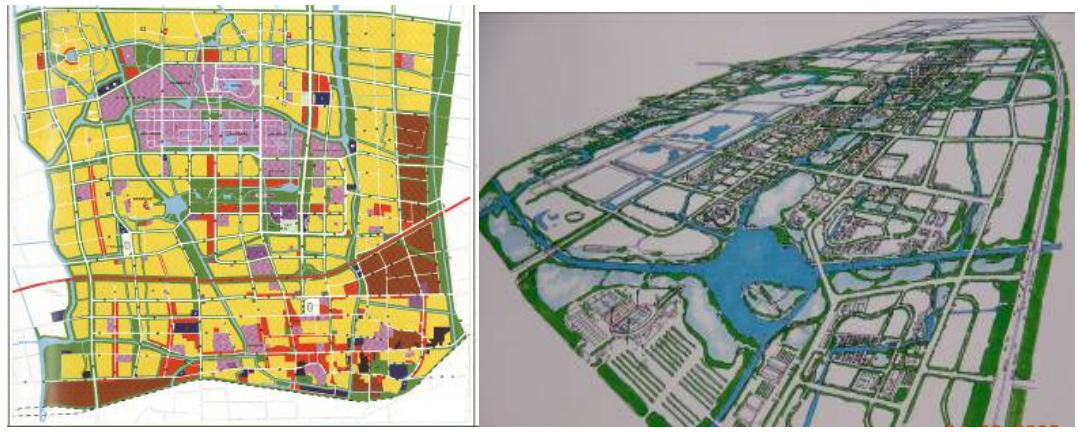


Figure. 9. Songjiang master plan in 2002 (60sq.km)

There are also one English style residential area with a church , called Thames Town, an area of 1 square kilometer in Songjiang New Town, is Located between the residential districts in the west and the central business district of the New City. With a single extended stretch of multi-functional path and the waterfront square serving as the main axis of the overall design, various public facilities like commercial facilities, clinic, hospital, supermarket and church are effectively congregated and buildings like detached houses, townhouses and multi-storey apartments accommodated within the one square kilometer area. Thames Town will offer a different home to its 8,000 residents.



Figure 10. The bird view of the Thames town in Songjiang.



Figure 11. Landscape of Songjiang old town.

Although the built environment is in high quality with wide road, spacious of green space, comparing with the Songjiang old town where the street is narrow, crowded with buildings, lack of green area but full of people, full of activities, the New town goes to another extreme, much more automobile dependence, less of living.

The Impact of 2010 Shanghai World Expo

It is estimated that there will be more than 70 million visitors to the Expo site, which located just beside the city center, in half year. Because the transport in city center is already saturated, any small increasing in the travel demand will induce a big drop in the performance. So that the improvement in urban transport will be give first priority. It is expected that large amount of investment will be put on transport infrastructure, both within city or between cities.

Huge amount of visitor, urban infrastructure especially for urban transport will have a fundamental development. In addition to the program for urban rail transportation, Yangshan Harbor phase I, the Pudong Airport phase II, Chongming linkage, a large

number of major infrastructure projects will further improve the city Urban infrastructure, greatly improving the city's traffic and transport capacity, and the public will travel more speedy and comfortable. It can be said that the Expo will be held in Shanghai will make Shanghai a world-class city. The impact of the Expo on Shanghai industry adjustment, infrastructure construction should not be underestimated.

Expo factors will further accelerate migration to the suburbs. Holding the Expo will make marked improvement in urban infrastructure, especially the central area and the suburbs, as well as transport links between Shanghai and its neighboring cities, which will be more convenient for suburban development. The transport constrains will be greatly eliminated. This factor will stimulate region's real estate development along transport corridor, and to further promote migration to suburb town. To avoid the undisciplined expansion in the previous, we need to control and guide the rational planning to form a moderate gathering, the more open and integrated spatial pattern.

Comments and Suggestions

With the development gradually moved from center city to the suburbs, and how urban planning, development policy can effectively guide the process, and creating a rational layout of urban space and promote the development of urbanization in the suburbs, has become an increasingly important issue. Therefore, the following three issues will take into consideration:

(1) With development emphasizing move to suburb area, the priority should be given to fast transport linkage between the city and the suburb, particularly the fast public transport, to speed up infrastructure construction in rural towns. Previous experience shows that transport has a great effect on the spread of industry and population from the city center to suburb. Improve rural infrastructure facilities; in particular to improve the urban transport infrastructure in the rural areas is to enhance the attractiveness of population and industry to suburb town.

(2) The theoretical model only provides the fundamental concept of the future; the blueprint proposal will always face difficulties in implementation later on. Local culture and social facts must be taken into the consideration to develop the regional spatial strategy. Encouraging the development in suburb town may also induce the long distance travel, if without the qualified public transport service, may bring about serious congestion, even the self-contained policy can not limit the freedom of people moving around

(3) Further enhance the strategy to relocate the distribution of population; a

rapid/convenient transport networks and industrial park are the two key facts in creating a growth pillar for attract living, integrated business functions to support town's growth. The design of the satellite town should be multi-function to avoid the only bed town. Some service and business activities should also be encouraged to locate in the suburb town, serving as the secondary service center to towns across over Shanghai region. It is great important to adjust the dislocation of residence and employment.

References

Shanghai Government. (1999). *Shanghai Master Plan* 1999–2020.

Shanghai Government. (2002). *Shanghai Urban Transport White Paper*, Shanghai People Press.

World Bank. (1994). *China*: *Land Management in Emerging Market*, China Finance Economy Press.

Guangzhou Comprehensive Transportation Development Strategy (2010-2020)

Xiaoyi Ma[1]

[1]Chief Engineer, Transportation Planning Research Institute of Guangzhou, 10/F, 80 Ji Xiang Road, 510030, Email: pow2006@163.com.

Abstract

Review the 30 years of the Guangzhou transportation development course. Combining the requirements for transportation from urban master planning, it prospects the development opportunity and challenge facing the city from the urban primacy index and rail transit function orientation. On the basis of analyzing deeply the idea of human-oriented, it makes clear the key work to be done in next 10 years. Strengthening the pivot position (external) and constructing a public transit city (internal) are the two basic tasks. It has made "124" Comprehensive the transportation development strategy. The strategic objective is constructing a harmonious transportation system, which is sharing, fluent, intensive, green, and fair. It has made two core policies, which were transport facilities leadership and public transport priority. It introduced the strategy idea briefly from four respects, such as integration, differentiation, intensification, and refinement.

Key Words: transportation planning; **c**omprehensive transportation; opportunity; challenge; development strategy

Introduction

December 2008 was a critical period when the reform and development of Guangdong Province faced a new situation and challenge. More importantly, it was on a historic node that marked China's 30th anniversary of initiation of reform and opening-up. This month also saw the consideration and principle passing of a document on an executive meeting of the State Council -- the "Reform and

Development Outline of Pearl River Delta Region (2008-2020)" (hereinafter referred to as the "Outline"). The Outline expressly requires Guangzhou to "exploit its advantage as a provincial capital city, enhance its position as one of China's central cities and comprehensive portal cities, improve its radiating and propelling influence, and lead the Pearl River Delta Region to build a closely integrated city cluster with a reasonable layout and complete functionality (Ma et al., 2009)."

With this new objective and positioning, Guangzhou has rare opportunities and challenges of transport development, and identified tasks for the decade following the 16th Asian Games to be held in the city in 2010. In order to meet the objective and requirements defined in the Outline, Guangzhou has to undertake its historical mission and give full play to its leadership in the Pearl River Delta (PRD) region. In doing so, the city should not only review the demand and its gap therefrom, but also conceive a comprehensive strategy of transport development for the next decade from two aspects: the region in which it is located and the territory it encircles. This will help to enhance its position as one of China's central cities (Wang et al. 2009).

A Three-Decade Review

Major events play an undoubtedly significant role in driving the transport infrastructures of any city. This is also the case in Guangzhou, where each of its three development stages in the past three decades was closely related to its hosting of a major sports event.

1980s – the 6th National Sports Games of China (1987). As one of China's pilot cities of reform and opening-up, Guangzhou began to convert from a planned economy system and became aware of its market economy at the beginning of 1980s. In addition to bringing the city's economy to a fast track for the first time, this also triggered the increase of transport demand. Guangzhou, as the political, economic, and cultural center of Guangdong Province, and the economic center and transport hub of South China region, found its outdated transport facilities could hardly meet such an increase. On this stage, Guangzhou's road transport was primarily characterized by (Dai and He, 1999) :

- Public transport was continuously shrinking, contrasting with the continuous growth of private transport. This resulted in an extremely unreasonable structure of transport modes. Although motorcycles increased by 4.7 times, from 36,000 to 208,000, and bicycles by 90%, from 1,476,000 to 2,800,000, the public transport passengers dropped by 29%; and
- The road supply facilities were too underdeveloped to accommodate the increasing road traffic. As a result, vehicles were slowed down and

traffic delays increased. In the urban center, the average speed was only 10-15 km/h for all vehicles and 6.5-8 km/h for those of public transport.

To address urban transport problems and support for the 6th National Sports Games of China to be held in the city, Guangzhou fruitfully explored the transport engineering theory and used it in guiding many of its practices like building crossovers, constructing elevated roads in the urban center, and managing the traffic. Helped by the hosting of the Games, Guangzhou moved its center eastward from the old area to Tianhe District. The construction of this new district released the traffic stress in the old area and propelled the transport development of the city.

1990s – the 9th National Sports Games of China (2001). After two decades of reform and opening-up, Guangzhou had become a more powerful hub having more influence in South China. With a GDP growth rate kept at about 13%, the city had long been one of the Top 3 cities in China. However, its urban environment and transport infrastructures had never caught up with its development. The weak transport infrastructures had made traffic jams a persistent topic of criticism and seriously restricted the social and economic development, as well as the international competitiveness, of the city. On this stage, Guangzhou's road transport was primarily characterized by (Pan et al., 2006):

- Motorcycles rapidly increased to nearly 1 million at the maximum, accounting for 60% of total volume of motor vehicles in possession. This not only irrationalized the structure of transport modes, but also made the traffic disorderly; and
- The transport system was battered by weak infrastructures and insufficient capacity.

To address the ever-increasing congestion of urban transport, Guangzhou abandoned its "firefighter's role" against traffic problems. Urban transport was strategically studied in programs like "Study on Guangzhou Transport Plan" and "Guangzhou's Implementation Plan to Improve Transport at Urban Center." In 1998, it even defined an ambitious objective of "a small change a year builds a medium change every three years and a big change by 2010" for urban planning, construction, and management, just at the right time before the 9th National Sports Games of China would be hosted in the city. Accordingly, RMB62.7 billion was invested in the city's infrastructural constructions. When these constructions were completed by the end of 2001, Guangzhou saw a "leap forward" in city construction, and traffic congestion was virtually removed, thanks to the birth of some important trunk roads and road facilities, an urban

transport network, and the longed-for subways (Ma 2004).

2000s – the 16th Asian Games (2010.) On July 1, 2004, Guangzhou was selected as the host of the 16th Asian Games 2010, the third non-capital city following Hiroshima and Busan. Although the city values the Games as a rare opportunity for urban construction and transport development, the city also recognizes a huge challenge that the Games will bring -- a surge to the ever-increasing transport demand of the city (Ma 2008). On this stage, Guangzhou's road transport was primarily characterized by:

- Its external transport hub is hard to match its position as a regional center.
- Cars have seen a rapid increase since motorcycles were banned in the urban center at the end of 2006. By the end of 2009, they had approached 1.1 million, 17% more than in 2008.

On this stage, Guangzhou has grasped advanced technological tools to plan for its transport. In addition to creating a mechanism to track the transport dynamically, the city begun to compile annual reports of transport development and update the traffic model on a regular basis. Based on a residential travel investigation in 2005, it made geographic information directly and completely accessible to its transport planning model. Moreover, Guangzhou has built many infrastructures for external transport, like a new airport, sea port, and railway passenger station, as guided by a superior transport plan. This led to a "leap forward" development of its regional transport hubs, and enhanced its position as a regional transport center. In addition, the construction of many new expressways, highways, and rails has given birth to the backbone (rapid rail transit system and expressway / highway system) of Guangzhou's urban transport system. In addition to inducing the adjustment of Guangzhou's layout of city functions, this also supports the networked expansion of Guangzhou's spatial structure consisting of multiple centers and several groups. By 2010, 80% of the Asian Games venues can be reached by subways and suburban railways, 85% will be on expressways and highways, and each of other venues can be accessed through two or more primary arterials.

Opportunities and Challenges

Since the 21st century was unveiled, Guangzhou, like all other Chinese metropolitans, has been plagued with traffic congestion, but more mildly than in Beijing and Shanghai, thanks to, among others, two noteworthy factors.

Primacy Index. Although Guangzhou is the primate city in the PRD region, its primacy has gradually lagged behind that of Beijing and Shanghai when the region

sees rapid growth (refer to Table 1 for detailed data). This is why it has been saved from the excessive stress of concentrated traffic. However, although one of the identified central cities of China, Guangzhou will be much more stressed by the external traffic as it's more closely related with other PRD cities and thereby becomes more primatial.

Table 1. Comparison of Primacy index among Beijing, Shanghai, and Guangzhou in 2008.

Categories	Primate City	Secondary City	Primacy index
Population Scale	Beijing (1695)	Tianjin (1176)	1.44
(X 10,000)	Shanghai (1889)	Suzhou (882)	2.14
	Guangzhou (1005)	Shenzhen (877)	1.15
GDP (X RMB0.1	Beijing (10488)	Tianjin (6354)	1.65
billion)	Shanghai (13698)	Suzhou (6701)	2.04
	Guangzhou (8216)	Shenzhen (7807)	1.05

Positioning of rail transit Although Guangzhou, Beijing, and Shanghai all have a long history of massive rail transit in China, their different ticketing policy has objectively caused different positioning of rail transit. In Guangzhou, the relatively high fare of rail transit (refer to Table 2) has suppressed the growth of cars to a certain extent, as proven by Guangzhou's residential travel investigation in 2005 (refer to Table 3 and 4). However, as the ground transport is more stressed, the pricing will have less effect on the positioning difference.

Table 2. Comparison of public transport indicators in Beijing, Shanghai, and Guangzhou at the end of 2008.

Indicators	Beijing	Shanghai	Guangzhou
Ticket system	In-station free transfer	Distance-based	Distance-based
Disposable income per resident	1.00	1.00	1.00
Relative subway fare	0.50	0.67	1.00
Passenger flow per km of subway	1.25	1.00	1.00
Passenger flow ratio of public transport to subway	1.08	1.03	1.00

Table 3. Guangzhou's rail passengers vs. educational background of its residents (He and Ma 2008).

Educational Background	Primary School	Junior Middle School	Senior Middle School and Polytechnic School	Junior College and Higher	Total
By Rail	14%	15%	30%	41%	100%
Full Sample	6%	46%	29%	19%	100%

Table 4. Guangzhou's rail passengers vs. income level of its residents.

Income Levels	No Income*	<4,000	4,000-10,000	10,000-25,000	25,000-50,000	>50,000	Total
By rail	22.2%	2.5%	10.5%	36.9%	27.9%	0.0%	100.0%
Full sample	18.8%	7.1%	21.7%	33.9%	14.6%	3.9%	100.0%

The high percentage of no-income rail passengers is because many retired cities enjoy their free rail ride.

Identifying the Policy Focus

As a concept to address urban transport problems, "people orientation and priority of public transport" has enjoyed the public support for more than a decade. However, "car traveler-oriented" policies have been followed in almost all Chinese cities since the mid 1990s and will continue in some cities for some time, because the limited urban transport resources have to favor travelers of a certain transport mode in a certain time period. Moreover, it has not been identified who should be the people to be oriented around. This has resulted in a conceptual confusion of transport. Although a "vehicle orientation" concept has been tried for interpretation by some experts. Such a concept is inherently debatable because it is man who has its hands on a steering wheel.

A view on the generalized value will help to get an understanding.

$$U = \sum_i q_i v_i$$

Where: U = Total travel value in a city; i = A certain travel mode (passenger cars, public transit, and non-motorized vehicles); q_i = Number of travelers in a certain travel mode; v_i = Average composite travel value of travelers in a certain travel mode.

In certain time periods, decision-makers of a certain regions may follow a "car traveler-oriented" policy because they believe that the total value of car travelers exceeds that of any other mode. Therefore, the "people" to be oriented around should be the travelers in a certain transport mode and must come with a definition of time and space limits. In Guangzhou, for example, the road has been made very slow by the great number of private cars for family use. Therefore, the total value of car travelers is obviously not as good as that of public transport travelers, which also outperforms that of non-motorized vehicle travelers because the city has to maintain a certain speed of motorization to keep its fast economic growth. Therefore, Guangzhou's transport policy should be oriented about public transport travelers

from 2010 to 2020.

Development Strategy

Strategic Objectives. With a conceptual vision of "sharing, fluency, intensiveness, green, and fair" (Ma et al. 2007), Guangzhou should build an integrated, multi-mode "harmonious transport system" to propel the sustainable development of economy, society, and environment in Guangzhou and even in the Pan-PRD region.

The "harmonious transport system" should be directed to enhance the regional radiating influence and cooperation, and harmonize the transport and land utilization. It should be backed by major regional transport infrastructures and urban road and rail networks to be developed at a faster pace. Motorization should be developed in a sustainable manner by focusing on public transport, which should be supplemented by motorized private transport. The public transport system should be built to include multiple modes in harmony with each other: rail transit as the backbone, conventional public transport as the main part, and taxies and water passenger transport as the supplement. All these are to meet Guangzhou's development requirements to build itself a modernized metropolitan suitable for starting and doing business and for living.

Core policies

Led by transport facilities. Guangzhou should implement a guiding ideology that requires "the transport infrastructures to lead integrated development of the PRD region." It should emphasize the openness and service functions of its regional transport system to help shape the one-hour PRD megalopolis. The city should also build a modern system of transport infrastructures that matches the integrated PRD development. Particularly, this includes:

- Priority should be given to developing moderately proactive regional major transport infrastructures to be co-built and shared. Regional transport infrastructures like Baiyun International Airport, Nansha Port, and Guangzhou South Railway Station should be brought into full play to lead and promote the adjustment of Guangzhou's spatial layout, as well as Guangzhou's integrated and harmonious development with transport systems in surrounding cities. This will drive the co-development in the PRD region, Guangdong Province, and even the Pan-PRD region.
- The service target and radiating scope of regional transport hubs should be identified. Accordingly, Guangzhou should keep sufficient reserve for future development and connectivity. The relationship among regional airports, ports, and railway and road hubs should be coordinated. This will intensify

the connectivity between external transport means and between external transport and the city. Guangzhou should also aggressively build an inter-city transport infrastructure. Moreover, the radiating scope of regional transport hubs should be intensified to help shape an integrated transport pattern across the region.

- A mechanism to coordinate regional transport development should be built on different levels - PRD region, Guangdong Province, and Pan-PRD region. A leading committee should be established to lead the regional transport planning and construction, and thereby continue to deepen the regional transport connection plan and propel the construction of Guangzhous transport connectivity facilities with Foshan and Dongguan.

Priority of public transport

Guangzhou should implement a "public transport traveler-oriented" guiding ideology and a "priority of public transport" policy on the present stage. Accordingly, priority of law, investment, tax, plan, construction, operation, and management should be given to developing public transport, enabling the latter to access reasonable land and funds, operate efficiently, and provide convenient transfer. Particularly, this includes:

- Prioritized infrastructural construction: Rail transit, the main backbone of urban passenger transport, should be built more quickly. Passenger transport service should be made more efficient by building more supporting facilities for transport connectivity, especially passenger transport hubs. For the development of public transport, land should be reasonably and gradually made available by, among others, constructing bus depots. Especially, those depots supporting large residential communities and public buildings should be perfected.
- Prioritized adjustment of network structure: The public transport plan should be developed and implemented based on a principle of "scientific planning and dynamic adjustment." The structure and capacity deployment of a bus (trolley) network should be adjusted whenever necessary, by following a guideline of "optimizing public transport network in the main urban area and filling the public transport vacuum in peripherals." This will further increase the network density and station coverage and shape a well-structured public transport system in which trunk and branch lines play in concert.
- Prioritized provision of operational right-of-way: A signal system should be scientifically deployed for bus-only lanes (roads) and bus-priority access, in order to improve the supervision and management on the measures that give right-of-way priority to the operation of public transport. An intelligent public transport system should be pushed forward, providing the public with high-quality, wide-coverage, massive, and rapid public transport service.

- Prioritized supporting measures: The public transport industry should receive more support of finance, tax, and policy. A system to allocate land for the use of public transport should be enforced. To compensate for policy-related loss that public transport enterprises may suffer as a result of undertaking social welfare tasks, financial subsidies should be provided. In addition, more resources should be invested in the scientific research for public transport development. A system of laws, regulations, and standards should be established and perfected. Moreover, the public transport industry should be more reformed by optimizing and reorganizing public transport enterprises through market means, encouraging their scale development, and keeping moderate competition on the public transport market.

Development Strategies

Servicing not only Guangzhou but also China (Ma et al. 2009) -- "integrated" transport strategy. Guangzhou should improve its comprehensive bearing capacity in an all-around way. Aiming at world-class core cities of megalopolis, it should intensify its core city function of "radiation, service, drive, and lead" to the one-hour PRD megalopolis. In doing so, it should not only consolidate its existing advantages, but also foster new and high-tech development domains. This will help it lead the region to acquire its place on the worldwide stage.

An open, comprehensive, and modern transport system should be built as guided by a development concept of "regional share" and "building a city-like PRD." The system should be spearheaded by three world-class hubs of Baiyun Airport, Nansha Port, and New Railway Passenger Station, and supported by a highway network and an inter-city rail transit network that connect Guangzhou to its surroundings. It will help Guangzhou become China's southern transport center, which will service the whole country, face the world from PRD, and intensify Guangzhou's function as a portal region.

From urban to rural – "differentiated" transport strategy. Different regions and groups should be sophisticatedly considered in terms of their traveling characteristics of residents and vehicles, and the characteristics of traffic flow, vehicle speed, and spatial variation of traffic volume. Based on such consideration, the transport policy and guidelines suitable to local conditions of corresponding spatial areas should be defined, in support of the urban layout that features functional separation. The following paragraphs provide detailed descriptions:

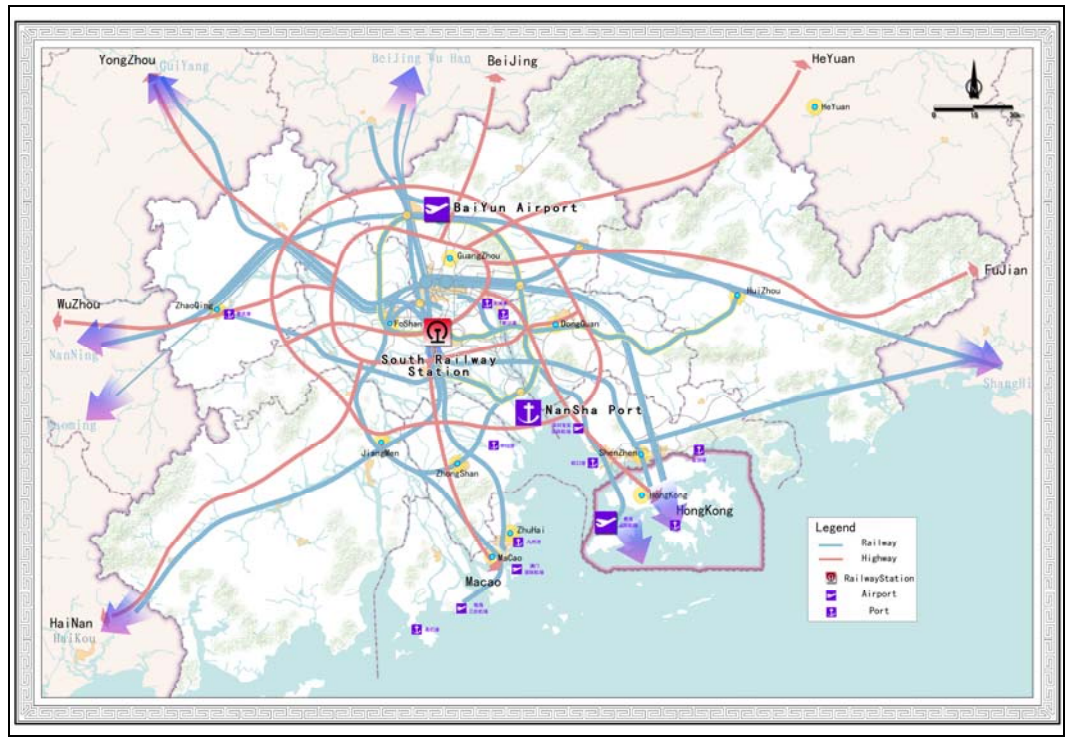


Figure 1. A regional network pattern of highways and rapid rails.

- Central region: To fit in with the style and features of the central region, where there are historic and cultural heritages being protected, the public transport system should be developed with rail transit as the emphasis and main part. The transport function of arcade blocks, unique facilities in the Lingnan region (a region in Guangdong province and Guangxi Zhuang Autonomous Region of China), should be appropriately utilized to create a superior environment for pedestrians. The use of cars should be strictly controlled through various manners, in order to keep an acceptable service level for road transport operation and maintain vigorous transport in the core area.
- Northern, Eastern, and Southern regions: The development priority should be a rapid public transport system consisting of rail transit and BRT (Bus Rapid Transit), and the convenient transfer between different public transport vehicles and between public transport and other transport modes. The quantity, as well as the usage level and scope of cars should be regulated through various measures that promote public transport.
- Conghua, Zengcheng, and Nansha regions: Different structures of transport modes should be built depending on the land-use characteristics in different areas. In addition, land development should be oriented about public transport. In this TOD (Transit-Oriented Development) mode, intensive, dense development should be conducted in central areas of new towns to

attract people to public transport. In general peripherals, car users should see relatively loose regulation. Accordingly, dense road networks and more sufficient parking spaces should be available. In addition, new peripheral towns should be supported by improved public service functions to minimize their reliance on the main urban area and to reduce the tidal traffic.

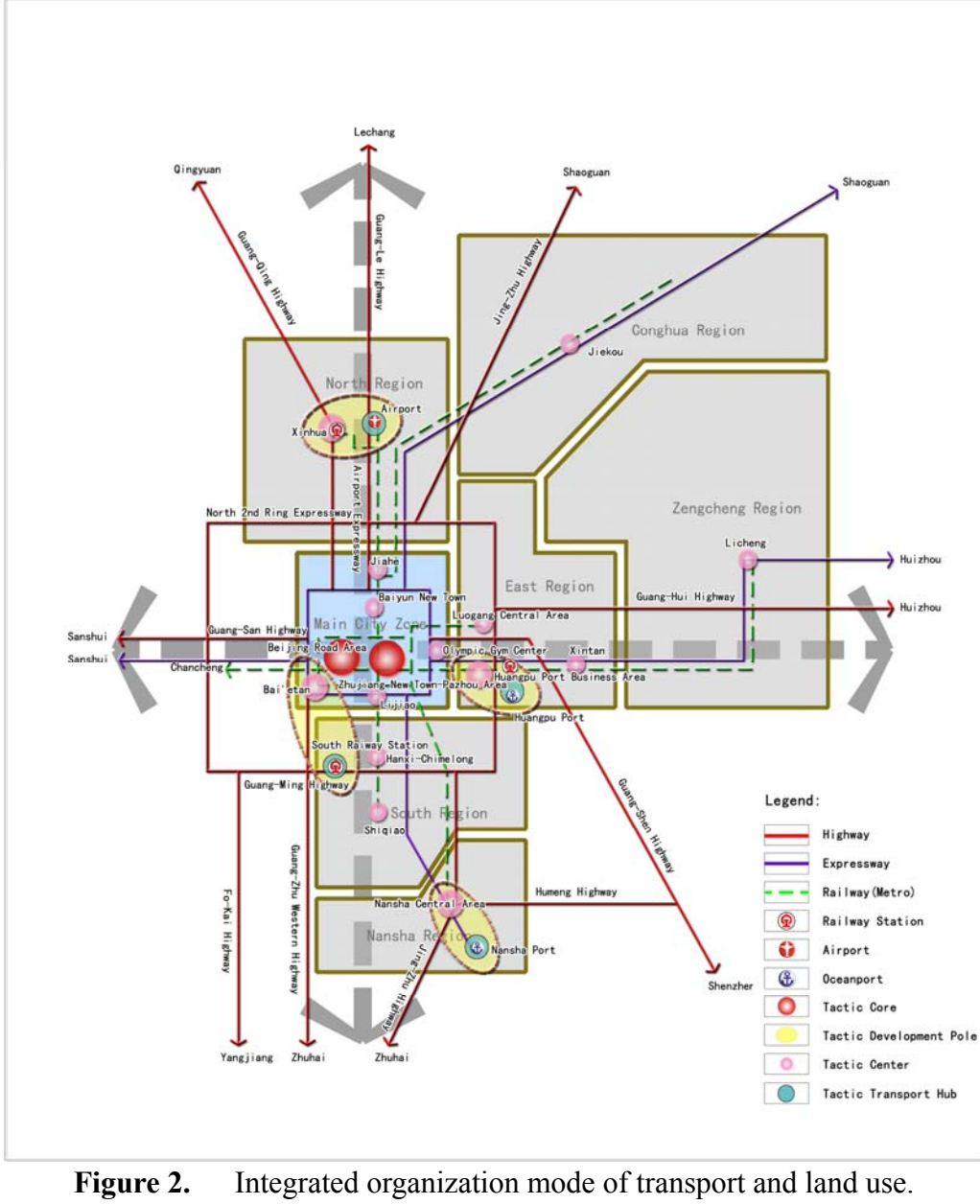


Figure 2. Integrated organization mode of transport and land use.

From roads to rails – "intensive" transport strategy. Constrained by the bearing capacity of resources and the sensitivity of environment, the "Rapid Rail Network" should be focused on providing public transport among cities, in the urban territory, and in the main urban area, and the "Highway Network" should more service private

and freight transport. This is to meet the objective of sustainable development and to implement the guidelines of "priority on public transport" and a "transit-oriented" spatial development strategy. It will suit Guangzhou's spatial layout that consists of "one main region and six suburban regions," and help build an intensive system of transport corridors and hubs.

From construction to management – "refined" transport strategy. In order to optimize the overall strategy of urban development and improve the functionality of the city, transport development should see a principle conversion. It should not focus on infrastructural construction anymore, but to put equal emphasis on construction and management. And at the urban center, management should become the focus. All kinds of transport facilities in the urban area should be more inspected and maintained on a routine basis, so that their efficiency and level can be improved. To maintain vigorous transport despite of concentrated traffic demand at the urban center, different kinds of demand management measures should be gradually rolled out. In the Central region, the development groups should be classified into three strategic zones, namely, the core group, peripheral group, and ecological group, based on the location, environment, and transport characteristics. Different transport policies and guidelines should be defined as appropriately for each group, in order to create a convenient transport connectivity between groups which functionally supplement each other, as well as unique transport spaces within each group.

Conclusion

Comprehensive transport plan is an important part of Guangzhou's master plan, which is being redefined, and a programmatic document that will propel the construction and management of urban transport in the city. Therefore, Guangzhou must have a global and strategic conception on the construction and development of its transport infrastructures, and scientifically plan and deploy its major transport infrastructures in the near and long term. Moreover, systematic solutions should be developed to address existing transport problems. The existing transport plans and facilities should be integrated to each other and all transport modes should be developed harmoniously. This will be the only way for the city to build a "harmonious transport system" which is "shared, fluent, intensive, green, and fair" in support for this central city of China.

References

Dai, C. H. (1999). "An exploration on the development path of urban transport." *Annual Conference Symposium of China Academic Committee of Urban Transport Planning*, 1-8.

He, C., and Ma, X. (2008). "Impact of ticket transaction policy on URT passenger volume in Guangzhou." *Urban Mass Transit*, page numbers?

Ma, X. (2009). "Recommendations for planning the integrated transport network of Pearl River Delta Region around Guangzhou." *Guangzhou Transport Planning Research Institute*.

Ma, X. (2004). "Enlightenment on the variations in the structure of Guangzhou inhabitant mode share." *Urban Transport of China*. February, 29-32.

Ma, X. (2008). "Transportation development strategic plan about 2010 Guangzhou Asian Games." *Journal of Transportation Systems Engineering and Information Technology*, August, 16-17.

Ma, X. (2007). "Strategic planning of urban transport development in Guangzhou." Guangzhou Transport Planning Research Institute.

Ma, X., and Wang, B. (2009). *Integrated Transport Plan of Guangzhou (Phase 1)*. Guangzhou Transport Planning Research Institute.

Pan, A., and Zhou, H. (2006). *The Path of Urban Transport*. China Building Industry Press, 74-75.

Wang, D. (2009). *Strategic Planning of Guangzhou Urban Master Plan*. Guangzhou Bureau of Urban Planning.

Development in Market Structure and Regulation of Urban Bus Services in Developing Countries and Countries in Transition

Brendan Finn[1] and Corinne Mulley[2]

[1] Senior Consultant, ETTS Ltd., 24 Marley Court, Dublin 14, Ireland
[2] Professor of Public Transport, ITLS, University of Sydney, NSW, Australia

Abstract

The market for urban passenger transport has experienced major change in many developing countries in Africa, Asia, and the Middle East, as well as in countries of political and/or economic transition in the Commonwealth of Independent States (CIS) and China. Such changes have included planned market opening to private operators and new entrants; unplanned market opening by the entry of unlicenced operators; privatisation and other changes to the ownership base of large public-sector transport companies; emergence of large-scale minibus and paratransit; national and urban policies and programs to upgrade the transport supply and quality. This paper presents relevant developments in Africa, Asia, Middle East, the CIS and China, and clusters them according to (i) changes in the role of the regulator and market structure; (ii) changes in the structure of the operator and of private sector participation; and (iii) changes in the transport supply. The paper then identifies critical factors leading to change in the urban transport sector, factors which can be identified with successful outcomes and issues associated with the development of the minibus, paratransit and informal sector which have played major roles in the urban transport sector of developing countries and countries in transition.
Key Words: urban transport; public transport; bus services; regulation; market development; developing countries

Context

The mobility of people is fundamental to their ability to participate in society. In developed countries, mobility in urban areas is assured for the majority of the population, in part due to high levels of car ownership, and in part due to stable public passenger transport systems. In the urban areas of most developed countries, public transport is provided by public or private operators who have medium- to long-term agreements with a transport authority for a defined set of services,

enforced legal protection against interlopers, and are usually subsidised to cover any losses due to the social dimensions of the services. The situation in developing countries and countries in transition is more varied. For working purposes in this paper, 'developing countries' are considered as those with a low average income and a low degree of industrialisation. 'Countries in transition' is used to include countries in which there have been major political and market shifts (e.g. former Soviet Union countries), changes to the structure of the economy (e.g. China), or fundamental change in the demographic or settlement patterns (e.g. India's and China's migration to urban areas), and where institutions, markets, laws and other frameworks are still in the process of adapting. Some countries can be both developing and in transition.

By their very nature, these countries (which account for the majority of the global population) face diverse challenges which involve change to society, economic and legal frameworks and relationships between these. Further, many of these countries lack the financial resources, and sometimes the political stability, to implement effective and efficient mobility services at affordable prices which meet the expectations of citizens. These challenges are most concentrated in the urban areas and their hinterlands due to the scale of population and the intensity of activity (although this is not to underestimate the sometimes extreme problems caused by lack of mobility for individuals and communities in rural areas).

The market for urban passenger transport has experienced major change in many developing countries in Africa, Asia, and the Middle East for a variety of economic, political and societal reasons; and due to fundamental political and economic transition in CIS and China. Such changes have included planned opening of the market to private operators and/or new entrants; unplanned opening of the market by the entry of unlicenced operators, especially where the licenced services become inadequate; privatisation and other changes to the ownership base of large public-sector transport companies; emergence of large-scale minibus and paratransit; national and urban policies and programs to upgrade the transport supply and quality. This paper concentrates on developments in the urban passenger transport sector in developing countries and countries in transition that have recently occurred, or which in 2010 are currently taking place. These changes impact on both the mobility of the citizens, and on the ability of the authorities and other stakeholders to influence the mobility services which are available.

For the purposes of this paper, the urban passenger transport sector can be considered to consist of three strands:

- The agencies that manage the market for the supply of transport services, which are referred to as 'the regulator' (this includes transport authorities);
- The entities of public and private form that operate the transport services, which are referred to as 'the operator';
- The transport services.

Three main types of development are observed, which are the focus of this paper:

- Changes to the market structure and the role of the regulator, and the basis of the relationships between the regulator and the operator;

- Changes to nature, format, ownership and structure the transport operators and means of participation of the private sector;
- Changes to type, structure, quality and scale of passenger transport services.

This paper focuses on positive developments (A good analysis of failings and what to avoid is provided by Gwilliam, 2005): the approach has been to identify relevant developments from available published sources, supplemented by direct observation. Since most of the developments are ongoing and are themselves changing, the focus has been on contemporary sources. Following an introduction to international trends in market structures and the role of the regulator, with emphasis on those in more mature systems, each of the three strands is considered in turn. The final section discusses issues arising from the presented developments, and key factors and/or enablers for evolution of an effective passenger transportation system.

International Trends in Market Structures and Regulatory Frameworks

Over the past two decades, there has been a very clear trend towards a structured market for the provision of urban public transport (This section is derived from more extensive reviews (Finn and Nelson, 2003 a,b)). The main features include:

- restructuring Institutional Frameworks as a coherent supporting structure;
- a clear separation of the planning and operational functions;
- opening of the markets to allow new entrants to offer services;
- procuring supported services by transport authorities, using market processes;
- corporatisation of formerly public-sector operating entities – i.e. transformation into entities that are structured as companies with associated corporate, governance and accounting principles;
- privatisation, joint-ventures and other means of modifying the ownership base of parastatal operators;
- making public assets available to both public and private sector bidders; and
- mobilising private investment for public infrastructure and services.

These trends and motivations have led to different market frameworks emerging, with greater or lesser opportunities for transport operators to participate, and with different relationships between the authorities and the operators. The market framework in a specific location usually reflects historical context, the national and local policies, and the capacity of the local government. In more mature systems, the various issues of policy, planning, development and integration of the passenger transport, regulation, financing, etc. are handled by a Transport Authority either for the Urban Area or for a broader Metropolitan Area. The Transport Authority is usually accountable to the governance structures of the local government, but it can be at regional or national level.

Change in Urban Public Transport in Developing and Transition Countries

This paper concentrates on the connections between changes to the regulatory framework, changes to the nature and form of the operator, and the nature of

transport supply. Examples of these are tabulated in Tables 1 to 3. These elements are clearly interconnected and developing countries and countries in transition do not show a unique response. However one characteristic of the experience of developing countries and countries in transition is common: the frameworks in which the urban transport sector exists change and often change rapidly. In some cases the main reason is because society itself is changing. In other cases, it is because a framework is lacking, or contains inherent weaknesses that need to be adjusted. In contrast, in mature systems where the frameworks are stable, there is often no fundamental change for decades and all participants understand the roles, relationships, opportunities and boundaries.

Table 1. Changes in the role of regulator/transport authority and market structure.

Change	Examples
Regulation of the Urban Passenger Transport is introduced or restored	In Ghana, the passenger transport sector had been self-regulating since 1983 as the minibuses became the dominant form of transport supply. In 2008, the function of regulation has been restored to Local Government. Bye-Laws were passed in each area to provide the legal basis. (Finn, 2008, 2009). In Manila, Philippines, while route franchising remains with LTFRB, since 2003 the city authority (MMDA) regulates the flow of buses along the principal urban orbital (EDSA) and at both their terminals and stopping places along EDSA.
Regulation of the Passenger Transport is transferred to the City	In Jordan, the Passenger Transport Regulatory Commission (PTRC) was the regulator for all passenger transport services throughout the country. In November 2007, all responsibility for urban passenger transport in Amman, including the regulatory role, was transferred to Greater Amman Municipality. PTRC continues to regulate services elsewhere in the Kingdom.
Regulatory Capacity is developed within the Local Government	In Accra and Kumasi (Ghana), Amman (Jordan), Tbilisi (Georgia), new Public Transport Departments have been established, or existing general departments have been restructured. Specific public transport and regulatory expertise is engaged and developed. (Finn, 2008). In Manila, MMDA has developed a unit, regulations, staffing and supporting technical systems to manage the bus routes operating on EDSA (30+ routes, 3000+ buses). This system is called the Organised Bus Routes (OBR).
Changes are made to rules for market entry and/or procurement	Accra and Lagos (Nigeria) will procure bus and feeder bus services for the BRT scheme under tendering, awarding Route Service Contracts to the successful applicants. In Tbilisi, in 2001 the unregulated minibus operators were formalised by opening a competitive tender process and awarding operating permits to organised operators
Regulators develop new models for procurement of services	In Recife (Brazil), a new model for competitive tendering has been developed to overcome three decades of wrangling among the different institutions, and establish a basis for a specified service, regulation of quality, and financing (Filho, 2007)

Table 1. Changes in the role of regulator/transport authority and market structure (continued).

Regulators develop new contractual basis for bus services, or amend existing ones.	In Kaunas, Lithuania, as part of major restructuring of the reorganisation and finances of the municipal bus and trolleybus companies, the City entered a 10-year public service contract with the municipal company, based on committed services and quality, subsidies, and investment program (Bruggeman, 2009). In South Africa, a system of subsidy contracts was established with the private bus operators for commuter bus services. In total there are about 115 such contracts, with differing payment bases. The Government is currently attempting amend all contract to payment for kilometres operated (Walters, 2009). In Santiago de Chile, as part of the TranSantiago BRT system, a new contractual and payment system was established for the operators so that they would participate in integration of network and tariffs (Munoz, 2007)
The Regulator makes direct intervention to restructure the operator sectors	In Sri Lanka, in 2005, the 11 regional state companies were reformed into a single entity. There are now approximately 19,000 private bus operators in addition. The National Transport Commission has initiated five pilot projects to achieve clustering of the operators in pilot areas, to improve their organisational capacity, use of resources, and service quality (Kumarage, 2007)
The Transport Authority makes vehicles available to the operator sector :	In Accra and Tbilisi, the Transport Authority or Government acquired vehicles to re-establish large-bus operations. In Accra and Tbilisi, these were initially second-hand buses which have limited benefit and many of these vehicles soon end up out of service. (Finn, 2008) In Dhaka (Senegal), the Transport Authority, using Government-negotiated loans, established a special purpose company to acquire vehicles and lease them to the operators who accept to be regulated and to operate under contract (IBIS, 2008)
The Transport Authority provides funding or financial incentives for bus purchase and/or technology upgrades	In Kaunas, the City of Kaunas arranged finance through EBRD for the purchase of new buses and associated support systems for the municipal bus company (Bruggeman, 2009). In South Africa, a scrappage grant has been implemented to incentivise bus operators to replace old vehicles, equivalent to about one sixth of the cost of a new vehicle (Walters, 2009). In many rural areas in Brazil, schools transport involves long journeys in difficult operating conditions. The Brazilian government has sponsored the development and deployment of more robust and suitable vehicles (Carvalho, 2009).
The Transport Authority finances new transport infrastructure to enhance the public transport system and to provide suitable facilities to operators	In Accra, Amman, Dar es Salaam (Tanzania), Delhi (India), Mauritius, and many other cities, the Transport Authority has funded the development of Bus Rapid Transit infrastructure and supporting facilities to improve operating conditions and to make the passenger transport attractive (Richmond, 2006, Torres, 2007, Finn, 2008, 2009) In China, the Municipalities constructed passenger interchanges, terminals and overnight parking by authorised operators (Dotson, 2007) In Amman, municipality has constructed bus and shared taxi terminals (Finn, 2009)

Table 2. Changes in form of operators and participation of the private sector.

Change in the form of the operator	Examples
Public Sector Bus Operators have their corporate and ownership basis changed	In China, the State Owned Enterprise bus companies in many cities have been transformed from administrative units of the Municipality into joint-stock companies, and inward investment is now accepted either as direct investment or the creation of joint-ventures (Dotson and Finn, 2008) In Kazakhstan, the first wave of public transport reform in 1996-7 allowed the municipalities to privatise their transport companies and to allow new market entrants. Most cities apart from Almaty and Astana privatised their bus companies. All cities allowed new market entrants. (Finn, 2005) In Tbilisi, the Municipal bus company has been transformed into a corporate entity, placed at arms-length from the Municipality, and has been considered for privatisation (Finn, 2008)
Private operators are given controlled access to the market	In Rostov-on-Don (Russian Federation), the private sector had achieved 63% of the market by 2005 within a controlled competition structure (Zyryanov and Sanamov, 2007)
Individual minibus operators form associations or co-operatives to gain route contracts, or to have access to facilities or finance	In Samarkand and Bukhara (Uzbekistan), the independent minibus operators were formed into associations at the initiative of the Municipalities. This gave them access to operating permits, and to maintenance and shared purchasing facilities (Gwilliam et al, 1999) In Tbilisi, in 2001, about 3,500 minibus operators consolidated into 64 companies to gain operating agreements on 223 lines. The companies were shell companies for convenience. There has not been further integration or consolidation of assets (Finn, 2008) In Dhaka, minibus operators formed into cooperatives to be collectively responsible for loan repayments on minibuses made available under a vehicle leasing scheme sponsored by government. (IBIS, 2008) The 17 East Jerusalem bus operators formed an association in 2002 and operate under common livery and co-ordination. This has given them access to two good quality bus terminal facilities and bus purchase grants from the Israeli Ministry of Transport. In Cebu, Philippines, the jeepney association VUDTRASCO secures operating franchises from LTFRB and its members then operate under the association franchise (c. 500 units).
The private sector participates directly in infrastructure and rolling stock development and financing	Various financing and investment instruments (PPP, DBOT, BOT, PFI) have been developed to mobilise private finance for major infrastructure project, e.g. LRT in Manila (Philippines) In Fuzhou (China), investors provided direct financing for buses without any participation in the companies or operations, and received their payback over the life of the vehicles

Table 2. Changes in form of operators and participation of the private sector (continued).

Private sector investors purchase vehicles and lease them to drivers.	*Tro-tros* and taxis in Accra, Jeepneys in Manila, *marshrutka* in Tbilisi, shared-taxis in Palestine are examples where private individuals purchased vehicles (usually second-hand) and then rented them to drivers who must pay an agreed amount to the owner In Samarkand and Bukhara, extended families have pooled savings to purchase minibuses, since finance from banks is either not available to them or too expensive.
Operating companies change their business model to gain efficiencies	Bangalore Metropolitan Transport Corporation (India), while remaining in the public sector, has restructured its business model along private sector lines. It has outsourced all non-core and many core activities, increased efficiency, and has become profitable (Torres, 2007) In Indore (India) the city has established a new company that procures all transport and support services from the private sector (Torres, 2007) In Dubai, the Public Transport Authority has contracted out the maintenance of all recently acquired buses, being about 1,400 of the total fleet of 1,800 buses. In Philippines, Ceres, the largest private operator of intercity buses, has established its own facility to build buses, manufacture spare parts, and overhaul major units.

Table 3. Changes in transport supply.

Change in transport supply	Example
Large-scale expansion of the conventional transport services to meet demand	In many Chinese cities, bus fleets and kilometres operated have more than doubled to meet population growth (Dotson and Finn, 2007) In Dubai, as it became clear that a roads-oriented strategy was failing and congestion levels were unacceptable, the bus service has been completely restructured and repositioned. An additional 1,300 new buses have been added to the fleet since 2002 and this will increase further in 2009-2010 to provide the feeder services to the 2 new Metro lines (Kaiser, 2007)
Introduce high quality passenger transport modes	Bus Rapid Transit has been implemented extensively in South America, Asia, and now systems are being implemented in Middle East and Africa
Large-scale paratransit services are established to replace large-bus services which have failed to expand to meet the increasing travel demand from population growth	*Collectivos* in Rio de Janeiro, Sao Paolo, Recife and other Brazilian cities have emerged to provide services from the poor areas (*barrios)*. In Rio de Janeiro, some of the routes have more than 700 vehicles. (Goncalves, 2005, Brasiliero, 2007) In St. Petersburg (Russian Federation), minibus services (*marshrutka*) were permitted on fixed-routes to complement the conventional services as the financial conditions did not support bus fleet expansion. By 2002, these services carried 370 million trips annually (or 14.1% of the total public transport market), more than the trolleybuses and almost as many as the trams. (Finn, 2003, 2009) In Manila, about 50,000 minibuses (Jeepneys) now provide an extensive network of services, and are preferred by many users due to direct services without transfers.

Table 3. Changes in transport supply(continued).

Large-scale paratransit services emerge to replace large-bus services as they go into decline	In cities of Kazakhstan, minibus services (*marshrutka*) replaced large-buses on routes where they ceased to operate, or offered insufficient service. In general, they did not attack routes where buses continued to operate. (Finn, 2005, 2008) In Ghana, Senegal, Tbilisi, and Santiago de Chile, paratransit which initially emerged to fill service gaps became the dominant form of transport and in some cases effectively 'killed' the remaining large-bus service
Shared taxis emerge to complement the large-bus transportation or to meet specific needs	In Amman, about 3,200 licenced shared-taxis now operate on 70 fixed routes from the centre to designated suburbs, on a fill'n'go principle. The shared taxis carry more passengers daily than the current underdeveloped large bus network In Ghana, in the rural areas and villages, shared taxis (cars) provide an equivalent service to the minibuses, and are the dominant form of public passenger transport (Finn, 2008)
Large bus services are restored when the city or operator regains financial and/or organisational capacity	In Ghana, the Government facilitated the emergence of a new large bus company, leveraging private and public financing. Metro Mass Transit now has in excess of 500 large buses, and operates services in the main cities and on intercity routes. It has regained 15-20% of the market share in Accra (Finn, 2008) In Kaunas, the refinancing of the municipal bus company's fleet (referred to in Table 1 above) had the specific objective to restore the large bus operations and reduce minibus operations. The market share of the large bus sector increased from 50% to 90% (Bruggeman, 2009). In Tbilisi, the Municipality has funded the acquisition of more than 600 vehicles for the Tbilisi Bus Company, being a mix of second-hand large buses and new mid-sized buses. TBC has restored service on all of the main routes, and has improved from a 5% market share in 2001 to close to 50%, at the expense of minibuses (Finn, 2008) In Cebu, Philippines, all public transport is currently provided by jeepney or taxis. BRT is planned which will introduce urban large-bus operations.

These tables show that the changes can be wide-ranging and that individual countries experience more than one type of change.

Discussion

The cases presented in this paper indicate a very wide range of contexts and mechanisms. Some have clearly been more successful than others, but each has been implemented with the objective either of strengthening the sector, or of making a practical response to challenges within the prevailing constraints.

In all cases, there is a stimulus to planned structural change within the passenger transport sector. The principal stimuli identified here are a leading political vision and/or policy, e.g. urban spatial development, economic regeneration; a need to respond to rapid growth in demand for passenger services due to population growth;

a need to respond to serious challenges such as traffic congestion or air quality; a recognition that major infrastructural projects (e.g. BRT) cannot be implemented until the regulatory framework or operator structure/quality is improved ; or a need to respond to serious degeneration of the available passenger transport services.

A number of contributing success factors are identified for developing urban transport services in general in cases where they are currently underdeveloped, fragmented or of low quality :

- Presence of strong political leadership and support
- Setting transportation strategies and programs within the context of national plans and strategies which have already been approved by Government
- Establishment of an effective regulatory framework defining responsibilities, roles, market access, and rights of initiative, based on sufficient legal backing
- Establishment of an empowered regulator with appropriate structures, capacity, continuity, and institutional memory
- Development of the human and organisational capacity of the Regulator
- Establishing and maintaining real consultation and collaboration with the key stakeholders, especially the operators and impact communities
- Organisation of a controlled transport market with quality-based entry, supported by an effective enforcement capacity
- Establishment of contractual or agreement-based relationships
- Creation of effective mechanisms to align supply with demand, and to stimulate investment/provision of increased resources where they are required
- Development of the business and operational/technical capacity where the operator sector is already fairly well structured.
- Where the operator sector is fragmented, consolidation to facilitate economies of scale, increase in professionalization, and access to finance
- Stimulus measures to encourage operators to upgrade and improve

The successful implementation of these factors confronts the minibus, paratransit and informal sector which has played a very important role in some, but not all, urban areas in developing and transition countries. Where they are already established, a number of factors determine their future development and, in some cases, their survival as a significant transportation mode:

- The attitude of the authorities, in particular whether they accept the presence of this transportation mode (as in Ghana, Kazakhstan, Philippines, South Africa), whether they wish to contain it in a reduced or transformed role (e.g. St. Petersburg), or to systematically replace it (e.g. Kaunas, Tbilisi).
- Whether the services are seen as a permanent feature of urban transport provision or something transient. For example, in Ghana the services have been a feature for a quarter century, are well understood by the population, and the routes are well known and as a result are now seen as a permanent feature; in Philippines, jeepneys are accepted by both authorities and passengers as an integral feature of urban transport and culture

- The willingness of the authorities to develop a formal regulatory framework within which the paratransit services can operate and develop, as part of their policy for the urban transport sector (e.g. South Africa, Philippines)
- The ability of this sector to evolve from operational collectives of individuals into companies which can invest in vehicles, rationalise and optimise resources, and develop their business capacity. Despite high revenues in the sector, business evolution has not observed and this will inevitably impact on the potential longevity of operation
- The ability and willingness of this sector to invest in continuous product improvement, especially in newer vehicles, professionalization of drivers, and better maintenance so as to evolve from a low-cost, low-quality product to something citizens would use by choice
- The determination to improvement service quality and reliability, and to integrate with other transport modes operating in the urban area. In South Africa, the paratransit is being integrated with the BRT. The restructuring of urban transport in St. Petersburg included the *marshrutka* as part of the planned supply. It occurs informally in Amman where the shared taxis operate from the bus terminals, and in Brazil when *collectivos* bring customers to Metro transfer points.

Conclusions

This paper has proposed a framework for classification of changes in the urban transport sector in developing countries and countries in transition. The presentation of evidence showing the way in which regulators and markets, the form of operators and the supply of transport have changed in response to contextual change have highlighted the very many different ways in which sectors respond. It reveals that change (whether legislative or policy) does not lead to a unique response.

The practice presented in this paper shows commonality in the desire for change but distinct differences in policy and outcomes. This allows the factors which are evident in successful outcomes to be identified. It also establishes the importance of the minibus, paratransit and informal sectors in the process of change in many countries, and the issues which affect whether this sector is likely to be a permanent or transitory feature of the urban transport sector supply.

References

Brasiliero, A. et al. (2007).“Regulatory Reform of the Urban Bus System in Recife (Brazil): Stakeholders and Constituency Building Process”, *proceedings of 10th International conference on Competition and Ownership in Land Passenger Transport*, Hamilton Island.

Carvalho, W., da Cruz, R., Câmara, M., and de Aragão, S. (2009). “Rural School Transportation in Emerging Countries, the Brazilian Case”, *Thredbo 11 Conference, Delft*, Netherlands

Dotson, E. And Finn, B. (2007). "An Overview of Ownership and Market Reform of Urban Bus Service in Chinese Cities", *proceedings of 10th International conference on Competition and Ownership in Land Passenger Transport*, Hamilton Island.

Filho, R.O. (2007). "Urban Transport in South America : Trends in Competition and Competition Policy", *proceedings of 10th International conference on Competition and Ownership in Land Passenger Transport*, Hamilton Island.

Finn, B. and Nelson, J.D. (2003a). "An international perspective in the changing structure of the urban bus market" , *Transit Planning and Development, Management and Performance, Marketing and Fare Policy*, TRB Research Record No.1799.
Finn, B. and Nelson, J.D. (2003b). "A Functional Model for an Urban Passenger Transport Authority", *Competition and Ownership in Land Passenger Transport : selected refereed papers from the 8th International Conference (Thredbo 8)*, Rio de Janeiro.

Finn, B. (2005). "Transition from Public Monopoly to Controlled Competition in a Post-Socialist Economy – Experience in the Republic of Kazakhstan". *Selected Papers from the 9th International Conference (Thredbo 9),* Lisbon, September 2005. (659-676). Elsevier, Oxford.

Finn, B. (2008), "Market Role and Regulation of extensive urban minibus services as large bus capacity is restored – case studies from Ghana, Georgia, and Kazakhstan". *Research in Transport Economics*, 118-125,Elsevier, Oxford.

Finn, B. (2009). "What can European and North American Cities learn from Large-Scale Paratransit in the Rest of the World?", *TRB 88th Annual Meeting*, Washington.

Finn, B; Arthur, B. A and Gyamera, S. (2009). "New regulatory framework for urban passenger transport in Ghannaian cities". *Thredbo 11 Conference*, Delft, Netherlands

Gonçalves Pereira, C. et al. (2005). "Paratransit Regulation in Rio de Janeiro : a Comparison of Two Surveys", *proceedings of 9th International conference on Competition and Ownership in Land Passenger Transport*, Lisbon.

Gwilliam, K.G., Meakin, R.T., Kumar, A. (1999). "Designing Competition in Urban Bus Passenger Transport. Lessons from Uzbekistan" *6th International Conference in Competition and Ownership in Land Passenger Transport*, Cape Town.

Gwilliam, K.G. (2003). "Bus franchising in developing countries: Recent World Bank Experience", *Competition and Ownership in Land Passenger Transport: selected refereed papers from the 8th International Conference (Thredbo 8)*, Rio de Janeiro.

Gwilliam, K, Zhu, L. And Finn, B. (2009). "Recent developments in Bus Transport in China", *Thredbo 11 Conference*, Delft, Netherlands

IBIS Transport Consulting (2008). "Dakar Bus Financing Study, Final Report", unpublished report for the World Bank.

Kaiser, J., (2007). "Dubai Public Transport Bus Master Plan – a new era of public transport services in the world's fastest developing city", *proceedings of 10th International conference on Competition and Ownership in Land Passenger Transport*, Hamilton Island

Kumarage, A.S. and Jayaratne, M.D.R.P. (2007). "Lessons in Ownership, Regulation and Management from 100 Years of Bus Transport in Sri Lanka", *proceedings of 10th International conference on Competition and Ownership in Land Passenger Transport*, Hamilton Island.

Kumurage, S. Weerawardena, J. and Piyasekera, R. (2009). "A concession model to promote rural bus services in Sri Lanka". *Thredbo 11 Conference*, Delft, Netherlands

Richmond, J. (2006). "Mauritius Transport Consensus Forums" Briefing note to Cabinet o the Government of Mauritius.

Torres, M., (2007). "Innovative Public-Private Sector Cooperation in Urban Transport in India", *TRB 86th Annual Meeting*, Washington

Walters, J. (2009). "Is the bus contracting system in South Africa leading to trusting partnerships between contracted parties? An analysis of funding issues and the impact on relations between government and operators". *Thredbo 11 Conference*, Delft, Netherlands

Zyryanov, V. And Sanamov, R. (2007). "Improving Urban Public Transport Operation: Experience of Rostov-on-Don", *proceedings of 10th International conference on Competition and Ownership in Land Passenger Transport*, Hamilton Island.

Notes:

(1) Many of the cited references are drawn from the Thredbo series of conferences. These can be found at the University of Sydney website at www.itls.usyd.edu.au.
2) Unreferenced items in Tables 1, 2, 3 above are from the author's experience.

An available International Freight Transport System with Rail Container and RORO/Ferry Ship

Guoquan Li[1]

[1] Senior Researcher, Transport Planning Lab., Railway Technical Research Institute, 2-8-38, Hikari-cho, Kokubunji-shi, Tokyo, 185-8540, Japan; Email: ligq@rtri.or.jp

Abstract

With the huge change of the international freights between China and Japan in contents and quality, it is needed to construct the new transport system with high certainty, punctuality, reliability and security for the logistics services with greater sophistication and diversity. This paper firstly analyzes the actual situations of China-Japan's international trade and relevant freight, and the relevant issues of conventional marine transport. Then, an intermodal transport system for these freights with the use of rail container and RORO/Ferry ship is described conceptually. Finally, the availability of the international freight transport system is confirmed through a case study based on the pots of Shanghai and Hakata.

Key Words: international freight; intermodal transport; RORO/Ferry ship; rail container

Introduction

As well-known, after 2000, with the remarkable development of China, the eastern Asia has become one of the most active areas in socio-economy of the recent world. In the meantime, the international trade surrounding to China have been grown into one of the most important world trades. In that, China-Japan's international trade changed greatly not only in quantity but also in quality. And the economic relations between the two countries have become closer and closer from the simple trade to the daily production processes and activities of factories & plants.

In order to be suitable for these situations, firstly, it is necessary to investigate the demands of shipper, analyze the issues of conventional international transport system, and create the new transport services with more efficiency and more rationalization between China and

Japan. Then, with the geographical advantages among north Kyushu regions in Japan and east coastal regions in China, in particular, an intermodal freight transport system is proposed based on Hakata Port and Shanghai Port, with the use of the rail container and the daily service of RORO/ferry transport between the two ports. Furthermore, the relevant feasibility and availability about the realization of the proposed transport system is preliminarily discussed.

International Trade Transition and relevant Freight between China and Japan

International Trade Transition between China and Japan. Figure 1 shows the transition trend of international trade volume between China and Japan from 1980 according to the Japanese foreign trade data. Because there are the differences of foreign trade statistics between the two countries, based on the Japanese data (Ministry of Finance: materials of trade statistics, 2007), the international trade volume imported from China was more than the volume exported to China, that is, the trade deficit in Japan appeared after 1990. But, according to the relevant data of Chinese foreign trade statistics (The Ministry of Commerce of PRC, 2007), the export volume of international trade to Japan is always less than the import volume from Japan. In any case, as for the international trade value between China and Japan, it is clear that it is intensely growing up both export/import volume since 1990. The trade value between the two countries reached about 28 trillion Japanese Yen in 2007. It was 3 times of the trade value in 2000 (Ministry of Finance: materials of trade statistics, 2007).

In addition, the international trade value between Japan and China (only in the mainland area) was over that between Japan and U.S.

Trade Structural Change. Table 1 shows that the change of international freight contents to export from Japan to China has been very great for the recent 20 years(Ministry of Finance: materials of trade statistics, 2007, Workshop for the flowing trend of the import & export freight, 2002). In 1980's, machinery products including household appliances such as Japanese television receivers, iron and steel, and automobiles were the main export goods from Japan to China, and the relevant occupation ratio in export trade was absolute high.

After 2000, the electron components, machine parts and precision instruments etc. have become the main export goods. According to the relevant statistics, it may be said that products are made or assembled in China has entered the mature phase.

On the contrary, in the main goods imported from China, the resource goods such as oil and coal, etc. were occupied the most in 1980's. The life consumption goods such as clothing, agriculture or fishery products were become the leading position in 1990's. After 2000, the instrumentation including household appliance, electron products, and personal computer accounted for over 30%, and then, the life consumption products including clothing took over 20%. It is clear that the international trade between China and Japan has fundamentally changed in contents and quality.

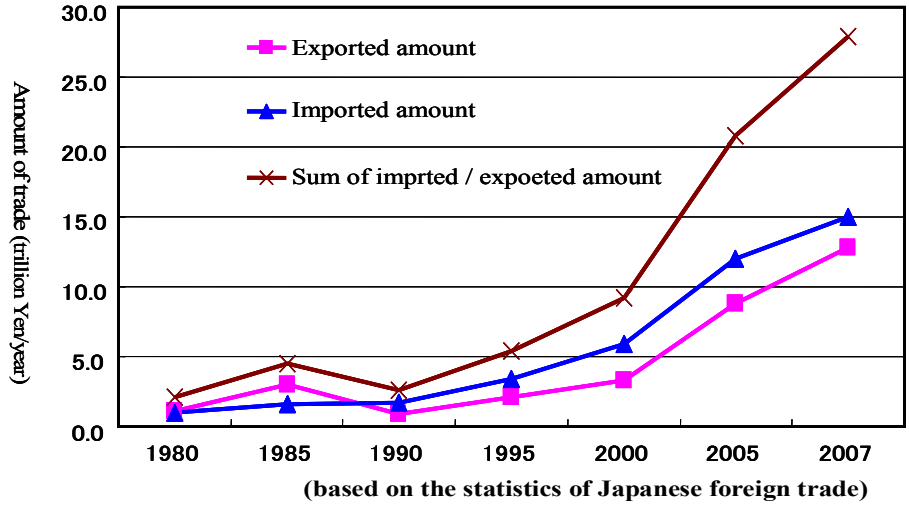


Figure 1. Amount of international trade between China and Japan.

Table 1. Commodities exported/ imported to/from China.

Exported commodities		
Percentage	% in 1985	% in 2003
Machinery	37.1	50.6
Inside: TV receiver	8.6	
Electric circuit arrangement		3.8
Textile machine		
Metalworking machinery		
Integrated circuits		7.5
Computer component		3.0
Iron and steel	25.6	6.5
Automobile	12.0	3.4
Automobile parts		3.1
Ship	3.0	
Plastic	2.8	3.9
Metal products	2.2	
Organic chemical compound	1.4	
Yam of synthetic fiber		5.1
Soft goods	1.3	
Pulp & paper		3.8
Precision equipment	1.0	
Others		4.8
	12.5	18.8

Imported commodities

Percentage	% in 1985	% in 2003
Oil	33.9	
Oil products	9.0	
Coal	2.8	
Clothing	7.1	20.5
Fish and selfish	2.6	3.1
Corn	5.2	
Fruits	1.2	
Vegetables	2.2	2.0
Cotton	1.2	
Cotton textile	2.9	
The hair of the Angola	1.6	
Metal products		2.2
Machinery		32.5
Inside: computer		11.4
Iron & steel		
Travel outfit and bag		1.
Toys and games		2.3
Footwear		2.8
Furniture		2.3
Precision equipment		3.4
Others	29.8	27

The present situations of the international trade between China and Japan have become very interesting. A new kind of international trade relationship has been formed. The parts and components made in Japan are exported to China. And then the relevant products are completed or assembled in China's factories. Finally these goods are re-exported from China to Japan or other countries. Therefore, with the promotion of the international division of labor in manufacturing industries, the freight transport pattern between China and Japan shall be changed to meet the new needs of shippers. The conventional marine transport was to move the products from factory to the location of consumers. The new type of transport is to move the parts and components from one factory to other plants located in different countries, and to re-move the products back or to another country. In order to increase the logistics efficiency and reduce the quantity of inventory of products, the transport system with the certainty, punctuality and safety is strongly wanted to meet the demands of the schedule of production line, and be suitable for the seasonable fluctuation of market.

Freight Transport for International Trade and Relevant Issues

Marine Container Transport. After 1978, the harbors in China have gradually been opened to foreign countries and their enterprises. Since 1990, many modernized ports have been constructed, and by now, the construction of modernized ports has entered the rush age

along the coastal areas from northern to southern regions.

In addition, with the predominance of coastal areas in Japan, in order to open up the international shipping route, almost each region with coast constructed one, or more one and multiple ports. By now, about 63 ports have operated the international scheduled services of marine container ships.

The marine container volume between Japan and China in 2005 reached over 2.8 million TEU, the increasing rate in the three years from 2003 (the volume was 2.2 million TEU) was 30%. The large part of these containers are concentrated on Shanghai Port, Tianjin Port, Qindao Port, Dalian Port, and especially, the container ratio of Shanghai port that occupied in the total volume was about 45%.

Although the distances of the shipping routes between the ports of China and Japan, are short, the transport time gets longer. With the changes of trade & economy relationships, the conventional patterns of marine container transport can't meet all the demands of freight shippers between China and Japan. Especially, with the international division of labor in manufacturing industries, more rapid marine transport service is required for component supply matching with the production line in order to keep the daily regular operations of a factory.

Issues In Conventional Marine Container Transport. Although the longest distance between the ports of China and Japan is about 2,000 n.m., and based on the ship technology, the sea shipping time for all sea routes among the ports can almost be controlled in less than 4 days, actually, most of real routes suffer more than 4 days (Japan Ship owners' Association, 2007).

For example, the shipping days in a container ship from the Tokyo Port / Yokohama Port to Shanghai port are about from 3 to 7 days although the distance between the two ports is only about 1,100 n.m. In addition, the sea route of the containership from Kobe port/ Osaka port to Shanghai port with about 800 n.m. needs about from 3 to 5 days. Especially, to the local port, there are a few of the sea transport routes more than 10 days among these Japanese local ports and the ports in China.

Furthermore, In the conventional marine container transport, there are the other issues such as longer operating time in the container terminal of port, few international routes connecting to the local container port, and that concerning the surface transport of the marine container freight etc.

As stated above, in the present situation of marine container transport, we have to face these issues such as shortening the shipping time of sea route, improving the loading/ unloading operation and handling process of container cargo at port terminal, increasing the numbers of shipping services through the consolidation of ports, supplying the convenient access to port, and easing the environmental problem, etc. To the international marine container

freight between China and Japan, it is dispensable to construct a more effective transport system as one of solutions with the use of geographic predominance of the near sea transport.

In addition, the Hakata port is justly located in the middle position of Shanghai and Tokyo in geographic space. The surface transport access such as railway, express way, national highway, the seaside road is relative convenience, besides.

Therefore, How to effectively use these advantages of Hakata port is really worth considering as one of the available solutions for China-Japan international freight transport, to match with the changes of shipper's needs.

Intermodal Transport System for International Freight with Rail Container and RORO/ Ferry Ship

Conceptual Framework. Hereby, with the use of nearing sea advantages such as shorter shipping time and possible frequency, we will propose and discuss the intermodal transport system based on the introduction of the marine daily service for international freight with the RORO / ferry ship and the network of rail container.

This system is not only suitable for the shippers' daily needs with the reasonable and efficient connection between RORO/ferry shipping and surface transport, especially the environmental-friendly railway, but also has the competitiveness with the aviation freight. The relevant characteristics of this system can be stated as follows.Establishment of the marine daily service with RORO/ferry ship.Simply and easily loading & unloading operations of rail container by top lifters or folk.Environment-friendly surface transport by nationwide rail network.

Marine Daily Service For The International Freight Transport Between China And Japan. In the case of the Japan coastal transport, there is RORO ship daily service for domestic freight transport between Tokyo and Tomakomai. Because a ship with the greatest speed 30 knots is adopted, this shipping route with about 570 n.m. can be connected within about 20 hours. Generally, in the case of a ferry, the length of shipping route with one day is about 650 n.m., and in the case of a RORO ship. It is about 600 n.m. And to a containership, it is less than 400 n.m.

In the spatial distribution, the shipping routes between the Chinese ports from Shanghai to Qindao, and the Japanese ports in the Northern area of Kyushu and Yamaguchi region are less than 600 n.m. The shipping route between Shanghai port and Hakata port with only 500 n.m. is less than that of Tokyo-Tomakomai. Therefore, it is possible that China-Japanese marine daily service will be realized only if the coastal shipping technology is used.

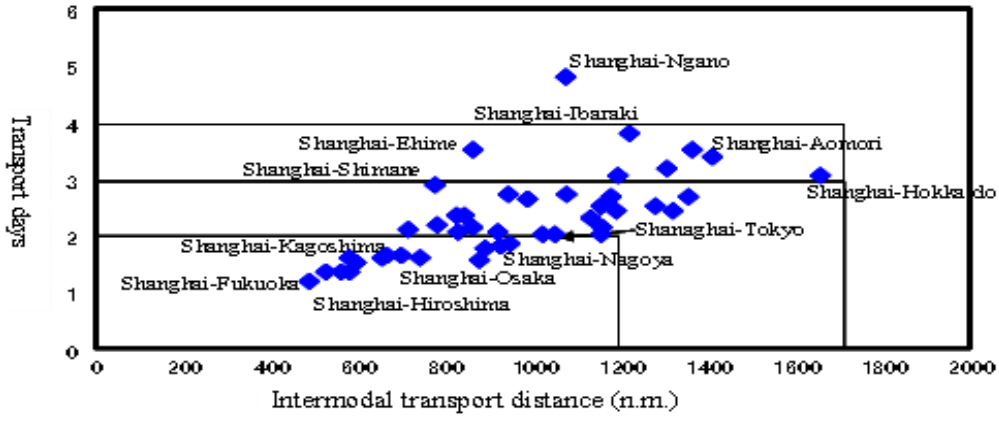


Figure 2. Intermodal transport distance and transport days between Shanghai and each region of Japan.

Based on the Hakata port as conjunction base of international intermodal transport system between China and Japan, the door-to-door transport time from each area of Japan to China can be greatly reduced as shown in Figure 2. Here, the door-to-door transport time of the system includes marine shipping time, access time between Hakata port and Fukuoka freight station, the surface transport time between Fukuoka freight station and each region of Japan, the relevant freight import & export proceeding time, and freight loading & unloading time at port and station, in addition, delivery & picking up time of freight around station and port. According to this, it is possible that the freight from all areas of Japan except Nagano prefecture can be transported to Shanghai within 4 days. In that, the freight from many regions of Japan such as all areas of Kyushu, Yamaguchi, Hiroshima, Okayama, Osaka, Nagoya, Gifu, Shizuoka, Tokyo, Toyama, etc. can be transported to Shanghai within 2 days, the transport time. Almost of areas can be linked with Shanghai in 3 days. The freight from Hokkaido area can be exported to China in 3 days through the intermodal system.

Analysis on the Realization of the Transport System

Analysis' Conditions. The realization of the system-related intermodal transport can be discussed based on the following conditions.

About 45% of the marine container freight between Japan and China is concentrated in Shanghai port.

The containerization rate in the total freight between Japan and China can be set up 20%.

With the RORO ship at the maximum speed of 30 knots, the shipping time of ship can be controlled within 24 hours, because the length of marine shipping route is less than 600 n.m.

To the surface transport, In the case of the middle or long range, the existing railway network is actively used for freight transport, and in the case of the short distance, the

delivery or picking up of freight is based on trucking.

The setting of the surface transport routes can use the freight railway schedule time table for references, from the rail station which is near to the port to the central district of each prefecture (The newspaper of transport, 2007).

The relevant fares or rates concerning freight transports are assumed by using the actual conditions based on freight tariff rate and various list of charges.

Estimations of the Freights for the System

Sharing Rate. Using these conditions for the transport system stated previously, we can estimate the sharing rate of the intermodal freight transport system from each prefecture to Shanghai outskirts area, by using the choice model of the TSL built by the marine technical safe bureau of MLITT of Japan as the formula (1).

$$P_k = \exp(W_k) / \sum_{k=1}^{n} \exp(W_k) \tag{1}$$

Where, *Wk*: disutility functions[7] for modal choice as follows:

$$W_k = -2.1224 - 4.2929 c_k - 6.5246 t_k$$

ck: transport expenses when using the system

tk: transport time when using the system

The relevant value of statistic test for each parameter of disutility function is -4.006, -5.768, and -2.672, respectively.

Freights of the System. After 1990's, with the marked changes of the contents of the international freight, the shippers want the new transport services to meet the international freight transport. According to the relevant investigations (Association of ship building and business, 2000), although about 53.8% of shippers was still without the changes in choosing the transport means, the shippers changing means from air cargo to marine container shipping was about 13.4% and from marine container to air cargo was about 32.8%. Actually, in the air cargo and marine container freight between China and Japan, there are many common items of goods. The average unit price of the air cargo is expensive with 10,000 Japanese yen per 1kg, and the average unit price of the marine container freight is low with 300 Japanese yen per 1kg (Ministry of Finance: materials of trade statistics, 2007, Workshop for the flowing trend of the import & export freight, 2002). It is easily judged that there are the massive niche needs between marine container and air cargo. Therefore, the freights of the constructed system can be calculated by the formula (2).

$$FVI_k = \sum_{i=1}^{m} (\alpha_{i1} \times \alpha_{i2} \times V_i \times P_k^i \times \gamma_k^i) \tag{2}$$

where, FVI_k: freights of the *k* system (ton),

α_{i1} : containerization ratio in the export/import commodities of the *i* region,

α_{i2} : concentration ratio from the *i* region to the one port appointed,

V_i : volume of the export / import commodities in the *i* region to/ from China,

γ_k^i : realization rate of the conversion from conventional marine transport to the *k* system in the *i* region.

Case Study. As analyzed previously, according to the spatial advantages of the Hakata port between China and Japan, the intermodal freight transport system based on the Hakata port will contribute to shippers' niche demands between air cargo and marine container.

Furthermore, according to national survey on import & export freight flow (Workshop for the flowing trend of the import & export freight, 2002), the transport demand to shift the marine freights between Japan to China to the intermodal freight transport based on Hakata port can be calculated as case study. Figure 3 is the calculating result for the freights of the system in the import commodities from China. It is clear that only if the small marine freights are converted in each region, there are enough freights volume for the system.

According to the results shown in Figure 3, when the kind of RORO/ferry ship with the loading capacity of 200 chassis is adopted in the system, and supposing that the loading rate is higher than 90%, the needed ships per week is shown in Table 2 corresponding to the realization rate of conversion. It is derived that the system can keep the marine daily service operated with the RORO/ferry ship.

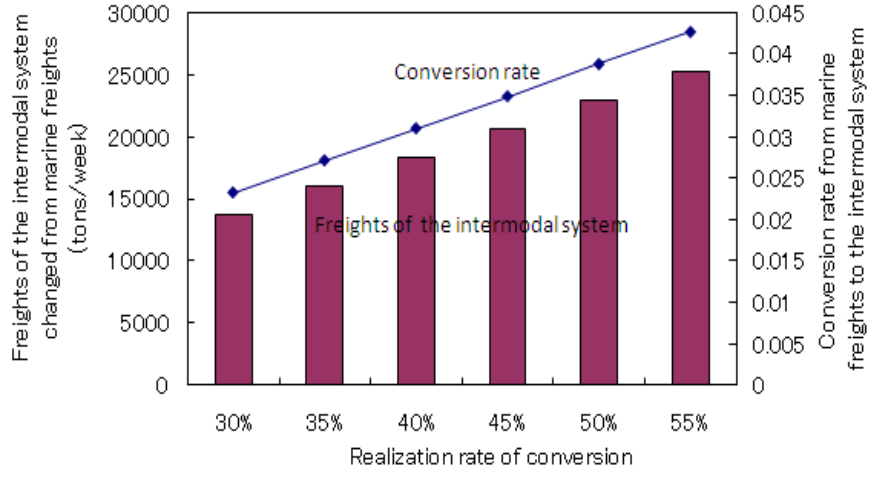


Figure 3. Freights' estimation of the international intermodal freight transport system.

Table 2. RORO/ferry ships needed in the system

Realization rate (%)	30	35	40	45	50	55
Ships /week	6	6	7	8	9	10

Note : loading capacity per ship is 200 chassis, loading rate is 90%.

Conclusions

In this study, based on the variations of international freight between China and Japan, the relevant issues concerning the conventional marine container transport system are analyzed. With the advantages of Hakata port in geography, an intermodal freight transport system in that the rail network is combined organically with marine daily service is proposed.

In addition, the realization conditions of the proposed system and its relevant effectiveness are discussed synthetically.

Finally, as a case study based on the ports of Shanghai and Hakata, the possible freights of the system shifted from conventional marine ships are estimated. In the meantime, the RORO /ferry ships per week corresponding to the estimated freights are derived.

References

Association of international forwarder: Handbook for international transport, Ocean commerce, Tokyo

Association of ship building and business: the investigations concerning specifications of the international voyage type TSL and relevant transport demands, 2000

Hakata harbor statistics: http://www.port-of-hakata.or.jp

Japan Ship owners' Association: Marine transport statistics handbook, 2007

Ministry of Finance: materials of trade statistics, Japanese trade statistics, 2007

Marine technology and safety bureau of the Ministry of transport: About the present situations around the international marine transport and the possibility to use TSL, 2000

The Ministry of Commerce of PRC: Total value table for import & export freight, 2007

The newspaper of transport: About JR freight's sea & rail, 2007

Workshop for the flowing trend of the import & export freight: The survey for the trend of the import & export freight. Association of Japanese duty, 2002

Ports Spatial Structure Analytical Method and Case Study

Shilin Han[1], Tian Xu[2], Tao You [3] and Liping Zhao[4]

[1] Associate Professor, [2] Master, [3]Associate Professor, [4] Professor, School of River and Coastal Engineering,Changsha University of Science and Technology, Changsha 410076, China; E-mail:hanshilin6@163.com

Abstract

Hunan Province plans to initiate larger scale construction of "3+5"city group on the basis of the Changsha-Zhuzhou-Xiangtan integration, in order to further enhance the radiation function of Changsha-Zhuzhou-Xiangtan integration, and to lead the development of surrounding areas,. The construction strategy is to make the central growth pole in Hunan Province's rising, which demands the ports of the pan-Changsha-Zhuzhou-Xiangtan "3+5" city group ports form rational port groups. The status and main functions of every port in port groups should be determined with primary-secondary order, clear levels, and complementary functions. Gini coefficient and Hofer index are utilized to make quantitative reasearch to the spatial structures of ports in the pan-Changsha-Zhuzhou-Xiangtan "3+5" city group, which could give reference for the port planning in the economic integration of "3+5" city group, fully exert the integral beneficial function of "3+5" city group ,and provide construction reference for other port groups in China.

Key Words: pan-Changsha-Zhuzhou-Xiangtan; "3+5" city group; port group; spatial structure

Introduction

Hunan Province plans to initiate larger scale construction of a "3+5"city group on the basis of the Changsha-Zhuzhou-Xiangtan integration, in order to further enhance the radiation function of Changsha-Zhuzhou-Xiangtan integration, and to lead the

development of surrounding areas,. The construction strategy is to make the central growth pole in Hunan Province's rise, which is also an engine of Hunan's development. The pan-Changsha-Zhuzhou-Xiangtan "3+5" city group takes Changsha、Zhuzhou、Xiangtan as the center, with Yueyang, Changde, Yiyang, Loudi, Hengyang as five surrounding cities, a total of 64 county-level administrative regions (including eight municipal districts), 156 streets, and 1087 township-level administrative regions. The "3+5" city group covers an area of 97,065 km^2, accounting for about 45.83% of the Hunan province. The gross domestic product(GDP) of the port group accounted for 75.7% of the Hunan province, and local financial revenues accounted for 76.6% of the Hunan province. The commuting distance among the three major cities of Changsha-Zhuzhou-Xiangtan is about a 40 minutes drive, and the commuting distance of five other cities away from the provincial capital Changsha are also within 1.5 hours. The key enterprises of the Hunan province is almost entirely concentrated in the eight cities. "The general report on construction of the "3+5" new type city group and industrial development ideas and countermeasures" reports that the "3+5" city group is the transport hub of Hunan province, and an important grain production partnership, energy raw material base, modern equipment, and high-tech industrial base. Among the "3+5" city group, the five cities of Changsha, Zhuzhou, Xiangtan, Yueyang, and Hengyang are all located on the beach of the Xiangjiang river; Changde is located in the tail of the Yuan River;Yiyang is located in the lower Zishui River valley. Because of the navigation suspension of Lian River and the navigational danger of Zhexi Dam in ZiRiver,the waterway transport of Loudi is less-developed, which constrains the rapid development of the local economy. The government is also active in the planning of local waterway transport development. The economic development of city group is closely related to waterway development, port construction, and management. In the process of construction, due to the conditions of the local government s own qualifications, ports of pan-Changsha-Zhuzhou-Xiangtan "3 +5" city group go their own way. In the aspects of resources integration and overall efficiency, they lack some considerations, and there also has been the disorderly competition phenomenon of competing and redundant construction. Overall planning of the port construction ofthe pan-Changsha-Zhuzhou-Xiangtan "3+5" city group is urgently needed for a new economic pattern of our province.

Port Group Construction Rresearch History and Background in China and Overseas

The construction history of overseas port groups is relatively longer. The present world-famous ports are almost constructed by many adjacent ports as large-scale combination port or port group, such as Japan Tokyo Bay Port group, Osaka port group, the British south coast "Port Union", located at the mouth of the Hudson

River New York - New Jersey port group and European Sea Ports Organization (ESPO) etc. These port groups have a rational division of labor and coordination, which fully reflects the complementary advantages in the port groups. Meanwhile, they also provided valuable experiences that can be used during Chinese port construction. As the formations of Yangtze River Delta, Pearl River Delta, and Bohai Bay economic zones, Yangtze River Delta port group, Pearl River Delta and Bohai Bay port group correspondingly come into being. Other places of the country are also making great efforts to strengthen the construction of the port group. Zhejiang province has accelerated the construction of the Ningbo-Zhoushan combination port; Fujian province accelerated the construction of the Xiamen hub port, and improved the Jiulong River Estuary of Xiamen Bay port system; Liaoning Province is also constructing the coast port system in order to promote the accomplishment of an economic overall mechanism in the northeast region.

Since World War II，western geographers have researched port groups (especially the competitive relationship of related ports about space), such as Germany's Ruhr industrial area port group and the U.S. Pacific coast port group. During the 1960s, American scholars Taaffe, Morrill, and Gould (1963) summarized the empirical mode of space structure evolution of the port group, and considered that the cargo flow will tend to concentrate along with development of a port freight transportation system network. Later, many scholars conducted empirical research, and Slack (1990) also considered that the development of technology will lead to goods flowing to pivotal ports or load centers. This outcome was questioned, and Hayuth (1998) researched the U.S. port system (containers), and come to the conclusion that the development of the port system will tend to disperse .The emphases of the foreign scholars research on port spatial structure are: 1) to appreciate the mutual competition of the related ports at hinterland and foreland, (the work outcome of Mayer (1957) and Weigend (1958) is a typical representative); and 2) to appreciate the changes concerning the degree of traffic concentration (Taaffe, 1963).

Domestic research in this area started relatively late. Since the 1990s, Chenhang (1997) first made an approach to the regional combinatorial about Chinese harbors. He emphasized making a special discussion and exposition about the formation mechanism, principle, conditions, and the law of development of the regional combinatorial about Chinese harbors. He considers that the main reason of port regional combinatorial is the competition between one another, and the main reason of this mutual competition is the supply of goods. The formations of port regional combinatorial are divided into five stages as follows: germination and emergence stage of the port; development stage of port; the initial development stage of port regional combinatorial; the formative stages of port regional combinatorial; the formation and continuous improvement stage of port regional combinatorial. From

the strategic perspective of cooperative development of the regional economy and construction of the port system, Li (2004) discussed the regional economic significance, condition, and basic direction of the construction of the large-scale and deep-water ports in Nantong. He also proposed the basic strategic measures for speeding up the construction. Zhang and Zhang (2001) made an analysis on the contents and significance of building the Taiwan Strait shipping port system. From an economic perspective, Chen (2002) analyzed the effect of the economic development of the Pearl River Delta port system on the Guangzhou port, Hong Kong port, and Shenzhen port. Zheng (1991) has studied the Chinese coastal port city system from the perspective of regions and cities. In their researches, the port system spatial structure and size combination became their concern. Cao (1997) has made a study on the spatial structure of Yangtze River port system, and he also depicted the state of spatial structure about the Anhui Province port system. Wang and Yang (2000) use the concentration factor of cargo flows to reflect the evolution of the port group's spatial structure.

In summary, scholars in China and abroad have researched the spatial structure of the port groups. But on the whole, domestic researches lag behind in the west. Researchers have studied each province and region on the port layout of the relevant region according to their own economic characteristics, but most of them are qualitative with a static description, and the quantitative and dynamic analyses are inadequate. They relatively lack theoretical exploration and model choice, which are in line with China's reality, and the study on pan-Changsha-Zhuzhou-Xiangtan "3 +5" city group spatial structure has not been completed.

Quantitative Analytical Method on Spatial Structure of Port Group

Two quantitative analysis methods are utilized in order to measure the spatial structure changes and strength of "3+5" city group ports in this paper.

Gini Coefficient Method. Italian statistician Gini put forward the Gini coefficient. It can quantitatively reflect the degree of concentration, which is a kind of socio-economic development indicator to measure the degree uniformity of the social income. Gini coefficient G is used to reflect the degree of polarization of wealth in a regional economy (Bian, 1997). This paper mirrored the basic principles of the Gini coefficient statistical methods to describe the concentration degree of cargo flows about the pan-Changsha-Zhuzhou-Xiangtan "3+5" city group ports, namely, the direction and intensity changes of spatial structure about the Pan-Changsha-Zhuzhou-Xiangtan "3+5" port group. This paper uses the following simplified Gini coefficient computational formula:

$$G=0.5\sum_{i=1}^{n}\left|x_i - y_i\right| \quad (1)$$

where n=the number of ports in the port group; x_i=the proportion of the port cargo throughput to the port group cargo throughput; y_i=when cargo throughput is absolutely uniform in distribution, the proportion of the port cargo throughput to the port group cargo throughput; $y_i=1/n$. $0\leq G\leq 1$, $1/n<HI\leq 1$, the closer G gets to HI, the more concentrative the space structure of the port system.

The larger the value of Gini coefficient G, the more unequal the distribution of cargo flows in the port group, and at this moment, the port group spatial structure tends to concentrate. When G = 0, it indicates that the throughput of each port is the same, and the port group spatial structure tends to strike a perfect balance (decentralized).

Concentration of Cargo Flows Coefficient Method. Hofer used the Hofer index to illustrate the issue of centralization in the 1950s doctoral thesis——"the concentration of iron and steel industry". Later, regional economics use the Hofer index to illustrate the degree of market concentration. In this paper, we define the Hofer index as the cargo flow concentration factor HI, which reflects the degree of cargo flow concentration. The calculation formula is:

$$HI=\sum_{i=1}^{n} p_i^{\,2} \quad (2)$$

where p_i =the proportion of the port cargo throughput to the port group cargo throughput; n=the number of the ports in the port group.

According to the formula we can find: $0<HI\leq 1$; when $HI\rightarrow 1/n$, the cargo flow tends to spread; when $HI\rightarrow 1$,the cargo flow tends to concentrate.

Analysis on Ports Spatial Structure of "3+5" City Group

Calculation on "3+5" City Group Ports' Spatial Structure. We can use cargo throughput, passenger throughput, fixed assets, gross output value of industry and agriculture, total profits and many other indicators as the characterization of port distribution. Cargo throughput and gross output value of industry and agriculture are comprehensive indicators. In this paper, we choose port cargo throughput as a fundamental analysis indicator, and we select "3+5" city group ports (Changsha port, Zhuzhou port, Xiangtan port, Yueyang port, Changde port, Yiyang port, Hengyang port) as the object of study. Because the port cargo throughput in the 1970s and 1980s is very low, we chose 1990 to 2003 as the basis for this research.

Table 1 shows cargo throughput of Changsha port, Zhuzhou port, Xiangtan port, Yueyang port, Changde port, Yiyang port, and Hengyang port from 1990 to 2003.

According to the method of the Gini coefficient and Hofer index, we can separately calculate and find the Gini coefficient and cargo flow concentration factor (HI), shown in Table 2, Figure 1, and Figure 2.

The dynamic characteristics and genetic mechanism.

Concentration is the overall trend of dynamic changes of spatial structure. The calculation results above (Table 1, Figure 1, Figure 2) indicated that the general trend of dynamic characteristics about "3+5" port group space structure gradually tend to concentrate. From 1990 to 2003, the Gini coefficient increased from 0.2367 to 0.2662, and the average annual growth rate was 0.0023; it reached its peak number of 0.3952 in 1997; cargo flow concentration factor rose from 0.2114 to 0.2580; the total increase was 0.0466, and the average annual growth rate was 0.0036.

In the process of port group's development, center ports often turn up due to the impact of the regional economic, consequently leading to the changes in the space structure of the port group. The internal mechanism is at the time of relative low level about regional economic productivity, and often only a small number of ports (or even individual ports) have a relatively convenient set of local acts systems, relatively advanced technology and equipment, relatively low transportation costs, and thus cargos flow to these most advantageous ports thus they become a center port.

Table 1. "3+5" city group port cargo throughput table(Unit: t).

Year	Changsha	Zhuzhou	Xiangtan	Yueyang	Changde	Yiyang	Hengyang
1990	3253597	1964060	1276218	6632511	1050396	2003566	1820698
1991	2918180	2257838	1495455	7566324	1034899	1877943	1839424
1992	3645504	2102046	2271963	9372612	1062012	1894229	1424748
1993	3926420	2207567	1542869	9652390	1081265	1552957	1396554
1994	3473197	1873366	1357003	9039091	1074076	1733861	1317102
1995	3359646	2014092	1119216	8036354	1085007	1880117	1579955
1996	3450000	1586000	1060000	8142000	1268000	1985000	1000000
1997	3874989	1313662	243281	7947099	1365522	447304	1513605
1998	4191584	1428357	1476052	6978553	1706835	1102000	1737855
2000	3110233	1675725	780600	9145800	1496000	1360000	1504600
2001	4422900	2055500	805900	7533000	1118600	1523800	1966000
2002	6918806	2386289	920144	8987997	1455281	1406000	2608046
2003	5442200	2549577	1543244	9689300	1839000	1312000	2140700

Data is from the statistics yearbook of shipping administration in Hunan Province.

Table 2. "3+5"city group port spatial structure quantitative analysis indicators.

Year	Gini coefficient	HI
1990	0.2367	0.2114
1991	0.2740	0.2248
1992	0.2854	0.2478
1993	0.3232	0.2660
1994	0.3173	0.2660
1995	0.2850	0.2429
1996	0.3144	0.2585
1997	0.3952	0.3021
1998	0.2874	0.2239
2000	0.3306	0.2828
2001	0.3030	0.2348
2002	0.3320	0.2398
2003	0.3047	0.2360

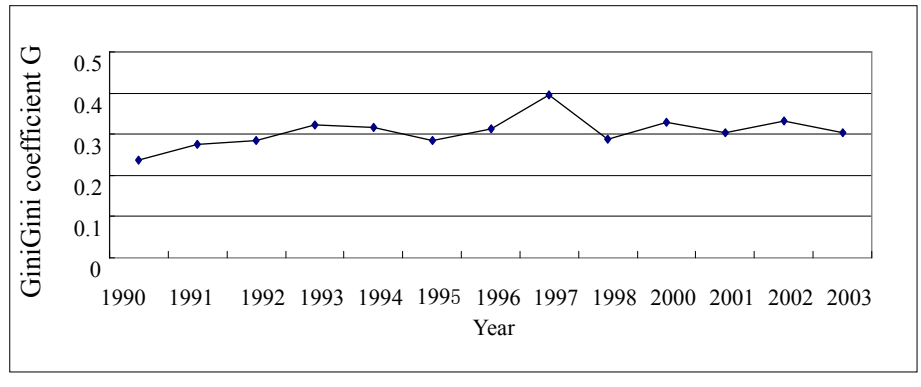


Figure 1. The Gini coefficient of "3+5" city group ports.

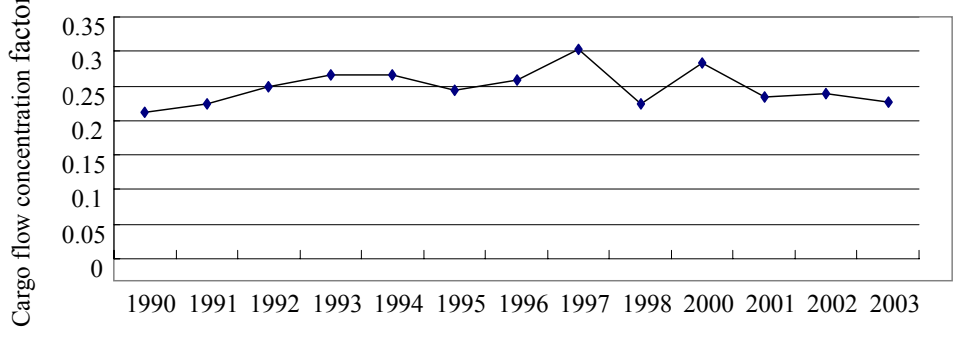


Figure 2. The concentration factor of "3+5" city group ports.

Table 3. The throughput proportion about Yueyang port and Changsha port.

Year	Yueyang port	Changsha port
1990	0.3684	0.1807
1991	0.3984	0.1536
1992	0.4304	0.1674
1993	0.4518	0.1838
1994	0.4549	0.1748
1995	0.4213	0.1761
1996	0.4403	0.1865
1997	0.4757	0.2319
1998	0.3747	0.2250
2000	0.4795	0.1630
2001	0.3877	0.2276
2002	0.3641	0.2803
2003	0.3952	0.2219

However, the developments of other ports are restricted, which lead to the cargo flow concentration.of the space structure. After the regional economics develops into a higher level, the space movements of goods are more selective because of the development of overall socio-technical technology system, and the operating expenses (such as labor costs, land-value etc.) of center ports constantly rise; the successive concentration of cargo flow will lead to a so-called "Non-economic scale." Meanwhile, other ports will pose a real challenge to the center ports, leading to the dispersion phenomenon.

On the whole, Yueyang port and Changsha port are the two most important modern ports in the Hunan Province. They both belong to the main river ports of our country. The throughput proportions of Yueyang port and Changsha port to the total throughput of "3 +5" city group are shown in Table 3. The proportion of Yueyang port's throughput to the total is 48% in 1997. And the proportion of Changsha port's throughput to the total is 28%. Yueyang port is the main river port in our country. It is located at the intersection of the Yangtze River and Dongting Lake, and along the Yangtze River we can reach the Chongqing and Shanghai port. Tracing Dongting Lake we can arrive at each of the following sites: Xiang, Zi, Yuan, Li River. The port's special railway lines connect with the Jingguang Railway, and the highways connect with the national road 107 and Jingzhu expressway; 168 km from the capital of the province-Changsha; 232 km from Wuhan, and 210 km from Jingzhou. It is the collecting and distributing centee of goods and materials of land-water transit or water-water transit, and it's the only port along the Yangtze River in Hunan with

good water depth conditions. With the accelerating economic growth of the Yangtze River Delta, which takes Shanghai as the center, west development and the development of economy in Hunan Province and Xiangjiang River water resources, the Yueyang port will develop better. From the analyses above we find that the Yueyang port is the most advantageous port in the Hunan Province, and it is also a center port, bringing about the fact that concentration about spatial structure of a "3+5" port group is the main aspect of the contradictions. As one of the 23 main hub ports in the country, Changsha port has a superior geographical location and water conditions. Also, Changsha is the capital city of the Hunan province, and the development opportunities are naturally greater than in other cities. The development of the Changsha port slowed the speed of the concentration of the "3+5" city group ports.

The process of decentralization is continued on the basis of the low level of space concentration. The evolution trend of port group space structure is a large-scale phenomenon .A lot of port system space structure over the world has gone through this evolution process, however, the centralized starting points are very different from each other. According to the calculation of the relevant information, in 1950, the Gini coefficient of the U.S. Great Lakes port systems was 0.516 and then gradually increased (Mayer 1957). But the highest Gini coefficient of a "3+5" port group, 0.395, occurred in the 1990s, which indicated that the concentration degree on space structure of "3+5" port group is rather low, also showing that the space structure of a "3+5" port group has room for further concentration. In recent years, shipping logistical economics of Yueyang has grown strongly. In 2008, the cargo throughput of Yueyang port reached 55 million tons, accounting for 1/3 of the entire port group, which is the eighth in the 28 major inland river ports. The total cargo throughput of the Changsha port in 2008 was 29.55 million tons, which increased 41.54% in 2007. Growth was experiences even during a year with a rarely seen ice disaster and the financial turmoil of the economy. The Changsha Xiangjiang River Integrated Hub project is planning a total investment of 5.732 billion yuan. This project commenced in October 2009. After its completion, following Hengyang, will be a 2000-ton channel, which will greatly enhance the navigation capacity of the Xiangjiang River, and the features and benefits of the Changsha port will also improve.

The velocity of concentration is very slow. From 1990 to 2003, the Gini coefficient of the "3+5" port group increased from 0.2367 to 0.2662. During these 13 years, the total increase of the Gini coefficient was 0.02975, and the average annual growth rate was 0.0023. The increase of cargo flow concentration factor was not much, respectively 0.0466 and 0.0036 in the 13 years, and the growth rate was slow.

Also, the implementation of the "3+5" city group economic integration is still in its initial stage. We are not responsible for the overall planning of the "3+5" port group. Therefore, each port is usually built based on its own needs of development, and it is difficult to overcome the department and regional fragmentation, each section having its own ideas. Therefore, at the rapid development of the Changsha port and the Yueyang port, the Zhuzhou port and the Xiangtan port are also accelerating the speed of the constructions, which slow up the speed of centralization of "3+5" city group.

In the process of agglomeration, spatial structure has a significant volatility. Shown in Figure 1, during the 1990s, the direction and intensity of spatial structure changes about "3+5" city group ports were obviously volatile.The dispersion phenomenon appeared four times, however, the speed and magnitude of the dispersion was not large. The general trend about dynamic changes of spatial structure of the pan-Changsha-Zhuzhou-Xiangtan port group gradually tends to concentrate. The Gini coefficient of the port group has increased from 1990 to 1997. The increase was rather slow from 1990 to 1995, from 0.2367 to 0.2850. After some of the ports changed from collective ownership to state ownership in 1995, the Gini coefficient ascended to a level markedly higher than before, and it reached its peak number of 0.3952 in 1997. Due to large floods at the Yangtze River basin in 1998, goods are basically transported to the neighborhood ports, and there is a greater decline in the Gini coefficient. There were small fluctuations from 1998 to 1999, however in 2003, the Gini coefficient had a greater degree of decline due to SARS. Because the concentration level of the spatial structure of "3+5"port group is not high enough, some dispersion phenomenon in the process of spatial structure evolution of port group appears in the context of lower concentration. Figure 2 also verifies the volatility of the gathering process.

Conclusions and Recommendations

Gini coefficient and cargo flows factor methods are utilized to make quantitative analyses on the throughput of "3+5" city group ports (Changsha port, Zhuzhou port, Xiangtan port, Yueyang port, Changde port, Yiyang port, and Hengyang port),which indicated that cargo gradually tends to concentrate on the Yueyang and Changsha ports. Despite some individual or special reasons, there are always fluctuations in the whole process of spatial structure evolution of the port group, and centralization is the main characteristic of spatial structure of pan-Changsha-Zhuzhou-Xiangtan "3+5" port group. Concentration is the overall trend of the port group's development. Along with the promotion of the "3+5" city group construction, which takes Changsha-Zhuzhou-Xiangtan as the center, Pan-Changsha-Zhuzhou-Xiangtan "3+5"city group ports should be formed as rational port group. "The planning of

Changsha-Zhuzhou-Xiangtan city group area" has made a plan for the ports of Changsha, Zhuzhou, and Xiangtan ports, however, this plan is mainly put forward in light of developing each port. Due to the lack of overall planning for the port group, each port does not systematically take into account the impact among all the ports. Under the good situation of Changsha-Zhuzhou-Xiangtan economic integration, it is proposed that the port construction of "3+5" city group should fully consider the impact among ports and put forward a reasonable and feasible "3+5"city group overall port plan according to its own characteristics; make a further improvement to the port plans that have been completed or are currently underway; give full play to the advantages of each port to achieve the greatest benefits of the port group; make a greater contribution to becoming a strong province and promoting rapid development.

Acknowledgments

This project was supported by the Scientific Research Fund of Hunan Provincial Education Department in China (No. 09B005).

References

Bian, Z. (1997). "The market economy statistics", *Shanghai University of Finance Press*, 325-332.

Cao, Y. (1997). "Preliminary study of Anhui Province port system along the Yangtze River]", *Geographical Science*, 154-162.

Chen, H. (1997). "The formation mechanism and the development process on the combinations of the harbor area", *Geography Journal*, 501-507.

Chen, Z. (2002). "Analysis on the Pearl River Delta port cluster development situation-the development strategy of Guangzhou harbor", *Jinan Journal (Philosophy and Social Sciences)*, 42-47.

Hayuth, Y. (1998). "Rationalization and deconcentration of the U.S. container port system", *The Professional Geographer*, 279-288.

Li, J. (2004). "Strategic thinking of Nantong coastal harbor development and construction", *Economic Geography*, 72-75.

Mayer, H. M. (1957). "The port of Chicago and the St. Lawerence Seaway", *Department of Geography research paper No. 49 Chicago: University of Chicago.*

Slack, B. (1990). “Intermodal transportation in North America and the development of inland load center”, *The Professional Geographer*, 72-83.

Taffee, E. J., Morril, R. L., and Gould, P. R. (1963). “Transportation expansion in underdeveloped countries”, A comparative analysis.” *Geography Review*, 503-529.

Wang, C., Yang, G., Shi, S.. (2000). “The space structure evolution characterics of the port system in middle and lower Yangtze River”, *Resources and Environment of Yangze River Basin*, 283-291.

Weigend, G. (1958). “Some elements in the study of port geography”, *Geographical Review*, 185-200.

Zhang, X., Zhang, M., Chen, K. (2001). “The analysis on construction of the Taiwan Strait shipping port system”, *Guangdong Institute of Public Administration Journal*, 72-76.

Zheng, H. (1991). *The Exploration on Port Cities*, Nanjing:Hehai University Press.

A Study on the Development of Urban Environment Measure's Evaluation Model

Tetsuo Morita[1], Hiroshi Sugita[2] and Shengchuan Zhao[3]

[1] Associate Professor, Department of Civil Engineering, Gunma National College of Technology, 580 Toriba-machi, Maebashi-shi, Gunma, 371-8530, Japan
[2] Chief Researcher, The Institute of Behavioral Sciences, 2-9 Ichigaya Honmura-cho, Shinjuku-ku, Tokyo, 162-0845, Japan
[3] Prof., Center for Transportation Research, Dalian University of Technology, Dalian 116024, China

Abstract

This paper develops a comprehensive model (MERS) to evaluate urban environmental measures which contribute to an eco-friendly city. The characteristic of MERS is a practical model which is based on Person Trip Survey data. MERS has two sub-models, Urban Activity Model and Evaluation Model. Urban Activity Model consists of Land Use Model and Transportation Model. Land Use Model estimates the volume of employee and populations. Transportation Model estimates transportation demand by mode, and car traffic volume on each road links. Evaluation Model consists of environmental evaluation model, quality of life evaluation model and economic evaluation model. Evaluation results can be represented on mesh unit or road network links. This paper will explain the framework of MERS and the results of a case study.

Introduction

Background and Aim. Greenhouse-gas problem is getting serious, and the damage by the abnormal weather like floods occurs at many places in the world. The researchers in the world conduct various studies to solve such an environmental problem.
This paper developed a comprehensive model to evaluate urban environmental policies and measures which contribute to an eco-friendly city.

There is a number of researches regarding to the development of the model to evaluate the effect of the urban environmental measures quantitatively. And there are these measures for various fields such as a transportation measure, land use measure, private sector measures. Therefore the model which can evaluate an effect by the measure from various points such

as a global environment, atmosphere environment, quality of life, generally is necessary. This research is trying to develop a model system which can evaluate the environmental measurers comprehensively.

This paper will explain the framework of the proposed model, i.e. Models to Evaluate Regional Renovation for Sustainability (MERS) and the results of a case study.

Target Area and Database. In this study Sendai Metropolitan Area (SMA) is target areas. SMA consists of 5 cities, 8 towns and 1 village. The population is 1,458 thousands in SMA and 1,030 thousands (approximately 71 percent) in Sendai city in 2008.The following model calibration uses data from Person Trip Survey conducted in 2002.

MERS's Characteristic and Structure

The Characteristics of Our Model. There are many models such as TRANUS, URBANSIM, MUSSA and TLUMIP, and those models are used for urban planning (Wegener, 2003). In Japan Person Trip Survey is carried out continuously on many urban areas. By Person Trip Survey daily movement of persons is understood. The characteristic of MERS is a practical model which is based on Person Trip Survey data. MERS has two characteristics. The first characteristic is that MERS can evaluate various environmental measures such as traffic measures, land use measures and private sector measures. The second characteristic is that MERS can evaluate from the view of environment, quality of life (QOL) and economy. Beside our model can display the result visually.

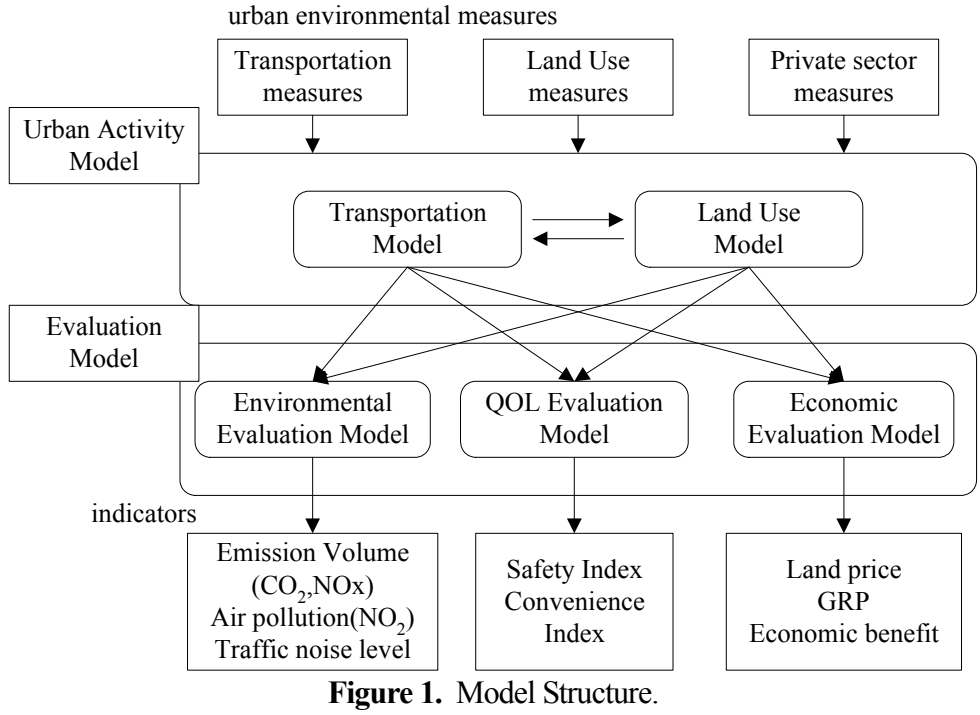


Figure 1. Model Structure.

The Outline of Model Structure. MERS has two sub-models, Urban Activity Model and Evaluation Model. Urban Activity Model consists of Land Use Model and Transportation

Model. And each variable in Land Use and Transportation Model is related mutually. Land Use Model allocates the total number of employee, populations for the each zone. Transportation Model estimates transportation demand by mode, and car traffic volume on each road links. Evaluation Model consists of environmental evaluation model, quality of life (QOL) evaluation model and economic evaluation model. Evaluation results from the 3 evaluation models can be represented on mesh unit or road network links (Figure 1).

Urban Activity Model

Framework. MERS have a dynamic structure shown in Figure 2, where discrete time intervals are represented as t=1, t=6, t=11(5-year interval), and so on. The land Use Model and Transportation Model influence each other through time.

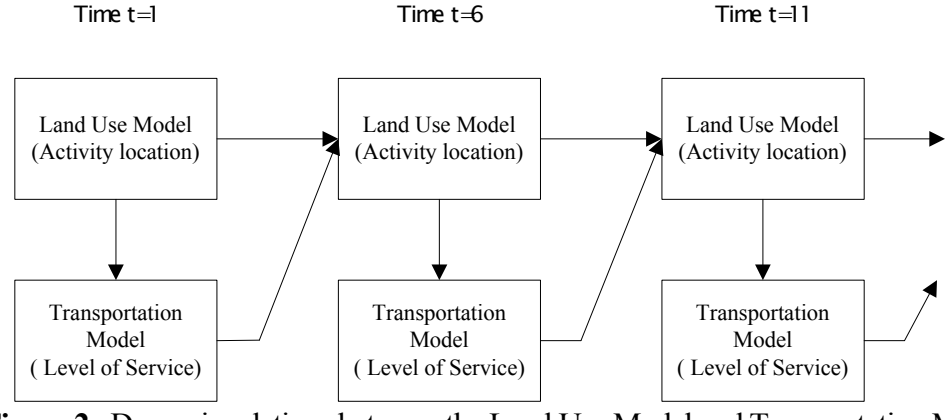


Figure 2. Dynamic relations between the Land Use Model and Transportation Mode.

Land Use Model. Many land use models were developed in Japan (Taniguchi et al, 1983 / Miyamoto et al, 1989 / Ueda et al, 1993). Land Use Model estimates the location of population and employee. And the size of location volume is determinate by accessibility, land price and space limitation.

Figure 3 is shown the relationship among the variables that determined the location of employee or population. The location volume of population is determinate by accessibility of commuter trip or shopping trip, land price and space limitation (population density). The size of population becomes large so that accessibility is higher. But size of populations become small so that land price level is higher, or the degree of space limitation is larger.

When population increases, the population density becomes higher. And the power to control population acts. In this way Land Use Model has a disincentive. Otherwise the location volume of employee is determinate by accessibility of business trip.

Land price level is determinate by accessibility of commuter trip or shopping trip and zoning. Accessibility of commuter trip or shopping trip becomes large according to an enhancement of transportation service level or increase of employee.

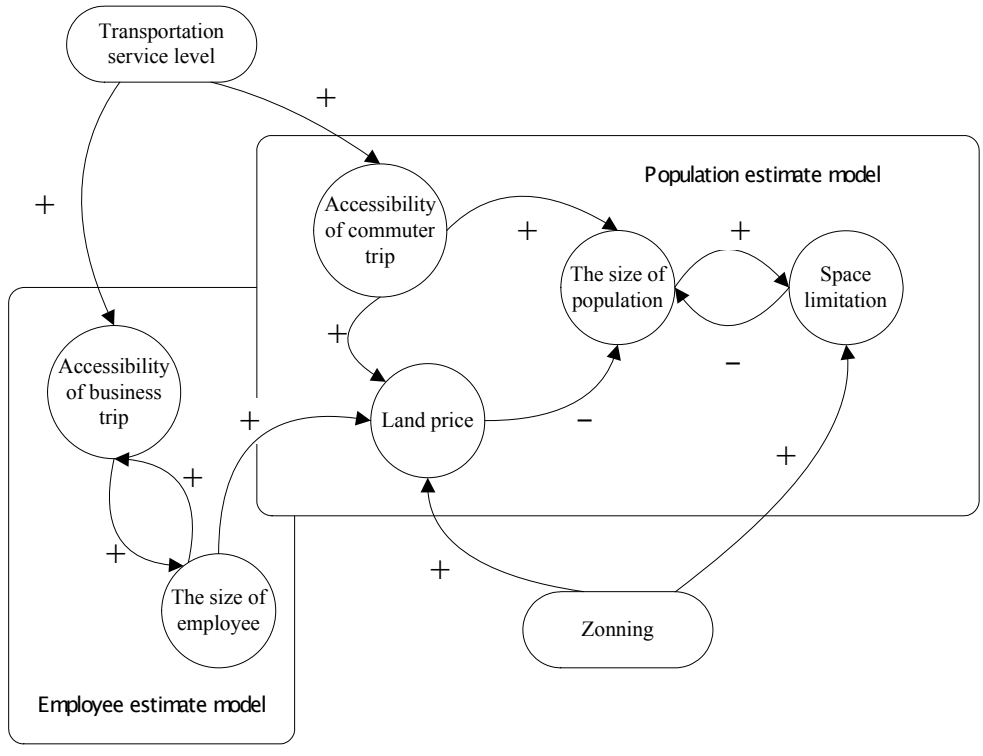


Figure 3. Interrelationship among variables (Land Use Model).

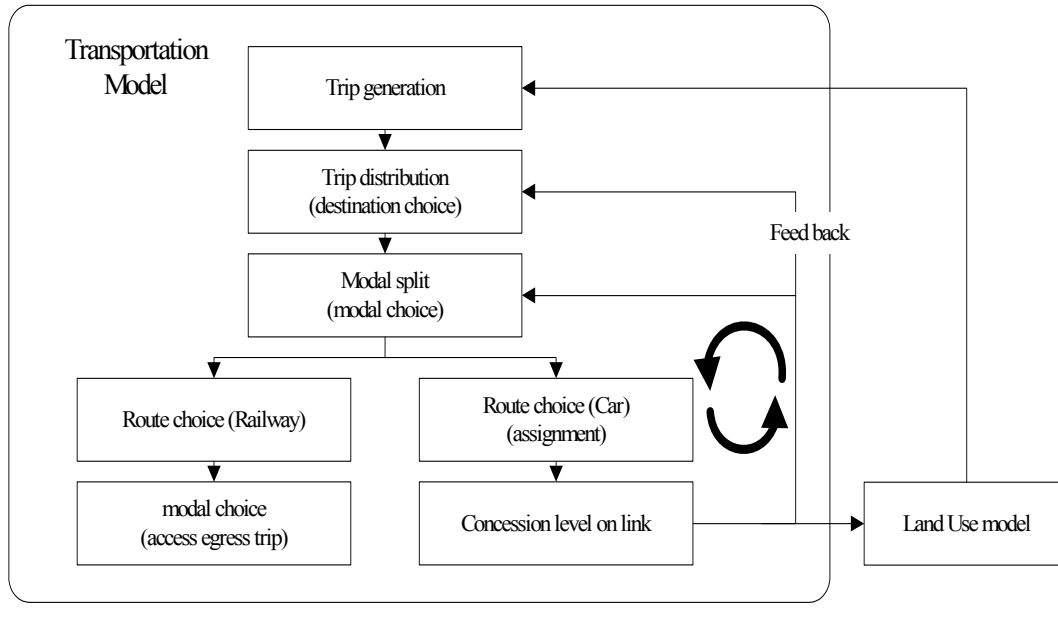


Figure 4 Dynamic relations between the Land Use Model and Transportation Model.

Transportation Model. Figure 4 shows the structure of transportation model. Transportation Model is the Travel Behavior Model which represents a series of travel decisions such as destination choice, modal choice, route choice and access egress mode choice by Nested Logit Model (Kojima, Yoshida and Morita, 2003).

Table 1. Level of service valuables used in travel behavior model.

Access Egress modal choice model	
Walk	Travel time on foot
Bicycle,Motorcycle	Travel time by bicycle or motorcycle
Bus	Riding time on bus, Waiting time and walking time
Kiss & Ride	Travel time by car
Park & Ride	Travel time by car, Parking fare, Capacity of parking facilities within 150m from station
Rout choice model	
Railway	Travel time on train, Access egress time and waiting time, Expected maximum utility associated with choosing access egress mode.
Modal split model	
Walk,Bicycle,Motorcycle	Travel time on foot
Car	Travel time by car, Parking capacity per employee, Area dummy
Bus	Riding time on bus, Waiting time and walking time
Railway	Expected maximum utility associated with rail route choice
Destination choice model	
Zone Characteristic	Size of zone area, Employee density, Non-employee density. Employee (Secondary industry) density, Employee(Tertiary industry) density, Area dummy (Business district, mall area), Expected maximum utility associated with modal choice, Size of population (Origin zone), Size of employee(Origin zone)

Table 2. Evaluation index of the Evaluation Model.

Model	Sub Model	Evaluation index	Space unit
Environment	CO_2 Emission	CO_2 Emission volume	zone, mesh
	Air Environmental	NO_x Emission volume	zone, mesh
		NO_2 Concentration	mesh
	Road-Noise Propagation	Road Noise Propagation level	zone
Quality of life	Quality of life	Safety index, Convenience index	zone
Economic	Land price	Land price	zone
	GRP	GRP	zone
	Economic benefit	Economic benefit index	zone

Car used trip which is estimated by travel behavior model is assigned on road network. Traffic volume and traffic congestion level on links is estimated. This traffic congestion level is feedback to destination choice process and modal split process till equilibrium status is obtained. The level of service valuables used in travel behavior model is shown in Table 1.

Evaluation Model

Framework. Evaluation Model consists of 3 models, i.e. environmental evaluation model, QOL evaluation model and economic evaluation model.

These models calculate various indexes using the output of Urban Activity Model, i.e. population, employment, traffic volume so on. Table 2 shows the evaluation index that is calculated by evaluation model.

Environmental Evaluation Model. Environmental Evaluation Model consists of CO_2 Emission Estimation sub model, Air Environmental Evaluation sub model and Road Noise Propagation sub model.

Emission estimation sub model. CO_2 Emission volume is estimated according to discharge from car and house/office facilities. And emission volume from car is calculated from Eq. (1) and Eq. (2).

$$Emit^{car} = \sum_i \sum_v \sum_t (Q_{t,i}^{veh} \times EF_{t,i}^{veh}) \quad (1)$$

$$EF_{t,i}^{veh} = \frac{a_1}{v_t} + a_2 \times v_t + a_3 \times v_t^2 + a_4 \quad (2)$$

$Emit^{car}$: CO_2 Emission volume from car

$Q_{t,i}^{veh}$: Traffic volume (vehicle distance)

$EF_{i,j}^{veh}$ CO_2 Emission volume per vehicle distance (km)

v_t : Velocity

t : time interval, veh : vehicle type, i : zone

The number of population and employee is translated to the number of household and office. By these value CO2 Emission volume from house and office is calculated from Eq.(3) and Eq.(4).

$$Emit^{house} = \sum_i \sum_h (H_i^h \times (\sum_f E_f^h \times C_f)) \quad (3)$$

$$Emit^{office} = \sum_i (O_i \times (\sum_f E_f^{office} \times C_f)) \quad (4)$$

$Emit^{house}$, $Emit^{office}$: CO_2 Emission volume from household and office

H_i^h: The number of household

O_i: The number of office

E_f^h: Energy consumption per household

E_f^{office}: Energy consumption per office

C_f: CO_2 Emission volume per Energy consumption one unit

h : House type, f : Fuel type, i :mesh

Air environmental evaluation sub model. Air Environmental Evaluation Model consists of Inventory Model, Meteorological (Weather) Model and Diffusion Model. These Models

are authorized by Ministry of the Environment of Japan. Inventory Model estimates NOx emission volume by mesh unit. And it becomes the input data of the Diffusion Model.

A diffusion model is a model estimating diffusion to the atmosphere of the quantity of NO2 estimated in inventory model.

Road noise propagation sub model. Road Noise Propagation Model is an energy based Model. Road Noise Propagation Model estimates the noise level of road edge which is occurred by automobile. The noise level is different by road structure and vehicle speed and traffic volume. This model is built based on the spread theory of the sound.

Quality of Life (QOL) Evaluation Model. Quality of Life (QOL) Evaluation Model has a covariance structure which can treat integrally personal attribute, environmental characteristic and life satisfaction (Yoshida et al, 1998).

The output of QOL Evaluation model is the index such as the convenience and safety, and this is the potentiality variable in the covariance structured model.

The covariance structured model converts the objective data such as a personal attribute or the local environmental characteristic into the abstract variable convenience, an abstract variable such as the convenience or safety and connects these abstract variables with subjectivity variables such as the satisfaction for the life. And then the structure of urban areas about QOL is clarified by these analyses. Table 3 shows the individual attribute and the environmental characteristic which is used in the covariance structured model.

Table 3. The individual attribute and the environment characteristics.

individual attribute	individual	age dummy(over60=1), student dummy(student=1), car ownership dummy(possession=1)
	housing	house flower space per capita, floor area ratio of house, owner occupant dummy(occupation=1), apartment house dummy(apartment house=1)
environmental characteristic	accessibility	travel time to down town by car, travel time to the nearest kinder garden on foot, travel time to the nearest elementary school on foot, travel time to the nearest junior high school on foot, travel time to the nearest general hospital by car, travel time the nearest railway station by car, travel time to the nearest block park. on foot, travel time to the nearest refuge on foot, travel time to the nearest shopping center by car
	transportation	the length of road (width is 3-5.5m) per zone area, bus service frequency at the nearest bus stop/distance to the nearest bus stop
	topography	average altitude above the sea level, farmland area per zone area
	building	total floor space per zone, average number of floors, floor space of factory per total floor space, population density, average growth ratio of population

Economic Evaluation Model. Economic Evaluation Model is a general equilibrium model. This Model estimates land price, gross regional product (GRP) and economic benefit index. For details of the Model(Hino, Yoshida, Morita and Semoto, 2004).

Simulation

Simulation Case.

Urban environmental measure. There are many urban environmental measures, such as transportation measures, land use measures and private sector measures. Table 4 shows the measures that can estimate effect / influence of measures in MERS. We calculate how change the environmental status, QOL and economy of Sendai Metropolitan Area (SMA), when these measures are introduced in SMA.

The choice of the measures. Land use measures (Urban structure). In most of Japanese urban city there is business district in the center. And residential area is located on around the center. And many people who work in center commute to there from suburban residence area daily.

Table 4. . Urban environmental measures.

	measures	content
Transportation	Road network	Road construction & maintenance
	Public transport	Grade up rail or bus service level
	TDM	Control of parking lots in CBD
Land Use	Trend type	Nothing
	Downtown area's residence increasing type	Promotion of the downtown area residence
	Sub urban center development type	Development sub urban center
Private sector	Housing	The introduction of saving energy house
	District heating and cooling (DHC)	Introducing district heating and cooling
	Tree planting in the city.	The roof tree planting of office building
		Creating the tree planting park in city area

If the scale of the city becomes large, business function on center district becomes large and the residential area expands to distance area. By these phenomena, many problems occur such as long-distance commuting or the traffic congestion. When we consider solving these problems from urban structure change, there are two methods. One is a policy to change the residential location from suburbs to inner city positively. Another is a policy to develop a sub center in suburbs and move part of business function concentrated on the center.

Table 5. Simulation case (urban structure).

Case	Urban structure
I	Trend type
II	The downtown area's residence increasing type (Type A)
III	Sub urban center development type (Type B)

At first we simulate these two policies using MERS. For comparison with these two policies, we simulate the case non policy is done (Table 5).

Combination measures: Transportation and Land Use. There are many urban environment measures. We carry out appropriate measures to make an eco-friendly city. Two cases are examined, the first case only has one measure alone, and the second case has several measures combined. If a combination is good, the second shows a bigger effect than the first. On the other hand, when the combination of the measures is bad, not only an effect as we expected is not appeared but also a problem that we didn't expect occurs.

Therefore we confirm how an evaluation index of MERS changed by the combination of land use measure and transportation measure.

Transport measures are road network investment, public transport service and the control of parking capacity. Table 6 shows the simulation case.

Result.

Land use measures (urban structure). The result of simulation is shown Figure 6 by radar chart. We display the deviation value[(1)] of evaluation value which can be compared with different evaluation index scale. We adjusted these values so that the outside of the chart became the high evaluation.

Table 6. Simulation case (urban structure & transportation measures).

Case	Urban structure	Transportation measures
II-1	Downtown area's residence increasing type (Type A)	Level up of public transport service
II-2		Level up of public transport service + Parking capacity control in CBD
III-1	Sub urban center development type (Type B)	Level up of public transport service
III-2		Level up of public transport service +Ring road construction

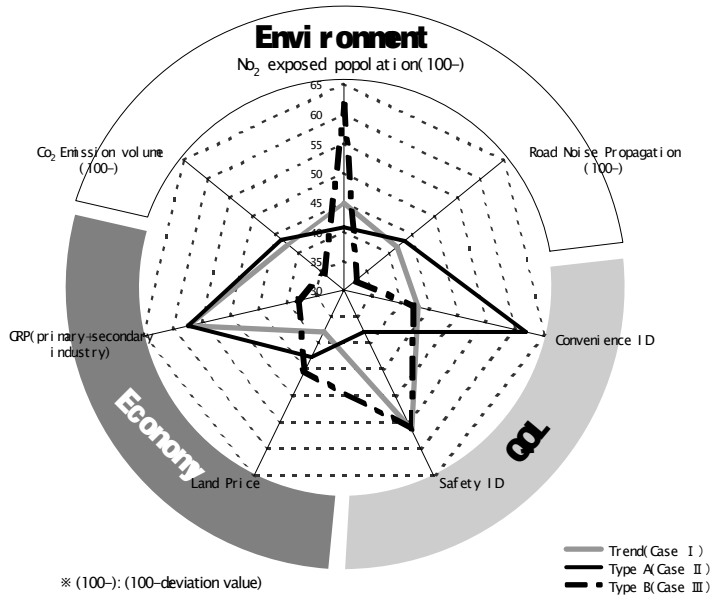


Figure 6. Result of simulation (Land use measures).

As for types A (Case II) and B (Case III), they have an opposite result. When type A (Case II) compares with trend type (Case I), the volume of CO2 emission and the area size of Noise Propagation decreases while Quality of Life and property value increase. Otherwise the number of NO2 revelation population increase, the safety of life decrease. When type B (Case II) is compared with trend type (Case I), the volume of CO2 discharge and the area size of Noise Propagation increase. Otherwise the number of NO2 revelation population decrease.

Combination measures: transportation and land use. The result of simulation is shown in Figure 7 by radar chart. For the transport measures, when the public transport service is improved, if Land use is Type A, the volume of CO_2 discharge and the area size of Noise Propagation decrease. And the property value increase (Case II, Case II-1). But if Land use is Type B the volume of CO2 discharge and the area size of Noise Propagation decrease while the property value doesn't increase (Case III, Case III-1).

The parking capacity control in CBD of Land Use Type A is effective in the reduction of the CO2 emission volume (CaseII-1, Case II-2), But In the case of Land Use type B the construction of ring road generate detour Traffic and the volume of CO2 emission increase (CaseIII-1, Case III-2).

From simulation results, it is confirmed that Type A (Land use) is suitable for reducing CO2 emission volume and increasing Quality of life (QOL), and Type B (Land use) is suitable for reducing the number of NO2 revelation population and increasing Safety of life..

In other words, under the same traffic measure, environmental effects are different by land use type.

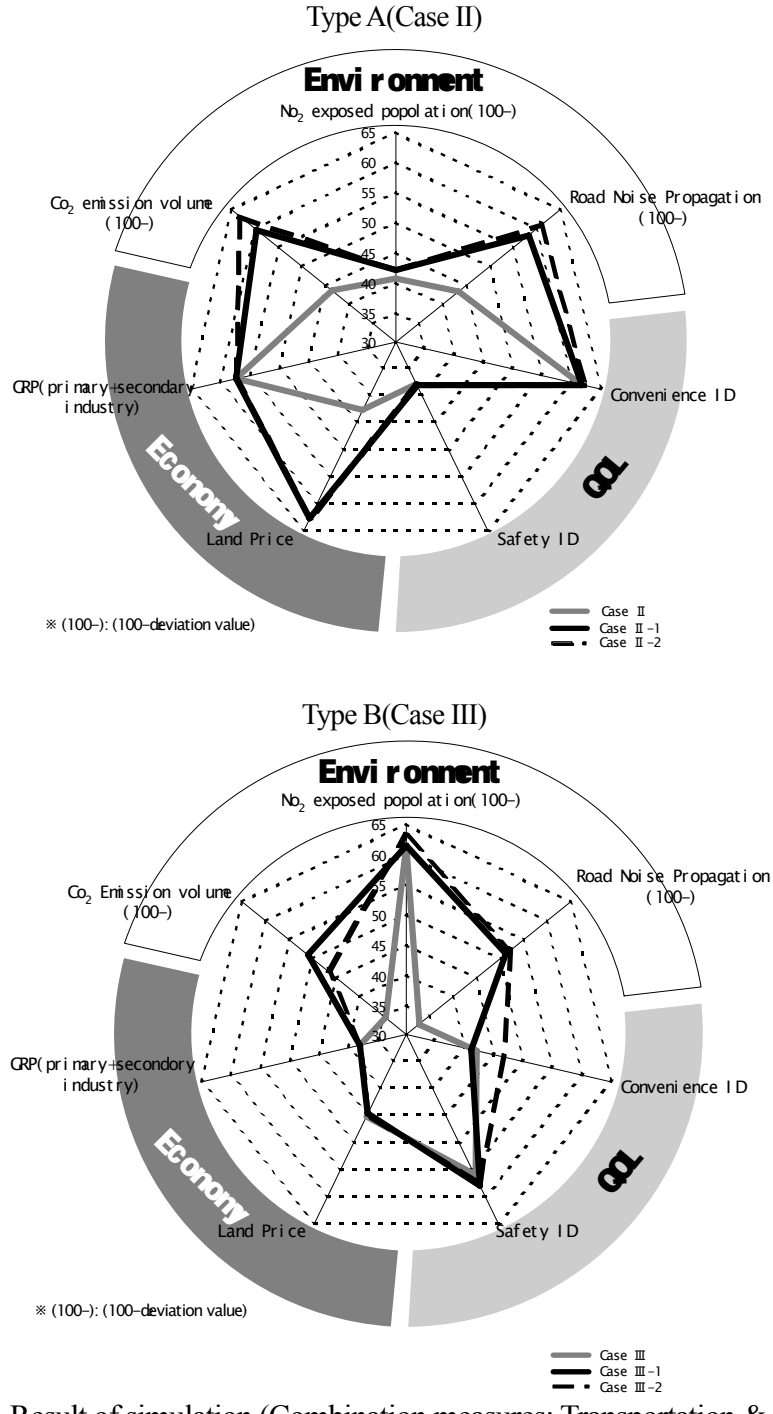


Figure 7. Result of simulation (Combination measures: Transportation & Land Use).

Conclusion

The proposed model in this paper can represent the urban activity change by land use measures and transport measures. The measures can be quantitatively evaluated by 3 evaluation models, i.e. Environmental Model, QOL model and Economic Model.

The proposed model can be also used to evaluate private sector measures such as housing measures, district heating and cooling (DHC) as well as tree planting in the city.

Note

deviation value=$(10 \times (x - \mu) / \sigma) + 50$

x : evaluation value, μ : average of evaluation value, σ : standard deviation

Reference

Hino, S., Yoshida, A., Morita, T. and Semoto, H. (2004). "A study of quantitative analysis on economic impact on the policy for urban structure and transport.",*Studies in Regional Science*, *The Japan Section of the Regional Science Association*, 34(3), 355-366.

Kojima, H., Yoshida, A. and Morita, T. (2003). "A study of urban structures and transportation programs to reduce traffic and environmental load.",*Journal of the City Planning Institute of Japan*, 38-3,553-558.

Miyamoto, K. and Kitazume, K. (1989). "A land-use model based on random utility / rent-bidding analysis (RURBAN).",*Selected Proceeding of The Fifth World Conference on Transport Research*,4, 107-121

Nakamura, H., Hayashi, Y., Miyamoto, K. and Uchiyama, H. (1983). "Computer aided land use - Transportation Analysis System (CALUTAS) for evaluation of infrastructure planning.",*Selected Proceeding of The Third World Conference on Transport Research*, 620-632.

Wegener, M. (2003). "Overview of land-use transportation models.",*The 8th International Conference on Computers in Urban Planning and Urban Management*, 20-40.

Ueda, T., Nakamura, H., Hiratani, K. and Tsutsumi, M. (1993). "Land use model based on the general equilibrium of land and building markets.", *Proceeding of International Conference on Land Problem and Urban Policy*, 183-198.

Yoshida, A., Suzuki, J. and Hasegawa, R. (1998). "Neighborhood-relevant measurement of quality of life.",*Journal of the City Planning Institute of Japan*, 33,37-42.

Cost Structure of Urban Bus Operations in Yangon, Myanmar

Hironori Kato[1], Nozomi Saito[2], Akihiro Inagi[3] and Phyo Thet Thet Htun[4]

[1] Associate Professor, [3] Master Student, Department of Civil Engineering, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan; E-mail: kato@civil.t.u-tokyo.ac.jp
[2] Overseas Consulting, Administration, Transportation, Department Division, Development Planning Department, Nippon Koei, Co. Ltd. 4 Kojimachi 5-chome, Chiyoda-ku, Tokyo 102-8539, Japan
[4] Doctoral Student, Department of Civil Engineering, Yokohama National University

Abstract

This paper analyzes the cost structure of urban bus operations in Yangon, Myanmar. The data regarding the revenues and costs of local bus operators are collected during the interviews with the bus companies, the bus control committees and individual bus owners. The operating cost includes the fuel cost, labor cost, maintenance cost, office cost and others. Then, the operating cost functions are estimated from the statistical approach. The results show that the marginal cost with respect to vehicle-km in the city-bus individual operators is much higher than in the truck-bus individual operators; that the city-bus individual operators have higher marginal cost with respect to vehicle-km than the truck-bus individual operators; and that the marginal cost with respect to number of vehicle for the city-bus individual operators is five-time higher than that for the truck-bus individual operators.
Key Words: urban bus; cost structure; Yangon; Myanmar

Introduction

Yangon is the country's largest city and the most important commercial center. As of 2007, Yangon City has about 4.1 million people in the area of 598.76 km^2 (United Nations, 2007). It grew rapidly both in terms of urban population and its physical scale. The City's population has been increasing at 15.1 % of the annual growth rate (United Nations, 2007). The city grew over the years radiating out to the outer city areas and suburbs along the main roads. The east west expansion of the City was constrained by the Hlaing River and the Pazundaung River. Government iniatives to

decentralize development started with the old suburbs of Okalapa (North and South) and Tharketa and more recently the development of New Towns in Dagon Myothit area, Shwepyithar and HlaingTharyar (Institute Sultan Iskandar and Malaysian Indisutry Government Group for High Technology, 2005). As the City grew, the travel demand, particularly the commuters also have also been growing rapidly.

There are mainly four types of transportation modes available in Yangon: private car, rail, taxi and bus. According to Zhang et al. (2008), the modal shares of private car, rail, taxi and bus are 3 %, 84 %, 6 % and 5 % respectively as of 2003. The vast majority of urban transportation relies on the bus transportation in Yangon. This is because of the difficulties of the car ownership, the low quality of rail service, and the regulation of motorbike and para-transit service in the central business district. Nevertheless, the urban bus transportation has a number of problems. Kato et al. (2010) point out the poor bus vehicles, inadequate bus network, unsatisfactory provision of service information, lack of financial support system and deficiencies in bus regulation system. These problems cause the traffic congestion on the main roads, the low traffic safety, the low quality of bus service, the high bus operation cost and the poor accessibility of the low-demand areas in Yangon.

This paper focuses on the cost structure of bus operators in Yangon. To the best of the authors' knowledge, this is the first attempt to analyze the cost structure of urban bus operations in Myanmar. Our research question is "how much does the operating cost vary among the different types of bus operators in Yangon?" As Kato et al. (2010) shows, the vast majority of bus service is operated by a number of small-scale individual bus owners in Yangon. This may lead to the high operating cost of bus service because the economies of scale do not work in the small-scale operators. This paper collects the empirical data of various cost factors during the interviews with the bus operators in Yangon and analyzes them by estimating the cost functions. Then, the cost functions are compared among different types of bus organizations. In our research, we survey the urban bus transportation in Yangon in the cooperation with the Government of Myanmar. Our research is a part of the study formulated in the Urban Public Transport Policy Framework (Kuranami et al., 2007). We interview the bus-industry regulators, the bus control committees, individual bus owners, and bus workers to collect the data. The local interview surveys were made during September 15 to 23, 2008 and during November 9 to 21, 2008.

This paper is organized as follows. The next section shortly summarizes the bus operation system in Yangon. Then, Section 3 reports the revenues and costs of bus operators on the basis of interview results. Section 4 presents the statistical analysis of cost functions of bus operations in Yangon. Finally, the achievements are summarized and the further research topics are shown in the final section.

Bus Operation in Yangon

Main Actors of Urban Bus Service Market in Yangon. There are the two types of local bus services in Yangon: the bus service given by bus companies and the bus

service controlled by the bus control committees (BCCs). The bus companies are the large-scale private bus operators who own the bus vehicles, hire the drivers/conductors and operate the urban bus service in Yangon. The two bus companies are now supplying the bus service in Yangon: the Golden City Link Co. (GCL) and the Union of Myanmar Economic Holdings Limited (UMEHL). The BCCs are the non-profit organizations which control the small-scale individual bus owners. The individual bus owners own the bus vehicles and lease the bus vehicles to the drivers and conductors. The individual bus owners must belong to one of the BCCs. There are a number of individual bus owners in Yangon. The bus operation and the fare are permitted by the Yangon Division Peace and Development Council. No subsidy program for the bus operation is given by the Government. No strategic urban transportation plan has been made officially so far.

Types of Bus Vehicles in Yangon. According to our observations and interviews with the local people, the local buses can be categorized into the following six types from the viewpoints of their function, design and capacity: city bus (air-conditioned); city bus (non-air-conditioned); mini bus; Dyna/Canter; Hilux; and other. Dyna/Canter and Hilux are the trucks which are redesigned into the passenger-use by adding the roofs and seats on the body. They are widely called the "truck bus" by local people in Yangon. The majority of buses in operation are the truck buses. Note that the air-conditioned bus accepts only seated passengers. The bus vehicles operated by the two bus companies are the "city bus" only. The UMEHL owns both the air-conditioned city buses and the ordinary city buses while the GCL owns only the city buses. The vehicles controlled by the BCCs are mainly the small-scale bus vehicles, including the Mini bus, Dyna/Canter, and Hilux.

Revenues and Costs of Bus Operators in Yangon

Revenue and Cost of BCCs. Table 1 shows the commission fee and cost which are collected in the interviews with BCCs. The main resources of BCCs are the commission fee collected from the individual bus operators. BCCs usually determine the commission fee by considering the fare revenue of individual bus operators. It accounts some 2 percent of the total fare revenue of each individual operator. BCCs pay simply the rent of their office and the salary of officers. The rent of office, including some types of costs such as the document cost, seems to be constant per month. The salary of officers is also almost constant per month.

Revenue and Cost of Individual Bus Owners. Table 2 shows the daily fare revenue and the average fare revenue per vehicle which are collected in the interviews with the individual bus owners. The individual bus owners earn only from their fare revenue. It shows that the daily fare revenue varies among the individual bus owners. There are the two reasons. The first is because the difference in the capacity of bus vehicle causes the difference in the number of passengers. The second is that the potential demand varies among different bus routes. It should be noted that the interviewees do not have any incentive to report the correct revenues to us. As they are required to pay the commission fee on the basis of total revenue, they have the

Table 1 Commission fee and cost of bus control committees

Bus Control Committee	Commission fee	Cost (Kyats per month)		
		Rent of offices	Salary of officers	Other cost
Western District BCC	2% of total revenue of individual operators	5,000,000	12,000,000	No
Northern District BCC	65 Kyats per vehicle per day	70,000	1,000,000 ++	70,000 (document cost)
Eastern District BCC	Different commission fees in different routes	150,000	1,000,000	No
Than Myan Thu BCC	Fixed commission fee per operation cycle	650,000	3,700,000	No
Ma Tha Ta BCC	2% of total revenue of individual operators	2,000,000	48,000,000	No

Table 2. Fare revenue and cost of interviewed individual bus owners

No	Bus control committee	Number of owned vehicles by type	Fare revenue	Fuel cost	Wage	Maintenance cost	Other cost
			Kyats/vehicle/day	Kyats/vehicle/day	Kyats/worker/day	Kyats/vehicle/month	Kyats/month
1	Northern	2 (Dyna)	N/A	N/A	N/A	N/A	N/A
2	Northern	4 (Dyna) +3 (Canter)	N/A	N/A	N/A	N/A	N/A
3	Northern	1 (Dyna)	N/A	N/A	N/A	N/A	N/A
4	Eastern*	1 (Dyna)	45,000	10,000	35% of revenue	N/A	N/A
5	Eastern*	1 (Canter)	35,000	10,000	25% of revenue	N/A	N/A
6	Than Myan Thu*	2 (Dyna)	30,000	10,000	25% of revenue	100,000	No
7	Than Myan Thu*	1 (City bus)	40,000	18,000	35% of revenue	500,000	No
8	Than Myan Thu*	1 (Dyna)	30,000	10,000	25% of revenue	300,000	No
9	Than Myan Thu*	2 (City bus)	40,000	18,000	40,000 ++	N/A	No
10	Than Myan Thu*	4 (Dyna)	30,000	8,000-10,000	30,000 ++	N/A	No
11	Western	4 (Dyna)	40,000-50,000	11,000	27% of revenue	350,000	Helper:150,000
12	Western	1 (Mini bus)	40,000	8,000	20% of revenue	10,000-100,000	No
13	Ma Tha Ta	20 (City bus)	100,000-140,000	27,000	30% of revenue	550, 000 + Mechanics:300,000	Lunch cost: 4,000/worker; Penalty: 25,000/vehicle
14	Ma Tha Ta	3 (Mini bus)	80,000-90,000	17,000	25% of revenue	350,000-450,000	Lunch cost: 4,000/worker
15	Ma Tha Ta	1 (Hilux)	50,000-55,000	9,000	22% of revenue	200,000-250,000	Lunch cost: 4,000/worker

* means that the interview was conducted in accompanying with the officers of BCC.

incentive to report lower revenue than the real revenue. Particularly, we should carefully look at the data collected during the interviews where the officers of BCCs accompanied with the individual bus owners.

The individual bus owners should pay the most of costs that is needed for the bus operation. They include the fuel cost, the wage of bus workers, maintenance cost, commission fee to BCC and other cost. First, the daily fuel cost per vehicle ranges from 30,000 to 55,000 in truck bus while it ranges from 40,000 to 140,000 in city bus. Second, many of individual bus owners apply the achievement-based salary system to pay the wage to drivers/conductors. The achievement rate ranges from 20% to 35% of total fare revenue. Third, the maintenance cost also varies among the operators. The maintenance cost of city bus seems higher than that of truck bus. It is noteworthy to point out that few individual bus owners hire their own mechanical engineers to maintain the bus vehicles. Exceptionally the bus operator No.13 hires the mechanical engineers. This is probably because this operator owns considerable number of city buses. Other operators do not have the mechanical staffs. This means that the quality of vehicle maintenance is quite low in these operators. They should hire the mechanical engineer temporally if the bus vehicle should be repaired. Finally, the individual operators often pay the lunch cost of drivers/conductors additionally. This reflects the local custom of labor market in Myanmar.

The bus owners must pay the registration fee every year to keep the operation license B. The annual registration fee amounts to 350,000 Kyats.

Revenue and Cost of Bus Companies. Table 3 shows the fare revenue and cost which are collected in the interviews with the two private bus companies. The GCL earns some 3,300,000 Kyats per route per day while the UMEHL earns 834,750 Kyats per route per day. This means that the GCL operates the route with larger

Table 3. Fare revenue and cost of bus companies

Companies		Golden City Link	Union of Myanmar Economic Holdings
Fare revenue	Kyats/day	Route1: 3,564,000 Route2: 3,069,000	16,695,000 (in 20 routes)
Fuel cost	Kyats/day	6,700,000	7,595,952
Salary of drivers/conductors	Kyats/worker/day	150,000 + incentives	Non-airconditioned:100,000(Driver); 40,000 (Conductor)
Salary of office workers/mechanics	Kyats/month	16,000,000	18,000,000*
Maintenance cost	Kyats/month	35,000,000	50,000,000
Office cost	Kyats/month	17,500,000	N/A*
Other cost		0	1,000,000 (Sundry expenses)
Number of drivers/conductors		180	1768
Average fuel cost	Kyats/vehicle/day	23,000	Air-conditioned bus: 28,860 Non-airconditioned bus: 23,088

Note: The salary of office workers/mechanics in Union of Myanmar Economic Holdings includes the office cost.

Table 4. Capacity and price of bus vehicle by its type.

Vehicle type	Capacity (passenger)	Price(million kyats)
City bus (air-conditioned)	50	20++
City bus (non-air-conditioned)	45–52	20–33
Mini bus	30	15–22
Dyna/canter	25–35	13–22
Hilux	12–20	12–20
Other type	25–35	12–15

passenger demand than the UMEHL. The daily fuel costs are 6,700,000 Kyats in the GCL and 7,595,952 Kyats in the UMEHL. The daily fuel cost is 23,000 Kyats per vehicle in the GCL while the daily fuel cost of air-conditioned bus is 28,860 Kyats and that of non-air-conditioned bus is 23,088 Kyats per vehicle. The difference of fuel cost between the air-conditioned and non-air-conditioned buses is because the air-conditioned city bus needs 10 compressed natural gas (CNG) tanks per day while the ordinary city bus needs 8 CNG tanks per day. Note that the price of a unit CNG tank is 2,886 Kyats. The constant salary system is applied to the wage of drivers and conductors in both companies. In addition to the constant salary, 6 percent of fare revenue is kicked back to drivers in the GCL. The UMEHL also gives the achievement-based additional salary, but the details are not available. Note that the definitions of the salary of office worker/mechanics, maintenance cost, and the office cost in the GCL are different from those in the UMEHL.

Cost of Bus Vehicles. Table 4 shows the capacity and the price of bus vehicle by its type. This shows that the price of bus vehicles is roughly in proportion to their size. Note that the individual bus operators must pay the import tax to purchase a new bus vehicle. For example, the import tax of city bus is 750,000 Kyats per vehicle. This is nearly equal to 15% of the price of bus vehicle.

Estimation of Cost Functions of Bus Operators

This section analyzes the cost structures of bus operators by estimating the cost functions. In general, there are the three approaches to empirically measuring cost functions: accounting approach, engineering approach and statistical approach (Small and Verhoef, 2007). We will apply the statistical approach to the empirical cost analysis in Yangon, Myanmar. The statistical cost studies pool information from various transportation agencies and/or time periods and use statistical inference to estimate the parameters of cost functions. There have been a number of researches using this approach. For example, Viton (1981) estimates a short-run cost function on a 1975 cross-section of 54 US city bus systems, using vehicle-miles as output. Button and O'Donnel (1985) estimates the cost function using 55 UK bus agencies and passenger revenues as output. Both of them use "translog" functions to estimate the short-run cost functions. Although the translog functions can relax the assumption of linearity in cost functions, it may be difficult for our analysis to apply them because of the lack of data. Thus, we assume the linear cost functions. The bus

operators are categorized into the following three types; bus company; city-bus-owned individual bus operators; and truck-bus-owned individual bus operators. This is because we expect that the differences in operating style among the three operators cause the differences in cost structure among them. The cost structures are compared among the three types of operators.

Formulation of Cost Function. In general, the cost function measures the minimum cost of producing a given level of output for some fixed factor prices (Varian, 1992) in a short run. It summarizes information about the technological choices available to the firms.

Suppose C^k(w, Y^k) be the short-run cost function of firm type k, where w is a vector of given factor prices and Y^k is the output of firm type k. It corresponds to the solution of

$$\text{Min}\sum_i w_i x_i^k \quad \text{subject to } Y^k = f^k\left(\mathbf{x}^k \middle| \mathbf{x}_I^k = \bar{\mathbf{x}}_I^k\right) \text{ and } \mathbf{x}_I^k = \bar{\mathbf{x}}_I^k$$

where $f^k\left(\mathbf{x}^k\right)$ means the production function of firm type k; $\mathbf{x}^k$ is a vector of input factors; and $\mathbf{x}_I^k$ is a vector of the fixed factors. The solution for each input x_i^k is a conditional demand function $x_i^{k*}(\mathbf{w}, Y^k, \bar{\mathbf{x}}_I^k)$, which represents the optimal amount of input. Then the cost function is:

$$C^k\left(\mathbf{w}, Y^k, \bar{\mathbf{x}}_I^k\right) \equiv \sum_i w_i x_i^{k*}\left(\mathbf{w}, Y^k, \bar{\mathbf{x}}_I^k\right)$$

In the context of Yangon's bus industry, the total vehicle kilometers of type k is regarded as the output Y^k. k means the organizational type including the bus company; city-bus-owned individual bus operators; or truck-bus-owned individual bus operators. We assume that the short-run variable input factors of bus operation consist of the fuel and the labor force while the short-run fixed factors consist of the vehicles and others. Then, the short-term cost function appears:

$$C^k\left(w_f, w_l, w_v, Y^k, \bar{x}_v^k\right) \equiv w_f x_f^*\left(w_f, w_l, Y^k\right) + w_l x_l^*\left(w_f, w_l, Y^k\right) + w_v \bar{x}_v^k + w_o \bar{x}_v^k$$

where w_f is the price of fuel, w_l is the wage of labor force, w_v is the unit cost relating to rolling vehicles, w_0 is the price of composite other input factors, x_f^k is the input of fuel, x_l is the input of labor force, $\bar{x}_v^k$ is the input of vehicles and $\bar{x}_o^k$ is the input of composite other factors. Note that we assume that the other factor $\bar{x}_o^k$ is the fixed initial cost, which is the constant for any operator of type k.

It is assumed that the output is in proportion to the inputs including fuel consumption and to the labor. Additionally, we assume that the fuel will substitute perfectly to the labor. Then we assume the Leontief technology for the short-term bus operation as:

$$f^k\left(x_f^k, x_l^k\right) = \min\left\{\alpha_f^k x_f^k, \alpha_l^k x_l^k\right\}$$

where α_f^k and α_l^k are the parameters. This satisfies $x_f^{k*} = 1/\alpha_f^k Y^k$ or $x_l^{k*} = 1/\alpha_l^k Y^k$. Finally, the cost function of bus type k can be rewritten as

$$C^k = \beta_Y^k Y^k + \beta_v^k \bar{x}_v^k + \beta_c^k$$

where $\beta_Y^k, \beta_v^k, \beta_c^k$ are the unknown parameters.

Empirical Estimation of Cost Functions

The cost data collected in the local survey with the bus operators is used for the empirical estimation of the cost functions. As shown earlier, we estimate the cost functions of the three types of bus organizations: the bus company, the city-bus-owned individual bus operator and the truck-bus-owned individual operators .

Table 5 shows the results of the regression analysis for the three cost functions. The dependent variable is the daily operating cost. Note that we use only the two variables in the bus-company's cost function: vehicle-km and constant as there are only two bus companies. For the estimation, we additionally assume that the fixed cost is zero in individual bus operators. This is because the interviews with the individual bus operators unveiled that they regard the initial cost as the sunk cost in the short-run(Although we attempted to introduce the fixed cost into the cost functions of the city-bus individual operators and truck-bus individual operators, the parameters corresponding to the fixed cost are estimated to be negative.). The adjusted R^2 is very high in these models. Figure 1 shows the relationships of the observed data versus the estimated data for city-bus and truck-bus individual operators respectively. These also show the fitness of the model is quite good.

First, the marginal cost with respect to vehicle-km in the city-bus individual operators is much higher than in the truck-bus individual operators. This is mainly because the performance of vehicles is higher in bus companies than in individual operators. The bus companies own the higher-quality bus vehicles which can run at a lower fuel consumption rate than the individual operators. Besides, the bus companies hire the technical staffs for maintenance while the individual operators do not. This difference impacts on not only the fuel efficiency but also the probability of engine troubles during the service.

Second, the city-bus individual operators have higher marginal cost with respect to vehicle-km than the truck-bus individual operators. This is probably because the fuel consumed by the city bus is higher than the fuel consumed by the truck bus. This reflects the difference of vehicle size or weight between the city bus and the truck bus.

Third, the marginal cost with respect to number of vehicle for the city-bus individual operators is five-time higher than that for the truck-bus individual operators. This is probably because the cost of parking facilities for city bus is more expensive than the cost of parking facilities for truck bus.

Table 5. Estimation results of cost functions.

Bus type	Bus company	City-bus individual operators	Truck-bus individual operators
Vehicle km	47.483 (−)	165.61 (1.68)	136.39 (2.59)
Number of vehicles		35 335.07 (1.67)	7 478.87 (0.59)
Constant	35 044 000 ((−)		
Number of observations	2	8	12
Adjusted R^2	1.00	0.79	0.82

Note: () means the t-statistic of corresponding variables

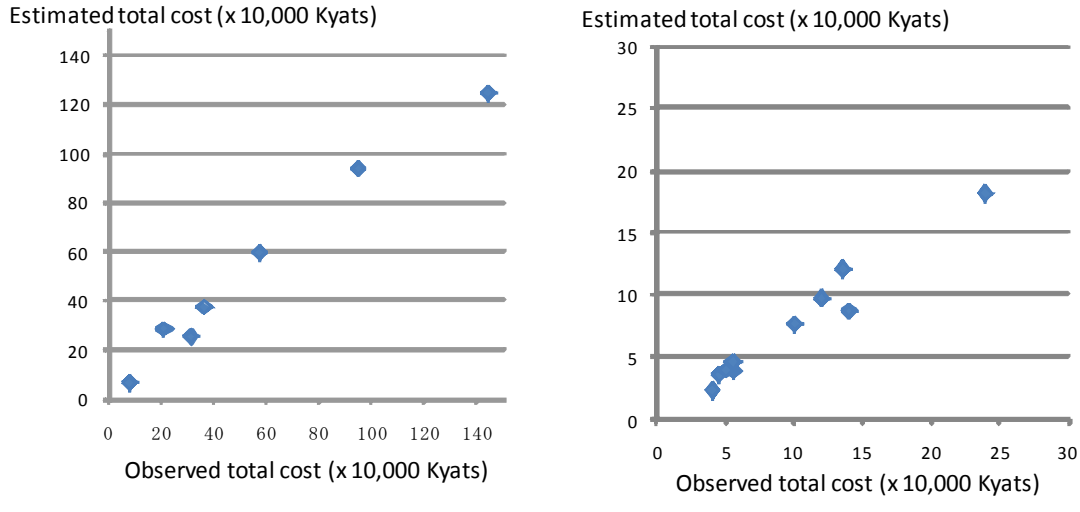


Figure 1. Observed cost vs. estimated cost in city-bus individual operators (left) and truck-bus individual operators (right).

Conclusions

This paper analyzed the cost structure of urban bus operations in Yangon, Myanmar. The data regarding the revenues and costs of local bus operators were collected during the interviews with the bus companies, the bus control committees and individual bus owners. The operating cost includes the fuel cost, labor cost, maintenance cost, office cost and others. Then, the operating cost functions are estimated from the statistical approach. The results show that the marginal cost with respect to vehicle-km in the city-bus individual operators is much higher than in the truck-bus individual operators; that the city-bus individual operators have higher marginal cost with respect to vehicle-km than the truck-bus individual operators; and that the marginal cost with respect to number of vehicle for the city-bus individual operators is five-time higher than that for the truck-bus individual operators. Several research issues are noteworthy. This is the first attempt to collect the cost data of local bus operators in Yangon. However, the collected data do not cover many of the individual bus owners. To estimate the cost functions with more sophisticated

method, more data are required. The quality of cost data should also be improved. If the accounting book or information of bus operators is available, the accuracy of the data becomes higher. It should be noted that the financial data is not well prepared by many individual bus owners.

Acknowledgements

The survey in Myanmar was made on the basis of Urban Public Transport Policy Framework in Japan-ASEAN Transport Framework under the financial support of the Government of Japan. Mr. Tadashi Nishimura and Mr. Ryo Masutomo (Japan Transport Cooperation Association), Mr. U Augn Myint (Road Transport Administration Department), Mr. Maung Maung Sue and Mr. Ko Yan Aye Tun (SM Tours & Transport Co., Ltd) supported our local surveys. We deeply thank them for their kind supports.

References

Button, K. J., and O'Donnel, K. J. (1985). "An examination of the cost structures associated with providing urban bus services in Britain", *Scottish Journal of Political Economy*, 32, 67-81.

Kato, H., Saito, N., Inagi, A., and Myint, U. A. (2010). "Regulatory framework and operational system of urban bus transportation in Yangon", Myanmar, *Transportation Research Board 2010 Annual Meeting*, January 2010. (accepted)

Kuranami, C., Shimomura, T., Ogita, S., Colette, A., Nakagawa, Y. and Kato, H. (2007). "Urban public policy framework (UPTPF): Case studies of Chiang Mai, Batam and Siem Reap", *Proceedings of the Eastern Asia Society for Transportation Studies*, 6, Dalian (China), September ,2007.

Institute Sultan Iskandar and Malaysian Industry Government Group for High Technology. (2005) *Yangon Strategic Development Plan*, Interim Report.

Small, K. A., and Verhoef, E. T. (2007). "The Economics of Urban Transportation", Routledge.

Varian, H. R. (1992). Microeconomic Analysis, 3rd edition, Norton.

United Nations (2007). World Urbanization Prospects, 2007 revision.

Viton, P. A. (1981). "A translog cost function for urban bus transport", *Journal of Industrial Economics*, 29(3), 287-304

Zhang, J., Fujiwara, A. and Thein, S. (2008). "Capturing traveler's stated mode choice preferences under influence of income in Yangon city", Myanmar, *Journal of Transportation Systems Engineering and Information Technology*, 8(4), 49-62.

Transport Infrastructure and Economic Development: An Empirical Study for China

Huang Qian and Harata Noboru[1]

[1] Professor, Department of Urban Engineering, School of Engineering, the University of Tokyo, Japan, E-mail: nhara@ut.t.u-tokyo.ac.jp

Abstract

There has been a fast development of transport infrastructure to support the economic growth in China, which calls for more study on the relation between them. This study employs a production function method and a VAR (Vector AutoRegression) approach to study the relation with national empirical data. The result of production function method is discussable because the elasticity of infrastructure to the output is negligible and may be contradict to common knowledge. The explanation for this includes the data limitation and potential unfixed effects. It is also shown that the more developed areas benefit more from infrastructure than undeveloped areas. The VAR method shows that the infrastructure counts for a significant part of economic growth and a shock in infrastructure might lead to significant short-term effect on economic growth but small long-term effect. The causality of the infrastructure and economic development is vague in statistic meaning but it seems more likely the infrastructure explains the economic growth but not reversely.

Key Words: Infrastructure; economic development; relationship; balance

Introduction

Since 1978 China transited to a market-based economy from planned economy, it has been experiencing a period of fast development. However, due to policy reasons, there has been imbalanced growth among regions, which has gradually become a serious problem. One way to deal with this situation is to help inland areas through appropriate infrastructure development, especially the transport infrastructure. A key to this issue lies on the relationship between the transport infrastructure and economic development. In China, large scale transport infrastructure is under construction, of which many are highways, high-speed railways and transport terminals. A traditional concept is “Want to be rich, build road first”, however, to

what extent the transport infrastructure could promote the economic growth might lack of consideration. In fact, despite the infrastructure itself as a stock, how it stimulates the economy besides the travel cost savings is on discussion worldwide. The present study joins the discussion on the relation between transport infrastructure and economic development through some empirical analysis on China.

Literature Review

There have been numerous studies on the transport infrastructure and economic development related issues in the past decades. This section briefly reviews some related issues.

Effects of Transport Infrastructure on Economic Development. There has been an underlying concept that transport infrastructure has the effect of promoting economic growth, especially among the politicians. However, according to Vickerman (2001) the correlation between transport infrastructure and economic growth is not so stable. It is difficult to assume a single causal direction between these two factors, regarding the high possibility of mutual interaction. According to Banister and Berechman (2001), it is widely agreed that the economic growth happens mainly due to capital, labor, etc. and only partly relying on the infrastructure improvement. Transport acts as a necessary condition for the growth to occur. Although there is no doubt about the direct effect that the transport infrastructure improvement contributes to cost savings of productive sectors, such as time saving, whether there is spillover effect, i.e. additional benefit generated from the infrastructure is discussable.

The Significance of The Effects. The great trend of quantified analysis of this issue could possibly dates back to the work done by Aschauer (1989), who estimated the macro effect of infrastructure investment on American economy. And his work was followed by many researchers, e.g. Munnell (1990), Ford and Poret (1991), etc. In these first trials big estimate results were derived. And this was criticized by other researchers that the high elasticity means unrealistic rate of return of infrastructure (>1). (Gramlich 1994). Yet some explanation was made that a first shock in infrastructure could cause great effect, however, after the basic infrastructure was in place, new investment would not cause much effect. (Hulten 1996). Considering the oversized results of previous work, Holtz-Eakin (1994) argued that results were substantially modified when econometrically taking into account state or region-level unobserved effects. Assumptions were made that the unobserved effects are time invariant. And fixed-effects and first difference regression was employed to reexamine the data. Positive estimates found were significantly smaller than those of Aschauer. (Romp and Haan, 2005) However, there was argument that the first difference will destroy the long term relationship. (Duggal et.al.1999)

China Related Studies. Due to the data limitations prevailing studies on China are mainly on macro level. Early study by Sylvie (2001) examined China data to find that the infrastructure might explain the regional disparities of economic development. In recent years, some related studies have also come forth in China. Xu

et al. (2007) formulated a two-stage correlation between highway transport and economic development. There are also many other researches on this issue (Liu et.al. 2005, Zhang et al. 2007).

Methodology

Two methods, a production function method and a VAR method are used to examine the effects of transport infrastructure on economic development.

Production Function Method. As is wildly used in previous studies, a Cobb-Douglas form production function is adopted:

$$P = AL^{\alpha}K^{\beta}G^{\gamma}$$

where P represents the output; A represents the total factor productivity; L for labor input; K for capital input; G for transportation infrastructure input. α, β, γ, are parameters. For total productivity factor, the Hicksian efficiency term is adopted to represent the technique development over time:

$$A = A_0 e^{\lambda t}$$

where A_0 is a constant; t is the time (year); λ is parameter; thus the following equation can be derived for estimation:

$$\ln P = \ln A_0 + \lambda t + \alpha \ln L + \beta \ln K + \gamma \ln G$$

VAR Approach. Vector autoregression (VAR) is an econometric model used to capture the evolution and the interdependencies between multiple time series. All the variables in a VAR are treated symmetrically by including for each variable an equation explaining its evolution based on its own lags and the lags of all the other variables in the model. It is a data oriented method that could help to analysis the relation of variables free of underlying theoretical assumptions.

Data Sets

Due to the limitation of available data, only macro analysis is made mainly on national level with two different kinds of methods. The data sources used include national statistic year book, transport yearbook, and transport infrastructure data from other literature (Liu et.al. 2007, Xu et.al. 2007)

Empirical Results

Production Function Method. To estimate the coefficients, the national data is used from year 1978-2004. And two different assumptions are considered here:

(1) $\alpha + \beta = 1$, which assumes that there is constant return of scale for labor and capital input while G serves as a total productivity factor.

(2) $\alpha + \beta + \gamma = 1$, which assumes there is constant return of scale for labor, capital stock and transport infrastructure. The G serves as an input factor.

Table 1. Coefficient estimates for $\alpha + \beta = 1$

Coefficients	Value	t-value
Constant	-.959	-2.819
α	.365	4.222
γ	-.067	-1.452
λ	.031	4.786

Table 2. Coefficient estimates for $\alpha + \beta + \gamma = 1$

Coefficients	Value	t-value
Constant	-1.486	-7.497
α	.370	4.842
β	.684	6.690
λ	.025	4.714

Table 3. Comparison between the two regions

	Yangtzs River Delta	Sichuan Province
F-value	33.218	44.281
γ	0.24	0.032

Overall, the results under two assumptions are all statistic significant, while assumption 2 seems to hold better judged by its t-test. However, it is interesting that the coefficients for transport infrastructure in both cases are quite small and negative (γ equals -0.067 and -0.054 respectively), which sounds unrealistic. Possible explanation for this small value might lie on the data. Since the data for transport infrastructure is quite limited, bias could occur during estimation. Another possibility is that there exist some significant unfixed effects which influence the overall economic development, thus a little further investigation into regional data was done. However, due to the limitation of regional data, only two areas: Yangtzs River Delta, which represents the east coastal region, and Si Chuan Province, which represents the west region are picked to make a simple comparison. The results show that the model could explain most of the data; however, the statistic significance is not very desirable. The contribution of transport infrastructure is quite larger in the east coastal area than the west area. This could be interpreted as that the infrastructure improvement in more developed area, where there is larger demand for transportation, could lead to larger benefits. In this sense, there could be a dilemma the infrastructure investment in more developed areas is actually needed because of its large demand but as a result to promote faster economic growth which widens the gap between regions.

VAR Method. The panel data used include national GDP, population in economic activities, capital stock and highway and waterway stock from 1978-2004. Unit root test shows all series are non-stationary and the orders of integration are different. This makes it impossible to make regression or even the cointegration test. So the log values are used. It is shown that log values are of integration order 1 at 10% level, thus cointegration test and VAR might be employed.

Since all the series are non-stationary, there might be spurious regression when directly applying VAR. To overcome this, Vector Error Correction (VEC) is used. An unrestricted VAR process is done in the first place to determine the number of lags. Since the sample size is not big, we only consider 3 lags at most. Based on the principle of minimum AIC (Akaike Information Criterion) and SC (Schwarz Criterion) value, lag of 2 is chosen.

Based on the assumption that some of the series are trend stationary cointegration test is run with linear intercept and trend. And the Trace test indicates that there is 1 cointegrating equation at the 0.05 level. Finally the VEC is estimated with lag 2 and cointegrating rank 1.

With the estimates of VEC, impulse response analysis is done. As is shown in Fig 1, one shock in the transport infrastructure G, the change in GDP converges to a positive value through vibration after about 25—30 time periods. This might imply that an infrastructure shock could cause a relatively significant increase in short-term growth; however, the long-term growth is quite small. On the other hand, a shock in GDP could results in relatively large increase in Infrastructure. Also variance decomposition analysis is done to separate the variation in one variable into the component shocks of other variables. As is shown in Fig 2, the shock of infrastructure does count a significant part of the GDP growth besides itself (15%-20%). And the shock in GDP leads to the most part (45%-50%) of infrastructure increase. What to be noted is that since log values of each variable are used and there is potential mutual causality between variables, the values in analysis here do not mean the elasticity of one variable to another. However, it still provides us with the information about the trend and the significance of relation between infrastructure and economic growth.

Table 4. VEC estimates

	D(GDP)	D(K)	D(L)	D(G)
CointEq1	0.092229	0.062915	0.005761	-0.15311
D(GDP(-1))	0.286663	-0.28943	0.1642	-0.4037
D(GDP(-2))	-0.26186	0.026653	0.527691	1.012451
D(K(-1))	1.31222	1.864384	-2.12884	0.35021
D(K(-2))	0.507916	0.006953	1.11883	-2.12332
D(L(-1))	1.309582	0.750301	-0.13722	-0.67718
D(L(-2))	0.631504	0.36706	-0.01157	-0.14014
D(G(-1))	0.185105	0.112161	-0.06928	0.282398
D(G(-2))	0.0457	0.075525	0.023651	-0.08668
c	-0.17344	-0.11753	0.073908	0.260751
R-squared	0.755729	0.876959	0.740344	0.665288
F-statistic	4.812592	11.08704	4.43527	3.091889

D(): denotes the difference of ().

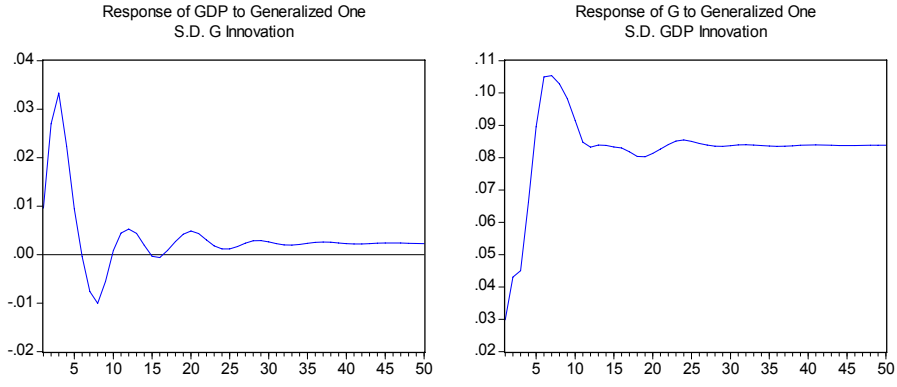


Figure 1. Impulse response results

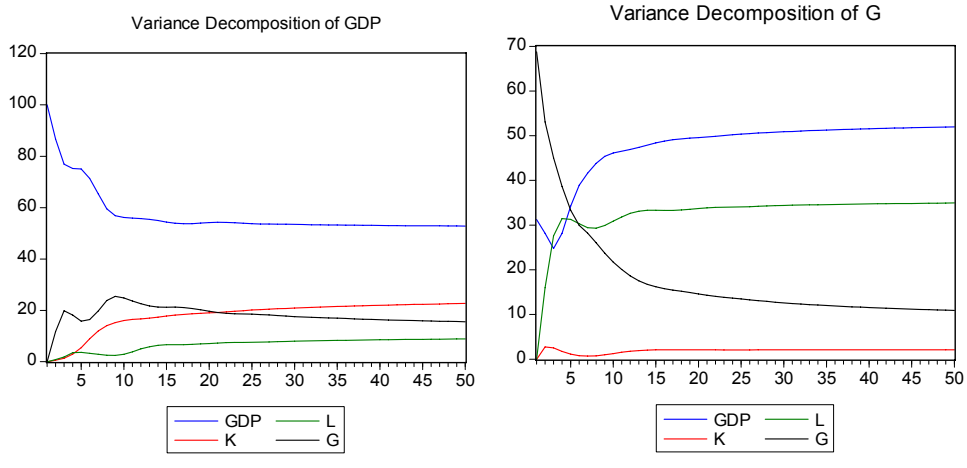


Figure 2. Variance decomposition results.

Table 5. Granger causality test.

Null Hypothesis:	Lag	F-Statistic	Probability
G does not Granger Cause GDP	8	3.14977	0.26326
GDP does not Granger Cause G	8	0.32175	0.89971

Finally Granger test is used to examine the causality between infrastructure and economic growth. In general, it is better to use more rather than fewer lags, since the theory is couched in terms of the relevance of all past information. As to transport infrastructures the life time is quite long, and it is shown in the impulse analysis that the influence of infrastructure could be significant in more than 10 time periods. So a lag of 8 is chosen, which is the max lag the data size permits.

We cannot reject the hypothesis: G does not Granger Cause GDP and also GDP does not Granger Cause G at 5% level. So the causality is still not clear in this case. However, it is more likely that G is explainable for GDP but GDP does not explain G.

Conclusion

The present study tries to provide some empirical analysis on the relation between transport infrastructure and economic development using China's data. The production function method and VAR method are adopted due to their different attributes. The results of the production function method is discussable since it gives a quite small and negative elasticity of infrastructure at national level, which is unrealistic based on our common knowledge. This may be due to the limitations of small samples and not sufficient data to cover all transport infrastructures. It also implies there might be some significant unfixed effects during the estimation, which needs further study. The comparison between two specific regions shows positive effect of infrastructure on the economic development, and also the more developed areas turn out to benefit more from the improvement of infrastructure than under developed areas. The results derived from VAR method also indicates that the infrastructure counts for a significant part of economic development. And it also shows that a shock in infrastructure could lead to a significant short-term effect on economic growth but quite small in the long run. The causality between infrastructure and economic development remains vague in statistics, which is also partly due to the small sample size. It seems more likely that the infrastructure explains economic development but not the other way round. Further studies are needed with more detailed and specified data.

References

Aschauer D.A. (1989). "Is public expenditure productive?" *Journal of Monetary Economics*, 23, 177-200.

Banister, D., Berechman, Y. (2001). "Transport investment and the promotion of economic growth." *Journal of Transport Geography*, Vol. 9, 209-218.

Duggal V., C. Saltzman and L. Klein (1999). "Infrastructure and productivity: a Nonlinear Approach." *Journal of Econometrics*, 92, 47-74.

Ford R. and P. Poret (1991). "Infrastructure and private-sector productivity." *OECD Economic Studies*, 17, 63-89.

Gramlich E.M. (1994). "Infrastructure investment: a review essay." *Journal of Economic Literature*, 32, 1176-1196.

Holtz-Eakin D. (1994), "Public-sector capital and the productivity puzzle." *Review of Economics and Statistics*, 76, 12-21.

Hulten C.R. (1996). "Infrastructure capital and economic growth: how well you use it may be more important than how much you have." *NBER working paper* 5847.

LIU Bing-lian, LIU Yong (2007). "Estimation in China cumulative amount of

highway and waterway capital." *Journal of Beijing Jiaotong University (Social Sciences Edition)*, 6(7), 44-48

LIU Bing-lian, ZHAO Jin-tao. (2005). "Empirical study on causality between transportation and regional economic development in China." *China Soft Science*, 2005(6)

Munnell A.H. (1990). "How does public infrastructure affect regional economic performance?" *New England Economic Review*, Sept-Oct., 2-22.

Romp W. and J. de Haan (2005). "Public capital and economic growth: a critical survey." *European Investment Bank Papers* ,10(1).

Sylvie D´emurger. (2001). "Infrastructure development and economic growth: an explanation for regional disparities in China?" *Journal of Comparative Economics* 29, 95–117

Vicerman (2001). report in the 119th ECMT round table, 2001.

Xu Haicheng, Li Jian, Yang Yan (2007). "Relationship between highway transportation and economic development in China" *Journal of Changan University (Social Science Edition)*, 9(22), 8-13

Xu Xianxiang, et.al. (2007). "Estimates of fixed capital stock by sector and region: 1978 —2002." *Statistical Research*, 24(5),6-13

ZHANG Wang, ZHOU Jian-an. (2007). "An analysis on the impact of the investment of infrastructures to the economic growth in China." *Statistics & Information Forum*, 22(3)

Estimating Greenhouse Gases from Roadway Transportation —Methodology Overview

Delwar Murshed[1]

[1] Ph.D., P.E., Senior Transportation Planner, Urban Planning Office, Washington State Department of Transportation, 401 Second Avenue South, Suite 300, Seattle, WA 98104; PH: (206) 464-1263; Email: DelwarM@wsdot.wa.gov

Abstract

Since Greenhouse gases (GHG) are believed to lead to global warming, emission of such gases are now considered an environmental threat that must be confronted. About 30% of the GHG emissions in the United States are from transportation sources. This is why, transportation planners, policy makers, and the public often wants to know what effects a transportation project or a policy decision will have on GHG emissions. How to estimate GHG emissions in planning stage of a project or at policy decision levels? The standard process involves running a travel demand model as well as a sophisticated vehicle emission model. Data requirements of most of these emission modeling tools strain the ability of travel demand models to produce inputs at sufficiently fine detail. Moreover, due to emission model's complexity, time-consuming data collection, and staff and project budget constrains, it may not be feasible for transportation professionals to apply such sophisticated tools for estimating vehicle emissions for all transportation projects. Rather a quick and easy approach would help planners/engineers to apply travel demand model for vehicle emission estimation. The purpose of this paper is to provide an overview of methodologies that can help estimate GHG emissions without requiring too much of difficult to find data. This paper will first provide an overview of the common approaches about applying travel demand modeling data for roadway vehicle emission estimation. Several quick and easy vehicle emission estimation approaches will then be described.

Key words: transportation; greenhouse gas; system-wide emissions; travel model data

Introduction

Once, all climate changes occurred naturally. However, during the Industrial Revolution, human activity began altering earth's climate and environment through changing agricultural and industrial practices. Before the Industrial Revolution, human activity released very few gases into the atmosphere, but now population growth, fossil fuel burning, and deforestation are affecting the mixture of gases in the atmosphere.

Roadway vehicles are one of the primary sources of harmful gas emissions. The emissions produced by vehicles fall into three basic categories—tailpipe emissions, evaporative emissions, and life-cycle emissions. The major pollutants from tailpipe emissions include hydrocarbons, nitrogen oxides (NO_x), carbon monoxide (CO), and carbon dioxide (CO_2). The evaporative emissions include gas tank venting, running losses, and refuel losses. The life cycle emissions are produced in the activities associated with the manufacturing, maintenance, and disposal of automobiles.

Many chemical compounds, including some gases from vehicle emissions, found in the Earth's atmosphere act as GHGs. These gases allow sunlight to enter the atmosphere freely. When sunlight strikes the Earth's surface, some of it is radiated back towards the space as infrared radiation or heat. GHGs absorb this infrared radiation and trap the heat in the atmosphere. Many gases exhibit these greenhouse properties. Some of them occur naturally in the atmosphere, while others result from human activities. Naturally occurring GHGs include water vapor, CO_2, methane (CH_4), nitrous oxide (N_2O), and ozone. Certain human activities, however, add to the levels of most of these naturally occurring gases. Human activities produce a number of GHGs including CO_2, CH_4, N_2O, hydrofluorocarbons (HFCs), perfluorocarbons (PFCs) and sulfur hexafluoride (SF_6).

CO_2 is released to the atmosphere when solid waste, fossil fuels (oil, natural gas, and coal), wood, and wood products are burned. CO_2 is a product of the complete combustion of hydrocarbons and has no immediate harmful effects to humans and is essential to plant life. However, the emissions of CO_2 are considered a pollutant because it is a significant GHG that contributes to global climate change. The sources of CH_4 are human activities that include livestock and paddy rice farming, land use and wetland changes, pipeline losses, decomposition of organic wastes in municipal solid waste landfills, and newer style fully vented septic systems. CH_4 is also emitted during the production and transportation of coal, natural gas, and oil. N_2O is emitted during agricultural (i.e., use of fertilizers) and industrial activities, as well as during combustion of solid waste and fossil fuels. Fluorinated gases such as HFCs, PFCs and SF_6 are very powerful GHGs, and are generated in a variety of industrial processes. Among all the GHGs, CO_2 is known to be the primary contributor to global warming.

Globally, transportation related emissions of CO_2 are growing rapidly. The use of petroleum as a fossil fuel for transportation dominates CO_2 emissions from this

source. In 1999, in the US, more than 30 percent of fossil fuel related CO_2 emissions were a direct result of transportation (GHG, 2008). With about two-thirds of this being from petrol consumption by motor vehicles and the remainder coming from diesel and jet fuel use in trucks and aircraft, respectively.

Since the global climate change has become one of the serious concerns in the world, the reduction of GHG emissions due to human activities is increasingly becoming important. As vehicle emission is one of the major sources of GHGs, the transportation industry as a whole can play a major role in reduction of GHGs.

Global Warming Potentials of GHGs

Each GHG differs in its ability to absorb heat in the atmosphere. HFCs and PFCs are the most heat-absorbent. Methane traps over 21 times more heat per molecule than CO_2 (EPA, 2006). However, according to the guidelines of the United Nations Framework Convention on Climate Change (UNFCCC), nitrous oxide absorbs 320 times more heat per molecule than CO_2 (UNFCCC, 1991). This ability of gases to absorb heat is measured by Global Warming Potentials (GWP).

GWP is defined as the cumulative radiative forcing - both direct and indirect effects - integrated over a period of time from the emission of a unit mass of gas relative to CO_2 as a reference gas (IPCC, 1996). Direct effects occur when the gas itself is a GHG. Indirect effects occur when chemical transformations of a gas produce other GHGs, or when a gas influences other radiatively important processes such as the atmospheric lifetimes of other gases.

GHGs with relatively long atmospheric lifetimes (e.g., CO_2, CH_4, N_2O, HFCs, PFCs, and SF_6) tend to be evenly distributed throughout the atmosphere, and consequently global average concentrations can be determined (EPA, 2003a). Table 1 shows GWP values of such GHGs in the atmosphere. The short-lived gases such as water vapor, CO, ozone, other ambient air pollutants (e.g., NOx, and VOCs), and tropospheric aerosols (e.g., sulfur dioxide products and black carbon), however, vary spatially, and consequently it is difficult to quantify their global radiative forcing impacts. GWP values are generally not attributed to these gases that are short-lived and spatially inhomogeneous in the atmosphere.

Table 1. GWP and atmospheric lifetimes of greenhouse gases (IPCC, 1996).

Name of gas	Atmospheric lifetime (years)	100-Year GWP*
Carbon dioxide (CO_2)	50-200	1
Methane (CH_4)	9-15	21
Nitrous oxide (N_2O)	120	320
Hydrofluorocarbons (HFCs)	264	11,700

*GWPs typically have an uncertainty of roughly ±35 percent.

In addition to communicating emissions in units of mass, one may choose also to use GWPs to reflect their inventories and projections in CO_2-equivalent terms, using information provided by the Intergovernmental Panel on Climate Change (IPCC, 1996). GWP values allow policy makers to compare the impacts of emissions and reductions of different gases. According to the IPCC, any use of GWPs should be based on the effects of the GHGs over a 100-year time horizon.

Why to Estimate GHGs?

GHGs are now considered an environmental threat that must be confronted. About 30% of the GHG emissions in the United States are from transportation sources (GHG, 2008), and such gases are generated in proportion to fuel consumed. Steps to mitigate emission of GHGs are similar to steps taken to reduce fuel consumption. The often-asked question is, to what extent does additional roadway capacity exacerbate the problem and what can be done to mitigate it?

Recently, GHG emissions and global warming are getting more and more public attention. For most transportation improvements or policy decisions, the public often wants to know what effect the improvement or the policy decision will have on GHG emissions. Transportation planners and decision makers want to know to what extent are solutions within the realm of transportation planning and to what extent are they in the realm of other policy areas? How to address this issue early in the planning process? How emissions are affected by congestion induced by population growth and maintaining economic vitality? Planners and engineers need to estimate GHG emissions for developing strategies to address these issues resulting from vehicle emissions.

How to Estimate GHGs?

How to estimate GHG emissions in planning stage of a project or at policy decision levels? Typically travel demand model is first run to generate traffic related input data. The process to apply travel demand model data to roadway vehicle emission estimation is generally straight forward, but how to get the emission factors/rates is the most crucial task. The emission factors generally come from sources such as a vehicle emission model, an emission curve/table, formula, or regression model (Liu et al., 2005; Williamson and Chidanamarri, 2000). Based on how emission factors are derived, GHG emissions from vehicles can be estimated using one of three methodologies:

- Emissions from vehicle emission models,
- Emissions from speed-emission rate relationships, and
- Emissions from fuel consumptions.

Emissions from Vehicle Emission Models

This approach is a formal/standard process that runs a travel demand model as well as a sophisticated vehicle emission model, such as MOBILE, MOVES, EMFAC, etc.

The MOBILE is a computer program that can help estimate emission factors of hydrocarbon (HC), NOx, CO, air toxics, gaseous sulfur dioxide (SO_2), ammonia (NH3), six hazardous air pollutant (HAP), and direct particulate matters (PM) from vehicle exhaust and brake and tire wear. Until recently the MOBILE series were USEPA's officially approved model for HC, NOx, and CO estimation (EPA, 2003b; EPA, 2004). Recently USEPA's Office of Transportation and Air Quality (OTAQ) has developed a new tool for modeling emissions. This tool is known as the MOtor Vehicle Emission Simulator (MOVES). This new emission modeling tool will help estimate emissions for on-road and non-road mobile sources, cover a broad range of pollutants, and allow multiple scale analysis. The latest version of this tool is MOVES2010 (EPA, 2009). The EMFAC (EMission FACtors) model helps calculate emission rates from all motor vehicles, such as passenger cars to heavy-duty trucks, operating on roadways in California. The latest version of this model is known as EMFAC2007 (ARB, 2007). This model calculates emission factors and emission inventories for HC, CO, NOx, CO_2, PM, fuel consumption, oxides of sulfur (SOx), and lead (Pb).

The inputs of these vehicle emission models are very complicated. For instance, MOVES model needs data such as vehicle age distribution for each type of vehicles, ambient temperature and relative humidity for each hour, off-network vehicle parameters (e.g., average vehicle population, start fraction, extended idle fraction, parked vehicle fraction), fuel characteristics, etc. (Audette et al., 1993). Although it is relatively more accurate for the vehicle emission estimation, it may not be easy for transportation professionals to apply such models for all types and scales of transportation projects because of emission model's complexity, time-consuming data collection, lack of skill and budget constrains.

Emissions from Speed-Emission Rate Relationships

This is a quick and easy approach. In this approach, vehicle emission is calculated using travel model outputs and the average emission rates which are directly derived from the speed-emission rate relationships. The USDOT (USDOT, 2005) and TRB (TRB, 1995) have developed a number of such vehicle speed-emission curves (USDOT, 2005). An article published in ACCESS in California shows a similar vehicle speed-emission curve (Figure 1).

The United States Federal Highway Administration (FHWA) developed a tool known as the Surface Transportation Efficiency Analysis Model (STEAM) to estimate vehicle emissions using speed-emission table. STEAM used information generated through the travel demand modeling process to compute the net value of mobility and safety benefits attributable to regionally important transportation projects. The tool helps develop quantitative estimates of natural resource usage (e.g., energy consumption) and environmental impacts (e.g., pollutant emissions - CO, NOx, PM10, VOC, and CO_2) (DeCorla-Souza and Hunt, 1998).

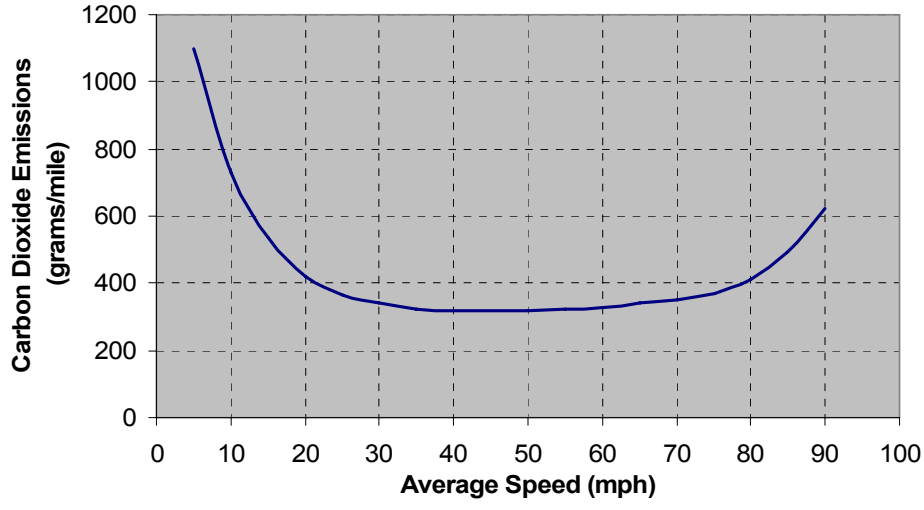


Figure 1. Vehicle speed and carbon dioxide emission rate (after Barth and Boriboonsomsin, 2009).

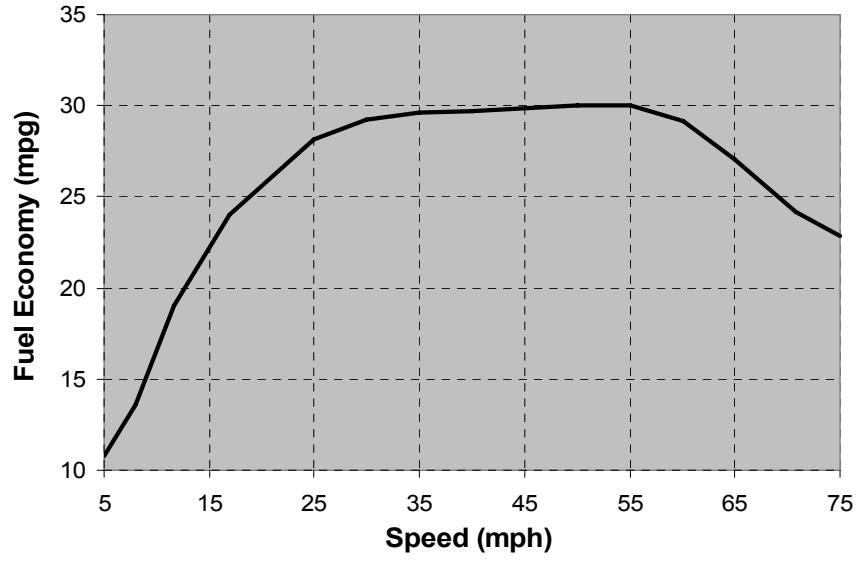


Figure 2. Vehicle Speed and Fuel Economy (after USDOE, 2008; West et al., 1999).

TRB vehicle speed-emission curve and STEAM speed-emission table are two of the most used sources of emission factors/rates for quick vehicle emission estimations. However, the TRB vehicle speed-emission curve was published in 1995 and the STEAM speed-emission table was published in 1990's. Are these emission rates still valid?

Based on the FHWA MOBILE6.2 sensitivity study, the emission rates are very sensitive to changes in travel behaviors such as vehicle miles traveled by hour of the day, vehicle speeds, and vehicle types and age distribution (USDOT, 2005). All these sensitive factors that affect vehicle emission rates have changed significantly in the last few decades. This is why, the EMFAC2000 has been updated to EMFAC7G,

and the updates include changes in basic emission rates, driving cycle adjustments, speed adjustment factors, vehicle fleet age distribution, and specific environmental parameters (Niemeier and Gao, 2000). The 1995 TRB speed-emission rate curve and STEAM speed-emission table developed in 1990's need to be updated to the current.

Emissions from Fuel Consumptions

This is another quick and easy vehicle emission estimation approach. In this methodology, total fuel consumption is first calculated using travel demand model outputs (vehicle type, link average speed, VMT), and fuel intensity (e.g., gallons/mile). The fuel consumption can also be estimated from a speed-fuel consumption curve (Figure 2). However, in order to estimate the fuel consumption more accurately, the vehicle speed-fuel consumption curves need to be developed for each vehicle type and fuel type.

Once total fuel consumption is known, emissions of GHGs (i.e., CO_2, CH_4, N_2O) can be calculated by multiplying the fuel consumptions and CO_2 equivalences. For example, Victoria Transport Policy Institute (VTPI) suggests CO_2 equivalencies as shown in Table 2. The STEAM User Manual documents fuel consumption and emission rates that can also be used to calculate total emissions (USDOT, 1999).

Table 2. Carbon Dioxide Equivalences (VTPI, 2006).

Fuel type	Greenhouse gases (grams/liter)*			Total CO_2 equivalent
	CO_2	CH_4	NO_2	
Gasoline	2,360	0.2273	0.3358	2,469
Diesel	2,730	0.0605	0.20	2,793
Ethanol 10	2,124	0.2273	0.3358	2,233
Ethanol 85	531	0.2273	0.358	640
Conventional aircraft fuel	2,330	2.19	0.23	2,447
Jet fuel	2,550	0.08	0.25	2,629

*Carbon dioxide equivalency of CO_2=1, CH_4=21, and NO_2= 310.

Limitations of Modeling Tools

In applying travel demand modeling data for vehicle emissions estimation, the actual vehicle operations, such as acceleration, deceleration, stop & go, cold/hot start (Audette et al., 1993; Gonzalez, 1995; Khan, 1999, and Elolahde, 2006), speed distribution, ambient temperature (Khan, 1999), vehicle mix (Lindmark, 1992), monthly/daily/hourly factors (Noriega and Florian, 2005; Elolahde, 2006), etc. are important to be considered in the process. Since most planning models predict average speed on a link, the variations of traffic condition cannot be captured. Some other concerns to apply travel model for vehicle emission estimation include missing VMT (intrazonal and drive to park & ride lot), level of detail (hourly, vehicle mix, hot/cold starts), and VMT impacts through pricing. Often travel forecasting models

cannot provide accurate travel speed data, and thus significantly affecting accuracy of estimated emissions.

To overcome this problem, the link speed data may be refined using geometric design, traffic characteristics, traffic delay, and roadside characteristics (Nesamani et al., 2006). The emission rate equations can also be developed accounting for vehicle operational characteristics in the intersections—idling, acceleration, deceleration, and cruise stages (Williamson and Chidanamarri, 2000). The cold startup emissions can be estimated using the number of trip origins multiplied by a cold start emission rate (NCRTPB, 2006).

Although MOBILE was USEPA's long-used emission estimation tool, it had no capability to estimate CO_2 which is the most important GHG due to human activities (EPA, 2004). The MOVES has the capability to estimate CO_2 emissions. The EMFAC model calculates emission factors and emission inventories for CO_2, but is developed and used only in California (ARB, 2007). The outputs of travel-demand models are inputs to the MOVES Model for estimating emissions. The MOVES model is a substantial improvement over the prior MOBILE series, but its demands strain the ability of travel demand models to produce inputs at sufficiently fine detail.

Factors Affecting Emission Rates

There are a number of factors that affect the rate at which any vehicle emits GHGs. Some of the most important factors are roadway type (e.g., freeways, arterials), vehicle type/size (motorcycles, cars, light trucks, heavy trucks, urban and school buses), engine size, vehicle age and accumulated mileage, fuel type (gasoline, diesel, CNG, LPG, ethanol E85, Methanol M85, electricity), ambient weather conditions (temperature, wind, precipitation), the maintenance condition of the vehicle (well maintained, in need of maintenance), vehicle emission control technology (LEV, tier 0, tier 1, catalyst, non-catalyst, uncontrolled), and how the vehicle is driven (long cruising at high speed, stop-and-go urban congested driving, typical urban mixed driving).

Newer vehicles generally use less fuel and produce less emission, while older vehicles generally use more fuel and produce more emissions (EPA, 2005). Vehicle age is a significant factor in the determination of motor vehicle emissions for several reasons, including:

- The increasing stringency of emission standards over time;
- The year of manufacture determines the level of emissions control technology built into the vehicles;
- For a given vehicle and year of manufacture, emission levels tend to increase with increasing vehicle age and accumulated mileage (APC, 2002). This reflects the effect of normal wear and tear on the efficiency of built-in emission control technologies (e.g., catalytic converters).

Another significant factor affecting GHG emissions is fuel type such as conventional fuel or alternative fuels. There are more than a dozen alternative and advanced fuels in production or use today (EPAct, 1992; AFDC, 2008). Although government-regulated and voluntary private fleets are the primary users of these fuels, consumers are showing a growing interest in them. These alternative fuels include:

- Methanol;
- Ethanol;
- Biobutanol;
- Compressed natural gas;
- Liquefied natural gas;
- Liquefied petroleum gas or propane;
- Ultra-low sulfur diesel;
- Biodiesel;
- Hydrogenation-derived renewable diesel (HDRD);
- Fischer-Tropsch diesel (FTD);
- E-diesel;
- Coal-derived liquid fuels;
- Biogas fuel;
- Biomass to liquids (BTL);
- Gas to liquids (GTL);
- P-series;
- Dimethyl ether (DME);
- Hydrogen;
- Electricity;
- Solar energy.

Review of literature (Murshed, 2009) reveals that biodiesel has the potential to reduce GHG emissions up to 75% compared to petroleum diesel. Methanol can reduce GHG emissions by 4% to 60% depending on the type of feedstock used to produce methanol, and its use in regular gasoline vehicles or fuel cell vehicles. Using ethanol instead of gasoline reduces GHG emissions ranging from 15% to 86% depending on the source of energy and the type of feedstock used for ethanol production. Compressed natural gas can reduce GHG emissions by 16% to 40% depending on its use in conventional or advanced technology vehicles. Liquefied natural gas can potentially reduce GHG emissions up to 40% depending on the source of the natural gas. Conventional alternative fuel vehicles using propane can provide up to 25% GHG reductions. Compared to gasoline, use of hydrogen in fuel cell vehicles can reduce GHG emissions by about 25% to 85% depending on type of feedstock used for hydrogen production. On a life-cycle basis, plug-in hybrid and electric vehicles provide over 50% and 75% reduction in GHG emissions, respectively.

Concluding Remarks

In performing GHG emissions estimation, ideally roadway, traffic operational, vehicle and fuel characteristics need to be considered. Although some of the more

recent models or tools are relatively more accurate for the vehicle emission estimation, it may be difficult for transportation professionals to apply such models for all types and scales of transportation projects because of emission model's complexity, time-consuming data collection, lack of skill and budget constrains. Using speed-emissions relationships or fuel consumption data to help estimate GHGs could be quick and easy approaches to evaluate conceptual alternative policies or transportation improvements.

References

AFDC. (2008). *Alternative Fuels and Advanced Vehicles Data Center*, U.S. Department of Energy, www.eere.energy.gov/afdc/fuels.

APC. (2002). *South East Queensland Motor Vehicle Emission Inventory*, Adam Pekol Consulting for the Environmental Protection Agency, Australia.

ARB. (2007). "Mobile Sources Emissions Inventory Program", California, EMFAC2007, Version 2.30, User's Guide, Air Resources Board (ARB).

Audette, A., Florian, M. and Wu, J. H. (1993). "Computing emissions for the montreal area with EMME/2." 8*th International EMME/2 Users' Conference*, Seattle, Washington.

Barth, M. and Boriboonsomsin, K. (2009). *Traffic Congestion and Greenhouse Gases*, ACCESS Publication.

DeCorla-Souza, P. and Hunt, J. (1998). *Use of STEAM in Evaluating Transportation Alternatives*, Federal Highway Administration, USA.

Elolahde, T, (2006). "Traffic model system and emission calculations of the helsinki metropolitan area council." 20*th International EMME Users' Conference*, Canada.

EPA. (2003a). *Inventory of U.S. Greenhouse Gas Emissions and Sinks*: 1990–2001, Final Version, EPA 430-R-03-004, U.S. Environmental Protection Agency, Washington, D.C.

EPA. (2003b). *MOBILE*6 *Vehicle Emission Modeling Software*, Mobile Source Emission Factor Model, http://www.epa.gov/otaq/m6.htm#m60, U.S. Environmental Protection Agency.

EPA. (2004). *Policy Guidance on the Use of MOBILE*6.2, Memorandum, U.S. Environmental Protection Agency.

EPA. (2005). *Emission Facts: Average Annual Emissions and Fuel Consumption for Gasoline-Fueled Passenger Cars and Light Trucks*, Office of Transportation and Air Quality, United States Environmental Protection Agency.

EPA. (2006). Environmental Protection Agency, USA, http://yosemite.epa.gov/oar/globalwarming.nsf/content/emissions.html.

EPA. (2009). *MOVES - Motor Vehicle Emission Simulator*, http://www.epa.gov/otaq/models/moves/index.htm, Office of Transportation and Air Quality (OTAQ), United States Environmental Protection Agency.

EPAct. (1992). *The Energy Policy Act*, http://www.eere.energy.gov/vehiclesandfuels/epact/, US Department of Energy.

GHG. (2008). *Carbon Dioxide—Sources, Sinks and Science*, www.ghgonline.org, Greenhouse Gas Online.

Gonzalez, R. C. (1995). "Traffic related pollution in urban planning—the case of madrid." 4*th European EMME/2 Users' Conference*, Brussels, Belgium.

IPCC. (1996). *Climate Change* 1995: *The Science of Climate Change*, Cambridge University Press. Cambridge, U.K.

Khan, A. M. (1999). "Issues in the use of emme/2 as a platform for the estimation of greenhouse gases." *First Asian EMME/2 Users' Conference*, Shanghai, China.

Lindmark, K. (1992). "Calculating vehicle emissions from EMME/2 assignment." 7*th International EMME/2 Users' Conference*, Vancouver, British Columbia, Canada.

Liu, J. J., Fan, B. Q. and Xia, X. M. (2005). "Study on the control measure of urban transport emission using EMME/2." 3rd Asian EMME/2 Users' Conference, Shanghai, China.

Murshed, D. (2009). "Effects of alternative fuels on greenhouse gas emission from transportation sources." *Proceedings of the 15th International Sustainable Development Research Conference*, July 5-8, Utrecht, The Netherlands.

NCRTPB. (2006). *Air Quality Conformity Determination of the 2006 Constrained Long Range Plan and the FY*2007-2012 *Transportation Improvement Program for the Washington Metropolitan Region*, National Capital Region Transportation Planning Board (NCRTPB), Metropolitan Washington Council of Governments.

Nesamani, K. S., Chu, L., McNally, M. and R. Jayakrishnan. (2006). "Estimation of vehicular emissions by capturing traffic variations." *TRB* 2006 *Annual Meeting* CD-ROM, Transportation Research Board, USA.

Niemeier, D. and Gao, O. (2000). *Mobile Emissions Inventories: EMFAC7G and EMFAC2000 Comparison by Air Basin*, UC Davis-Caltrans Air Quality Project, No. 00108.

Noriega, Y. and Florian, M. (2005). "Computing vehicle emissions and display on a fine grid." 19th International EMME/2 Users' Conference, Seattle, Washington.

TRB. (1995). *Metropolitan Highways: Implications for Air Quality and Energy Use*, TRB Special Report # 345, Transportation Research Board.

UNFCCC. (1991). *United Nations Framework Convention on Climate Change*.

USDOE. (2008). *Driving More Efficiently*, U.S. Department of Energy (USDOE), http://www.fueleconomy.gov/feg/driveHabits.shtml.

USDOT. (1999). *Surface Transportation Efficiency Analysis Model (STEAM) User Manual*, US Department of Transportation, Federal Highway Administration.

USDOT. (2005). *Sensitivity Analysis of MOBILE6 Motor Vehicle Emission Factor Model*, US Department of Transportation, Federal Highway Administration, http://www.tdot.state.tn.us/mediaroom/docs/2005/emission_reductions.pdf.

VTPI. (2006). *Transportation Cost and Benefit Analysis*, Victoria Transport Policy Institute.

West, B. H., McGill, R. N., Hodgson, J. W., Sluder, S. S. and Smith, D. E. (1999). *Development and Verification of Light-Duty Modal Emissions and Fuel Consumption Values for Traffic Models*, Oak Ridge National Laboratory, Oak Ridge, Tennessee.

Williamson, D. and Chidanamarri, S. *Air Quality Aspects of Traffic Management*, University Transportation Center for Alabama, Report No. 00108, 2000.

Adaptation Strategies for Low Carbon Transport by 2050

Hector G. LOPEZ-RUIZ[1]

[1] Dr., Researcher, Transport Economics Laboratory- University of Lyon - 14, avenue Berthelot –F-69363 Lyon Cedex 07, FRANCE; Email: hector.lopez-ruiz@let.ish-lyon.cnrs.fr / hlopezruiz@gmail.com

Abstract

Today, numerous works conclude that transport seems to be completely coupled to economic growth. Therefore, as a direct consequence of economic development, transport sits today as one of the major final energy consumers and one of the most important sources of carbon dioxide emissions. Furthermore, in the absence of major technological change, this unsustainable situation will most undoubtedly get worse in the future. In this paper we analyze what different types of public policies aiming at sharp reductions in GHE emissions imply on passenger transport and how they can be linked to new behavior patterns affecting consumption and in fine freight transport. For this, we use the transport issues in the long term (TILT) model's core microeconomic choice model. In this manner, we explain the interest of a time management approach to adaptation strategies for GHE reductions.
Key words: greenhouse emissions, long term, scenario, transport, sustainable development

Introduction

Today, numerous works conclude that transport seems to be completely coupled to economic growth. Therefore, as a direct consequence of economic development, transport sits today as one of the major final energy consumers and one of the most important sources of carbon dioxide emissions. Furthermore, in the absence of major technological change, this unsustainable situation will most undoubtedly get worse in the future. It is in part, from this preoccupation that spawned the Kyoto Protocol, ratified by the European Union Member Countries, which was a vital first step towards global emission reductions. This has pushed the EU Environment Council to adopt the conclusion that, in order to achieve stabilization in the environmental effects of oil consumption in transport activities, developed countries should reduce

emissions by about 15–30 % by 2020 and 60–80 % by 2050, below the base year levels (1990). On the basis of the EU recommendations, French public authorities have adopted a reduction objective of 75% from the 1990 level.

This reduction objective is not an easy task, especially since transport activities in France are at the origin of 25% of the county's total energy consumption and are responsible for, at least, 30% of its carbon dioxide emissions. In essence, 80% of these emissions come from road transport. Even tough we have seen the birth of new technologies in the transport sector and that we have witnessed a growing "social and entrepreneurial conscience", unfortunately, these environmentally friendly products and/or services have observed a very slow implementation.

What is more, in the effort to counteract global warming and oil dependency, new automotive technology prospects have been increasingly presented as possible solutions to the need for drastic reductions in greenhouse emissions. However, recent studies like those of: A. SCHEAFER, et al, (2009); D. SPERLING & N. LUTSEY (2009) and LET-ENERDATA (2008), have clearly underlined that technological progress cannot be effective if it is not accompanied by deep organizational and behavioral changes (especially if we are aiming at a very important reduction in emissions).

In this sate of matters, transport activities as we know them clearly pose an environmental problem that will most certainly lead to increasing constraints on mobility through public action. Consequently, there is a growing need in finding straightforward answers to questions such as: How will policies oriented at reducing fossil fuel use affect passenger mobility? But more importantly, how will these changes influence passenger behavior as well as general consumption patterns and how can adaptation strategies best integrate this?

In order to answer these questions we will analyze two existing long term mobility scenarios for the French economy built in 2008 by the LET-ENERDATA study group (Y. CROZET, H.G. LOPEZ-RUIZ, B. CHATEAU, and V. BAGARD). These scenarios were built using a backcasting methodology and they depict different options available for a 75% reduction in GHE emissions for the French economy. Furthermore, we will carry out our analysis using the microeconomic module of the TILT (Transport Issues in the Long Term) model which was used to build these scenarios.

The TILT model has been designed to be a long-term equilibrium model centered on defined behavior types -in which the speed/GDP elasticity plays a key role- in order to determine macroeconomic transport demand estimations. In LOPEZ-RUIZ (2009) this model was enriched to also be able to asses the system's sensitivity to public policies, investment needs in infrastructure and economic impact of different public policies whilst taking into account microeconomic choices. In this manner, the TILT model combines a macroeconomic and microeconomic structure in a backcasting approach that facilitates long-term planning through five modules: A macroeconomic

module, a microeconomic module, a vehicle fleet dynamic & technology evolution module, a public policy module and an impact assessment module.

This paper particularly focuses on the microeconomic side of the TILT model in order to precisely analyze what different types of public policies might imply on passenger transport. Firstly, we will take a look into the past tendencies for passenger transport and how they let us understand its evolution patterns for mobility in France. This analysis will be followed by an explanation of how the TILT model assesses passenger transport's evolution on a microeconomic basis and how changes in passenger mobility impact behavior in consumption patterns. Lastly we will point out how, accordingly to changes in behavior, public action should develop adaptation strategies for GHE reductions through a time management approach.

Main Tendencies in Passenger Transport and Model Assessment

Although an important number of transport policy mixes are imaginable in different types of situations, their results come down to: optimizing transport use; encouraging modal shifts and/or adjusting transport distances. These variations are induced on the system through adjustments in key transport variables as well as on the system's capacity to promote and integrate new technologies (motor and non-motor).

In this sense, the implementation of different policy mixes entails a (re-)optimization of agents' (passengers and firms) choices based on transport cost and the opportunities offered by transport services. In this manner, social effects of policies will most likely be linked to welfare variations associated to trade-offs between transport expenditure; house/firm location; accessibility; transport money budgets and in fine consumption of goods/services.

This (re-)optimization will not only translate into the system as a differentiation in modal and spatial evolution in accordance with technology use. It will also bring about social effects on a public level: contribution to emission reductions; new tax revenues/expenditures; infrastructure investment needs; overall impact on the economy (production, employment, productivity, etc.) as well as equity issues between different territories and population groups.

In this manner, in order to assess policy effects on future transport activities, a good understanding of the present situation is required. What is more, it is very important to identify how past tendencies have contributed to the current state of the actual transport structure.

Main Tendencies in Passenger Transport. Currently, transport activities in France are largely road based with a large part being linked to personal vehicle mobility. Nearly 85% of all trips in France are made using private vehicles and average distance per car have been fairly stable up until the last few years. What is more, for the last 10 years, car mobility has actually decoupled from mobility growth.

Table 1. Time passed on different transport modes in France.

In min/pax per day	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	AAGR 1993–2003
Road	26.4	26.7	27.1	27.2	27.6	28	28.4	28	28.8	28.8	29.4	1.10%
Rail	3.3	3.2	2.9	3.1	3.1	3.2	3.3	3.4	3.4	3.4	3.3	0%
Air	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.20%
Soft modes	12.8	12.5	12.3	12.1	11.9	11.7	11.5	11.3	11.1	10.9	10.7	–1.70%
Total	42.6	42.5	42.4	42.5	42.7	43	43.3	42.8	43.4	43.2	43.5	0.20%

Source: Les comptes des transports

Table 2. Effect brake-down for evolution in transport times.

In %/year	Traffic effect	Speed effect	AAGR 1993–2003
Road	1.4	–0.3	1.1
Rail	0.6	–0.6	0
Air	0.2	0	0.2
Soft modes	–1.7	0	–1.7
Total	0.5	–0.2	0.2

Source: Les comptes des transports

The main explanation for the slowdown in average distances and passenger kilometers in this period is due to a change in level of service. Indeed, although transport times (Table 1) have been fairly stable during this period and speeds on road transport services have actually increased, congestion problems are largely responsible for the growth in time spent on road modes.

As we can see in Table 2, the congestion effect on road transport has been negative and although average road speeds have gone up, its effect is limited and has not been able to keep transport times from going up a little over 11%. Although times have increased roughly 3 min per passenger per day, this translates into loosing 27 million man-hours per year.

The upside to these observed higher transport times is an increasing modal shift, over the last few years, favoring public transport services. This shift has been translated into a small GHE reduction but, even though it is encouraging, it is far from what is needed in order to reach the desired reductions in fossil fuel consumption and GHE emissions (the objective in France is -75% by 2050).

In light of this, and in order to further the GHE reductions, public authorities in France have resorted to the use of economic incentives to ensure the widespread use of new cleaner automotive technologies and increase the match in current technology supply and consumer demand. As a result, a fiscal program has been set up to try to influence the private vehicle market in order to renew vehicle fleets and to curb emissions from this sector.

Although this program has had a lot of success, studies like those of F. CUENOT show that if average weight on new cars in Europe had stayed constant from 1995 to

2005, the 2008 target of 1998s voluntary agreement would have been met and new regulation aiming at better averages in GHE emissions would probably have been unnecessary. Furthermore, F. CUENOT shows that if, in the following years, vehicle weights do not increase, new GHE reduction objectives might be attained by 2015 through a combination of strategies referring to car weight and motor technology.

In the French case, a "GHE bonus-malus" program started in 2008 based on malus payments or bonus rebates for new cars. These bonus/malus are calculated according to fiscal power units based on the car's maximum power and its CO_2 emissions. This program has been quite successful and although, two years might be a short time span to get to any firm conclusion, we find it interesting to observe (Table 3) that growth in total car sales in France has been fairly stable but skewed. As we can see in Table 3, when we look at sales according to fiscal power units, we can notice that the bonus/malus program combined to increasing fuel prices has resulted in a decrease in sales for mid-sized cars as sales for smaller and bigger cars have both gone up.

Table 3. Evolution for car fleet, fuel prices, consumption, distances and other modes.

	AGR 1998-2008	AAGR 1998-2008	% of fleet in 1998	% of fleet in 2008
Total car fleet	16%	1.5%	-	-
New 4 FPU cars	65%	5.1%	15%	23%
New 5 FPU cars	–1%	–0.1%	28%	26%
New 6 FPU cars	–9%	–1.0%	28%	24%
New 7 FPU cars	–64%	–9.7%	19%	7%
New 8 FPU cars	290%	14.6%	3%	10%
New 9 FPU cars	42%	3.6%	3%	4%
New 10 to12FPU cars	10%	0.9%	4%	4%
New 13 to16FPU cars	66%	5.2%	1%	2%
New 17 FPU or more cars	187%	11.1%	0%	1%
Total new cars	6%	0.5%	-	-
Average fuel prices	55%	4.5%	-	-
Average fuel consumption L/100 km	–10%	–1.1%	-	-
Average distance per car per vear (in km)	–9%	–0.9%	-	-
Total fuel consumption for cars (millions of m^3)	–5%	–0.5%	-	-
GPKM in private cars	6%	0.6%	-	-
GPKM in buses	14%	1.4%	-	-
GPKM rail	32%	2.8%	-	-

Data sources: Les comptes des Transports Fichier central des automobiles(FCA) and Ministère de I'Écologie, de Energie,du Développement durable et de la Mer.
Data available on 03/2010 showed results only until 2008

GPKM= Billons of passenger-kilometers
FPU=Fiscal Power Units (which are calculated on the basis of the cars max. power and CO_2 emissions).

H.G..LOPEZ-RUIZ 2010.

Table 4. Passenger-kilometer growth.

	AGR 1998–2008	AAGR 1998–2008
Private cars	6%	0.6%
Buses	14%	1.4%
Rail	32%	2.8%
All	9%	0.9%

Data source: Les comptes des Transports

H.G..LOPEZ-RUIZ 2010

Moreover, the slowdown on mid-sized vehicles has not been translated in a reduction in total car sales or total vehicles and thus, has not been a key element in fuel consumption reductions. Rather, total fuel consumption has gone down because of increased sales for smaller cars and, additionally, a drop in average fuel consumption by larger cars.

This translates into the fact that, even though fuel prices have gone up and consumption averages have gone down, total fuel consumption has stayed relatively high because bigger cars with relatively "affordable" -in relation with the car's final price tag- malus payments have gone from an 11% market share to 21%.

In this setting, reductions in average travel distances by cars and modal shifts towards public transport –table 4- can be largely explained by growing fuel prices, which are not entirely under the control of a determined public policy but rather fluctuations on international markets.

This means that we resume the situation in three effects:

- fuel prices and the "GHE bonus-malus" scheme have had little or no impact on total car demand;
- fuel prices and the "GHE bonus-malus" have had ad unconvincing impact on reductions stemming from vehicle size effects;
- fuel prices have had an important impact on vehicular use and a shift towards public transport.

Furthermore, this implies –as we can see in figure 1- that the cost of personal vehicle use has increased and that economic policies aiming at influencing vehicle size -and not vehicle demand/use- have resulted in:

- a decrease in car use -with lower consumption averages;
- a slow shift towards public transport.

These changes in the transport structure have certainly implied a reallocation of household funds. As more and more passengers shift from private transport to public transport services and/or walking, a slow reallocation of transport money budgets has begun to take place.

Presently, this change in behavior -and its inherent reallocation of money budgets- has not been sufficiently important to make a difference in general consumption

patterns. Nevertheless, as the effects of public policies aiming at multimodality and eventually decoupling increase, reallocation effects will most certainly have a big impact on the economy and will most certainly change consumption patterns.

Indeed, as we can see in figure 1, in time, the budget reallocation effects will have an important impact on consumption: ranging from investing in new location strategies to reduce transport costs or using extra money stemming from the savings made by shifting from private personal transport to public transport in new tea cups, more leisure or even financial services. This reallocation of household budgets will certainly have an impact on transport activities (be it freight or passenger), especially on public transport activities.

What is more, modal split changes stemming from public policy imply other effects. As a matter of fact, since different transport means have different averages speeds and levels of service, money budget reallocations will be accompanied by a reallocation of time budgets.

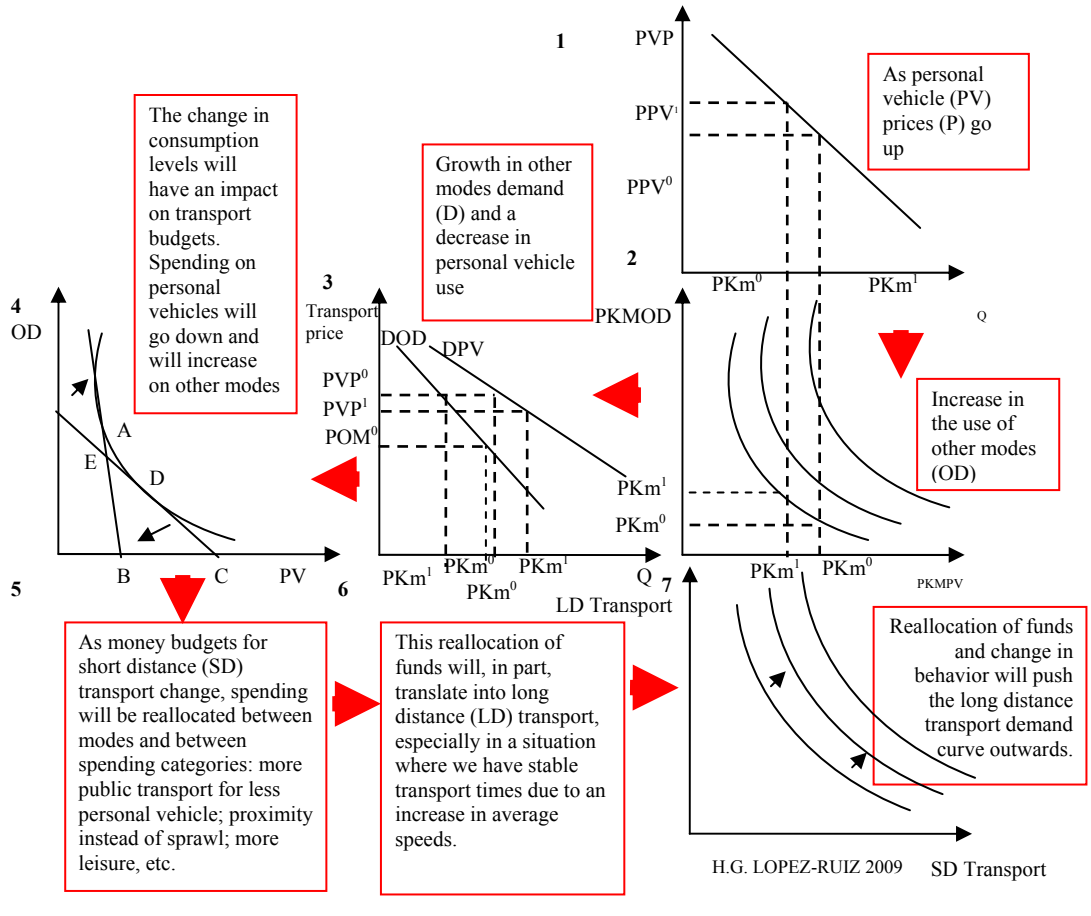


Figure 1. TILT logic for passenger transport evolution

In this manner, as long as the inherent principle of the actual transport policy will be based on the fact that rendering private transport more expensive than public transport will cause a "controlled" modal shift, the results stemming from the policy's macroeconomic changes will most certainly influence the microeconomic choices in other domains.

In order to assess how change in behavior will influence adaptation strategies in the long term, it is necessary to take into account how changes in passenger behavior will undoubtedly have an effect on transport services and how this can be used to plan integrated freight/passenger policies leading towards GHE reduction adaptation strategies. The TILT model's microeconomic module might offer some insight into how to take into account defined behavior types in order to determine long-term policies as per their impact on behavior.

TILT model developments. The TILT microeconomic module lets us understand past tendencies -the coupling between growth and mobility- and allows us to asses demand for transport services as well as trade-offs linked to policies and infrastructure use on different geographical scales through its main equation:

$$a_{ij}(t) = \frac{\left[Lateness\ index(t)\right]^{\alpha} * \left[\frac{\sum\ goods\ and\ services\ available}{\text{T * mode cost}}\right]^{\beta}}{\sum_{1 \in Ni}\left\{\left[Lateness\ index(t)\right]^{\alpha} * \left[\frac{\sum\ goods\ and\ services\ available}{\text{T * mode cost}}\right]^{\beta}\right\}} \quad \forall j \in N_i \quad (1)$$

where T equals time and is the sum of accessibility time (AT) plus transport time (TT)

The TILT microeconomic module is largely inspired by developments done on ant algorithms (DORIGO et al, 1999) and their application to freight and passenger transport (LOPEZ-RUIZ, 2009). This module relies on the idea of a representative agent that optimizes its transport choices by taking into account opportunities available (defined as the sum of Goods and Services Available -from hereon in called GSA- that can be consumed in a period of time, S. LINDER 1970) and cost in respect to a certain level of service on infrastructure (measured through a lateness index, LOMAX, 1997).

The inherent logic of the microeconomic module is particularly useful in technico-organizational public policy assessments as it enables an analysis based on the idea that public policies are implemented as increasing/decreasing constraints on the system in view of getting to a certain objective. Consequently, this facilitates the building of scenarios where a wide variety of social effects on different levels and aspects are comprised. In this manner, the TILT model is capable of giving insight on how changes in the transport structure linked to environmentally oriented public policies might influence passenger behavior in the future.

Policy and its Implications on Behavior and Overall Adaptation Strategies

In 2008 the LET-ENERDATA study group built three exploratory scenarios using a backcasting methodology in order to give insight on future transport options in France. This project required the specification of a desired future (-75% in emissions by 2050) and then different scenarios were specified by identifying the diverse equilibriums possible that allow the attainment of the specified future. From these possible equilibriums, the three that best depict the range of solutions available were chosen and presented. These three scenarios are:

- Pegasus - promoting strict technology standards
- Chronos - promoting green multimodality
- Hestia - promoting the decoupling of transport activities and economic growth (GDP).

Without going into the specifics of each scenario, their results were obtained by modeling a mix of different policies aiming at sensible changes in transport behavior and new motor technologies. Each of these scenarios implies different characteristics and thus different types of results concerning modal shares, fuel consumption, GHE and infrastructural needs.

The LET-ENERDATA scenarios are presented as a series of increasing constraints that are coherent with the attainment of the 75% reduction objective. In this manner, the first scenario, Pegasus is based on an inelastic market structure largely dependent on private vehicles and it aims at showing that a 75% reduction in emissions can not be solely obtained through the introduction of new hybrid motor technologies. Consequently, this result is used to illustrate the need for other options concerning emission reductions, more precisely the need to look for changes in transport behavior through policy. Accordingly, the Chronos and Hestia scenarios are then presented as different possibilities to attain the desired GHE reductions through multimodality and/or decoupling from GDP. The two policy scenarios imply changes on behavior and consumption patterns that can be very illustrative on the evolution of money and time budgets and, if analyzed, can offer insight on adaptation strategies based on time management.

Multimodality Policy. The multimodality scenario, introduces public policies where constraints on speed and emissions come into play as a signal aiming at changing behavior patterns. In consequence this scenario implies an increase in road transport costs that are coherent with the carbon values that were being proposed, in 2009, by public authorities in France (A 32 tax per ton of GHE is supposed to bring in over 8 billion Euros in tax revenues by 2010. This tax is supposed to go up to 350 by 2050).

In this scenario, the rise in road transport cost translates into a sharp increase in the use of rail and public transport for personal and freight mobility. This, in turn, implies that the average overall speed in the system goes down and transport times go up. Consequently, if we suppose the market structure and its inherent elasticities

do not change in time, household and firm transport money budgets would invariably go up in a very sharp way (more or less depending on car use elasticity).

Furthermore, since this scenario is based on a large increase in transport times and transport cost, this implies that in order to stay within certain logic, the system's adaptation should be accompanied by substantial investments in rail infrastructures, which will account for an increasing part of total trips.

Therefore, in order for this scenario to be viable, we should suppose that the population would:

- accept a very sharp increase in public transport and rail subsidies;
- accept that transport money budgets will go over the 15% mark;
- accept a transport time budgets would go up.

This situation seems particularly difficult because it translates into: paying more (for car users) for lower transport speeds (an all modes) and thus loosing potential value added time (VAT) that could be spent increasing revenue. In this setting, as we can deduct from Eq.(2) as transport times and costs go up, the choice model is driven to a situation where moving from point *i* to point *j* is less and less attractive because of the time spent on doing it or the cost that it represents.

$$a_{ij}(t) = \frac{\left[Lateness\ index(t)\right]^{\alpha} * \left[\frac{\sum GSA}{AT + \Uparrow \mathrm{TT} * \Uparrow \text{mode cost}}\right]^{\beta}}{\sum_{l \in Ni} \left\{\left[Lateness\ index(t)\right]^{\alpha} * \left[\frac{\sum GSA}{\mathrm{AT} + \mathrm{TT} * \text{mode cost}}\right]^{\beta}\right\}} \qquad \forall j \in N_i \qquad (2)$$

In this manner, unless the representative agent has the choice of being able to use a transport mode that is fast and cheap (at the same time), mobility would actually become a burden. However, in the absence of a solution linked to new transport services, increasing the sum of goods and services (opportunities) linked to transport activities might help to counterbalance the situation of lost VAT by pushing forward the idea of a value added transport time (VATT). Where the notion of VATT refers to transforming the time we spend on transport into value added time (not necessarily monetary value).

In short, by rendering transport time more productive a strong multimodality scenario aimed at drastic emissions reductions should imply a change in the way that goods and services are proposed and thus consumed. What is more, these changes might help to channel reallocation of funds in household budgets, which would in turn reinforce multimodality policies by rendering public transport more value intensive and thus productive.

It would then seem that an increase in transport times and/or cost would imply a high demand for productivity solutions linked to transport activities, much like what we can see today in transport systems with intensive use of public means. Consequently,

it would be possible to offer goods and services that transform increasing non-value added transport times into value added transport times.

Decoupling Policy. If we follow the same line of reasoning in a decoupling scenario, the logic of adaptation changes drastically. The decoupling scenario is very influenced by proximity services, the public policies at play are largely related to spatial planning and infrastructure investment.

In the decoupling logic, the main trade-off is directly linked to localization strategies and production organization aimed at decoupling transport distances form GDP growth. This entails a densification of main cities and production sites which would in turn; translate into a sharp increase in the use of urban and regional road networks. In this manner road investments needed are less important than in other scenarios but they are totally concentrated on urban and regional infrastructure.

Unlike the multimodality scenario, transport cost characteristics in a decoupling logic lead to more stable transport money budgets (around 15% of total budgets) because transport distances grow at a slower pace. Nonetheless, in order to render this type of scenario viable a very big increase in road transport costs followed by a very sharp decrease in rail and public transport costs (that should be largely subsidized) is needed.

In this case, if there is a sharp increase in transport times and/or cost, the normal response would be a very fast increase in demand for proximity solutions as consumer behavior is modified. This would imply new types of consumptions logics and services.

Indeed, in a multimodality scenario, a change in cost implies paying more for the same opportunities with higher transport times, thus the need to counterbalance through new opportunities. In a decoupling scenario, due to the fact that constraints are even higher than in a multimodality scenario, opportunities must increase even more in order to counterbalance overall distance reductions.

Historically, market power has enabled transport operators and planners to develop the offer on a transport networks based on an “access to the network” logic and asses their services levels using indicators similar to the UK’s PTAL index (Public Transport Accessibility Level –TFL, 2006) where accessibility to geographical locations in a given time is calculated without taking into account opportunities. In a decoupling scenario, this logic would be radically modified.

In the case of a decoupling scenario, very high use of proximity public transport would imply a decrease in price elasticity of demand (PED) in public transportation and thus cause a shift in market power. This change in market power will most undoubtedly profit users instead of transport operators. Consequently, the system would be pushed towards a change in how opportunities are conceived by users and planners.

Indeed, as more and more people use public transport services, market power would shift and we could assist to the emergence of a logic based on a completely different basis where level of service would be measured based on "time it takes to access opportunities", much like the Access To Opportunities and Services (ATOS) index (COOPER, S. WRIGHT, P. et al, 2009) proposed by Transport for London in order to improve planning measures.

As with any change in a system, modifying how planning is conceived implies that a new set of choices would be proposed to users and, as a result, the choice model would be lead to change into something reassembling Eq. (3), where consumption is differentiated. Consequently, a certain number of goods and services could be subject to changes in the way they are proposed and thus consumed.

$$a_{ij}(t) = \frac{\left[Lateness\ \ index(t)\right]^{\alpha} * \left[\frac{\sum \Uparrow GSA}{\Updownarrow AT + \mathrm{TT} * \text{mode cost}}\right]^{\beta} * \left[\frac{\Updownarrow AT + \mathrm{TT}}{\sum \Uparrow GSA * \text{mode cost}}\right]^{\sigma}}{\sum_{1 \in Ni} \left\{\left[Lateness\ \ index(t)\right]^{\alpha} * \left[\frac{\sum GSA}{AT + \mathrm{TT} * \text{mode cost}}\right]^{\beta} * \left[\frac{AT + \mathrm{TT}}{\sum GSA * \text{mode cost}}\right]^{\sigma}\right\}} \quad \forall j \in N_i \quad (3)$$

In Eq. (3), we identify two types of consumptions. Firstly we identify products and services that require a given transport time to be accessed and consumed and for which the choice model will still be based on a trade-off involving time, cost and opportunity (even though it would be ideal to have a pool within walking distance for everyone in Paris, it would be highly improbable that spatial availability would permit this). Secondly, we identify products and services answering to new proximity strategies aimed at counterbalancing higher costs and lower travel distances. Indeed, as market power shifts and planning decisions will be increasingly taken on the basis of ATOS, a certain number of goods and services will face a new choice model in consumer decision, where, in short, either they are accessible in a reasonable time-frame and they can be consumed or they won't be consumed (which would lead to a lack of opportunities and thus an imbalance in the system).

This effect is reinforced by the fact that in a decoupling configuration, even though long-distance travel is largely based on high speed rail, a large part of mobility is reserved to urban and regional services. This can only be possible if adaptation strategies are completely aimed at offering increasing proximity opportunities.

In this manner, the main difference between the presented GHE reduction scenarios lies in time management adaptation strategies. In a multimodality scenario, the adaptation strategy should be based on increasing Value Added Transport Time (VATT) whilst in a decoupling, scenario, adaptation will be linked to increasing Value Added Accessibility Time (VAAT). Where the notion of VAAT refers to using the time we do not spend accessing goods and services for their consumption to create value (not necessarily monetary value) in other activities.

Conclusion

It is widely known that whilst public policy might be aimed at reducing GHE emissions and its effects on the environment, the impacts of these actions will be different for passenger and freight transport. Through this paper we have aimed at showing how these differences will operate according to different public policies and how this will impact consumption behavior in the long term. In addition, we would like to identify three main points on the relationship between policy and their impact on adaptation strategies based on time management:

Adaptation Leads to Systemic Changes. On a regional and interregional level "low-emission high speed transport" for passengers is something that already exists and that will most certainly keep on being promoted as a solution to GHE. On an urban level, the fact that most big cities in France are well equipped in public transport will most certainly bring about big changes on how public transport is conceived and how it might be better utilized.

Change Needs Flexibility. Moreover, freight transport solutions will be a key part of progress made in GHE reductions on passenger transport. Indeed, big changes in modal split on passenger transport will undoubtedly have to be accompanied by proximity services that imply proximity freight transport solutions. In this manner, associating freight policy strategy and passenger policy strategy will be important in order to gain flexibility whilst continuing to open new markets for new types of behaviors and consumption patterns. These integrated policies will undoubtedly require careful planning based on adaptation strategies based on VATT and VAAT.

Europe Wasn't Built in One Day. Nevertheless, the actual lack of interoperability and/or flexibility and/or reliability of non-road modes will most certainly play a very big part in the continued success of road transport in Europe for years to come. Even though all of these problems can be overcome, it will most certainly take more than a decade before they start to be seriously addressed and in order to start dreaming of a serious integrated reduction strategy. However, the enormous potential for information technology solutions in VATT and VAAT adaptation strategies should not to be underestimated as it offers great opportunities for low-cost adaptation with great return value possibilities.

Acknowledgments

The author would like to thank Bertrand Château (ENERDATA) and Prof. Yves Crozet (LET-University of Lyon) for their comments concerning the refinement of certain aspects of the micro-module. This work was supported by the ADEME and the French Ministry of Transport.

References

Banister, D., Hickman, R., and STEAD, D. (2006) *Looking Over the Horizon: Visioning and Backasting* (VIBAT).

Banister, D., Stead, D., Steen, P., Åkerman, J., Dreborg, K., Nijkamp, P., and Schleicher-tappeser, R. (2000). *European Transport Policy and Sustainable Mobility*, London.

Commission des comptes des transports de la Nation. (2007). Ministère de l'Écologie, de l'Energie, du Développement durable et de l'Aménagement du territoire, Paris.

Conseil Général des Ponts et Chaussées. (2006). Démarche Prospective Transports 2050. Ministère des Transports de l'Equipement du Tourisme et de la Mer. Paris
Consortium VLEEM. "VLEEM 2." FINAL REPORT, EC/DG Research. 2002.

Contrôle général économique et financier & Conseil General des Ponts et Chaussés (2006), Rapport sur la comparaison au niveau européen des coûts de construction, d'entretien et d'exploitation des routes, Paris.

Cooper, S., Wright, P., and Ball, R. (2009). "Measuring the accessibility of opportunities and services in dense urban environments: experiences from London." *European Transport Conference*.

Crozet, Y., and Lopez-ruiz h. G. (2010). Passenger mobility and climate constraints in France. In consideration for *WCTRS 2010*.

Cuenot, F. "GHE emissions from new cars and vehicle weight in Europe; How the EU regulation could have been avoided and how to reach it?" *Energy Policy*, In Press.

Dorigo, M., Di Caro, G., and Gambardella, L. M. (1999). "Ant Algorithms for Discrete Optimization." *Artificial Life* 5(3): 137-172.

Let-enerdata (Y. Crozet, H.G. Lopez-ruiz, B. Chateau, V. Bagard). (2008). Comment satisfaire les objectifs internationaux de la France en termes d'émissions de gaz à effet de serre et de pollution transfrontières ? Programme de recherche consacré à la construction de scénarios de mobilité durable. Rapport final. PREDIT, Paris.

Linder, S. (1970). *The Harried Class of Leisure, New-York and London Columbia*, University Press.

Lomax. T. Turner. S. et al. (1997). NCHRP Report 398. *Quantifying Congestion Final Report*.

Lopez-ruiz H. G. (2009). "Quantifying the Economic Impact and Infrastructure Needs for the French Scenarios Aiming for a 75% Reduction in GHE Emissions by

2050: How vehicle market Trends May Impact Policy Strategies and Investments." Presented at the *European Transport Conference* in 2009.

Lopez-ruiz H. G. (2009). *Environnement & Mobilité 2050: des scenarios pour le facteur 4 (-75% de GHE en 2050)*. PhD Thesis of University of Lyon, October 2009.

Lopez-ruiz, H. G. (2010). "Sustainable Transport in France: Is a 75% Reduction In GHE Emissions Attainable?" To be published in the *TRR Journal of the TRB* in 2010.

Schafer, A., et al. (2009). *Transportation in a Climate-Constrained World*, MIT Press, Cambridge Massachusetts.

Schafer, A., Victor, D.G., (2000). "The future mobility of the world population." *Transportation Research Part A*, 34, 171-205 09.

Sperling, D., and Lutsey, N. (2009). *Energy efficiency in Passenger Transportation*, The Bridge, Summer 20. National Academy of Engineering.

TfL (Transport for London). (2006). *Transport Assessment Best Practice*. Guidance document.

Emission Volume, Air Quality, and Health Effects in Japan: The Cases of Tokyo, Kawasaki, Osaka, Amagasaki, and Kita Kyusyu

Keiko Hirota[1] and Shigeru Kashima[2]

[1] Doctoral candidate, Department of Civil Engineering, Graduate School of Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan;
Email: khirota@jari.or.jp
[2] Prof., Department of Civil Engineering, Graduate School of Chuo University, 1-13-27 Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

Abstract

This paper aims to verify the causal relationships among emission volume (mobile sources and stationary sources), air concentration, and health effects (1974-2006) in certified areas of Tokyo, Kawasaki, Osaka, Amagasaki, and Kita-Kyusyu, which were certified according to Japanese Pollution-related Health Damage Compensation law. Our first hypothesis was that an increase in the total emission volumes of suspended particulate matter (SPM), NOx, SO_2, and CO in these cities corresponded to a deterioration in the air quality. The second hypothesis was that the number of certified patients increases with the deterioration in air quality. To support the results of this correlation, the number of new patients, recovery, and mortality were assessed by a simple general equilibrium model. Finally, the comparability of the estimated results with the track record was estimated on the basis of the presence of high t-values.
Key words: emission volume; air quality; health effect; pollution-related health damage compensation system in Japan

Introduction

Due to the rapid economic growth in early 1960s, the air quality in the industrial sector of Yokkaichi, Mie prefecture, rapidly deteriorated. Residents in Yokkaichi had higher rates of morbidity and higher prevalence rates of respiratory and circulatory diseases. The health effects caused by air pollution became a social issue (Amagasaki, 1974; Kawasaki, 1994; Kita Kyushu, 1974; Osaka, 1974; Tokyo, 1978). The affected patients sought compensation from the Japanese government and the industrial sector.

Among the potential polluters, a major issue was the apportioning of the compensation. There was strong opposition from the industrial sector at the beginning of discussions among the polluters (Matsuura, I, 1984).

In the late 1960s to early 1970s, environmental laws and standards were established to address these issues (ERCA, 2008). In 1967, the Pollution Prevention Basic Law was established. This law established permissible limits of each contaminant in air (SO2 in 1969, CO in 1970, suspended particulate matter (SPM) in 1972, and NO_2 in 1973). The policies for reducing air pollution initially focused on stationary sources, such as industries.

To address the social issues associated with the health effects observed since the early 1960s, the Pollution-related Health Damage Compensation Law was enacted in 1973 to protect the affected patients by providing prompt and fair compensation to cover damages and by implementing the necessary programs for the welfare of these patients. The Pollution-related Health Damage Compensation and Prevention System was established in June 1974 in accordance with the Pollution-related Health Damage Compensation Law. The system focused on the following points:

(1) the law provides for civil court cases;
(2) the "polluter pays" principle;
(3) financial support for certified patients.

In 1974, the certified areas and diseases for the compensation system were determined by the government. The system was implemented at the areas shown in Figure 1 (Class 1). This law is responsible for determination of the groups that should bear the expenses in the Class 1 designated areas, where health damage has occurred; classification of the compensation; certification of patients with the 4 pollution-related respiratory and circulatory diseases (pulmonary emphysema, asthmatic bronchitis, bronchial asthma and chronic bronchitis), and implementation of welfare services to facilitate the recovery of patients with health damage (20). If patients with these 4 diseases in the certified areas satisfy certain other conditions (such as extended residence in the certified area, non-smoker status, illness established by a medical doctor), they are considered as certified patients. The purpose of the established system was to collect emission charges from 9,000 factories and offices on the basis of their SO_2 emission volumes. SO_2 emission charges from polluting industries (80%) and a vehicle tax (20%) covered the compensation costs.

After the 1970s, the focus of emission reduction policies shifted from the industrial sector to non-industrial sources, namely, household and mobile sources. In 1977, the Committee of Pollution Prevention focused on the city level and on emissions from households. In 1980, emission reduction from mobile sources was re-enforced. In September 1987, amendments were made to the compensation system. In March 1988, the Pollution-related Health Damage Compensation Law was revised due to the reduction in air pollution. This revision was intended to assist local public organizations with health consultations, health examinations, and functional training;

to allow research and studies on related issues; and to disseminate knowledge regarding health maintenance in local communities. The revision also ended the acceptance of new patients for certified diseases. Thus, the focus of the system shifted from compensation to activities for prevention of air pollution.

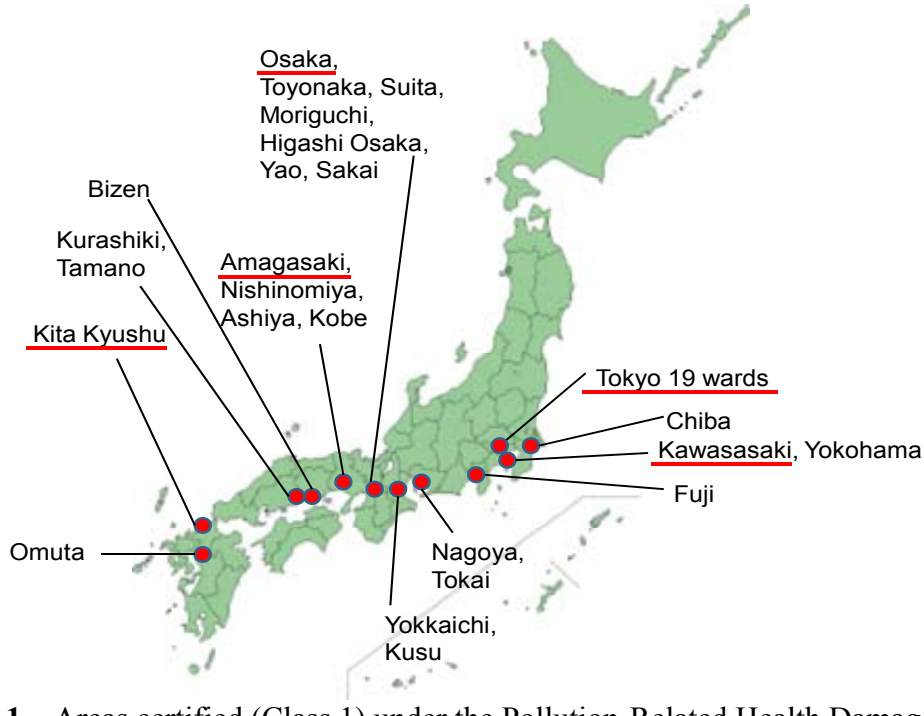


Figure 1. . Areas certified (Class 1) under the Pollution-Related Health Damage Prevention system (the research areas assessed in this paper are underlined)

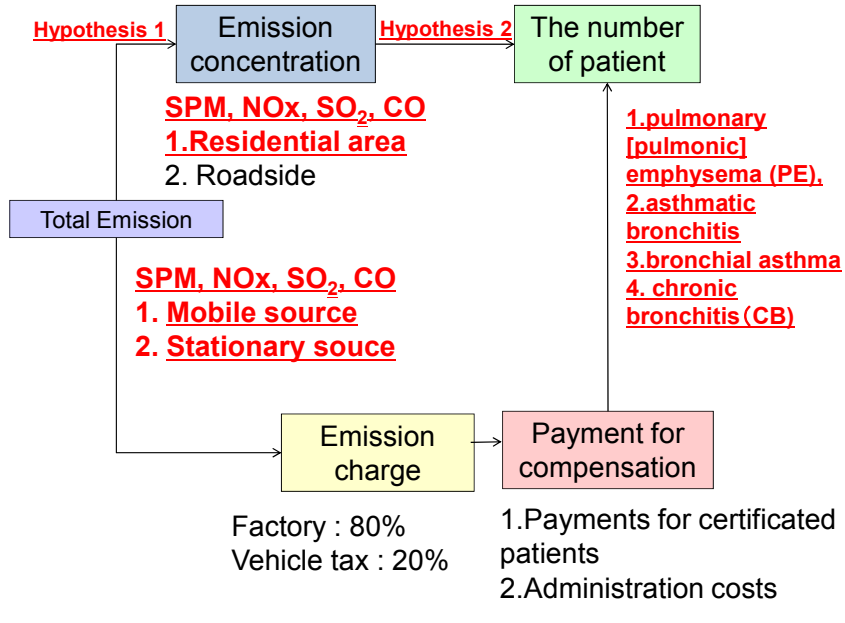


Figure 2. Causal relationships among total emissions, emission concentrations, and the number of certified patients (research focuses are underlined in bold).

Causal Relationships

In this study, 5 selected certified areas—Tokyo, Kawasaki, Osaka, Amagasaki, and Kita Kyusyu—were analyzed. Figure 2 shows the causal relationships among total emissions, emission concentrations, and the number of certified patients (research focuses are underlined). In this paper, we employed 2 hypotheses to assess the relationships; these hypotheses are stated below.

Hypotheses:

1. Total emission volume affects air quality (Hypothesis 1).
2. Air quality influences the number of certified patients (Hypothesis 2).

Total Emission Volume Affects Air Quality (Hypothesis 1). Figure 3 shows the causal relationship between total NOx emission (stationary and mobile sources) and NOx air quality (residential area, annual average) in Tokyo, Kawasaki, and Osaka. Figure 4 shows the causal relationship between total particulate matter (PM) emission (stationary and mobile sources) and PM air quality (residential area, annual average). The data for emission volumes (stationary and mobile sources, NOx and PM) were obtained by the respective city governments. We used SPM data without conversion to PM in order to show track record data.

Total PM, NOx, SO2, and CO emission volumes from stationary sources have been measured since the early 1970s in Tokyo, Kawasaki, Osaka, Amagasaki, and Kita Kyusyu. The total NOx and PM emission volume from mobile sources have been primarily measured since the 1970s to evaluate policy implementation. Tokyo, Kawasaki, and Osaka are not only certified areas according to the Compensation Law, but are also total PM and NOx emission reduction areas; therefore, data on emission volumes from mobile sources in these areas are also available. Amagasaki also belongs to a total Emission Reduction Area. Emission volume data for all cities in Hyogo prefecture are available (Hyougo, 2003). However, the data for Amagasaki alone are not available. Kita-Kyushu does not belong to a total PM and NOx emission reduction area; therefore, only data on total emissions from stationary sources are available for these areas (Kita Kyushu, 1982).

Data on total NOx and PM emission volumes (stationary and mobile sources) in Tokyo (23wards)(Tokyo, 1977; Amagasaki, 1976; Tokyo, 1994, 1996, 2000, 2004, 2005, 2009; NIES, 2008; Tokyo Metropolitan Government, March, 1971; Tokyo, 1973, 1983, 1995, 1998).Kawasaki (Kawasaki, 2008; Kawasaki, 2009), and Osaka (Osaka,1975, 1976, 1991, 2001, 2004). during 1972-2006 are available from various annual environmental reports by city governments. There were 2 outliners from Tokyo and 1 from Osaka. The 2 outliners of Tokyo are from 1980 and 1985; there are 2 estimated values of emission volumes for 1980 in 2 different reports of Tokyo Metropolitan Government. There are 4 estimated values of emission volumes for 1985 in 4 different reports of Tokyo Metropolitan Government. We used the highest values. The outliner of Osaka was 1972, the earliest year of emission volume estimation. It is the highest emission volume among data of Osaka.

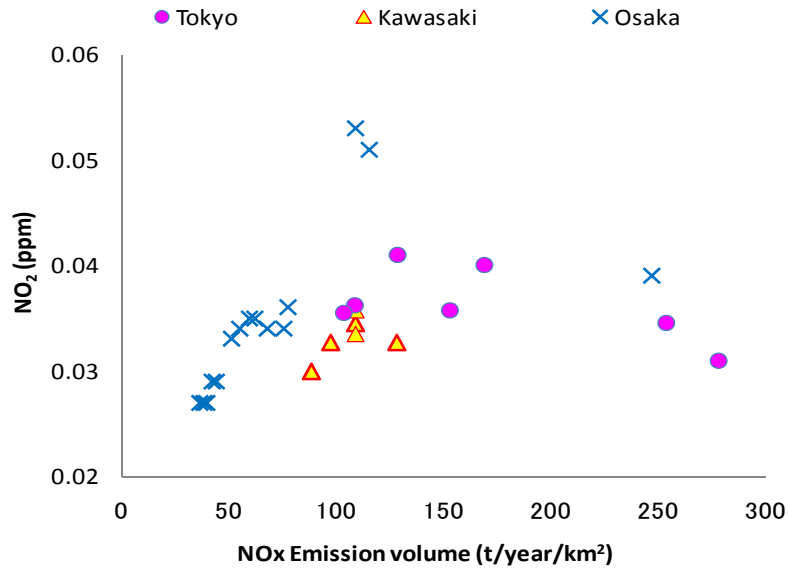


Figure 3. Total NO_x emissions and NO_x concentrations in Tokyo, Kawasaki, and Osaka.

Data sources: emission volume (Tokyo, 1977,1978, 1983, 1994, 1995, 1996, 2000, 2004, 2005, 2009; Kawasaki , 2008, 2009; Kita Kyushu, 1974; Matsuura; Osaka, 1971; 1976, 1991, 2001, 2004, 2009; Amagasaki, 1976).
air quality data (Kawasaki , 2008, 2009; Kita Kyushu, 1974; Osaka, 1971; 1976, 1991, 2001, 2004, 2009).

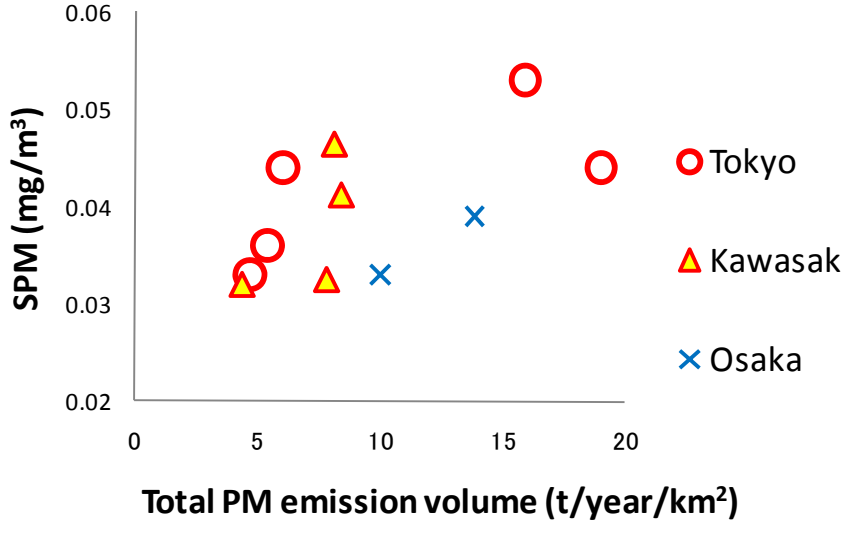


Figure 4. Total PM emissions and SPM concentrations in Tokyo, Kawasaki, and Osaka.

Data sources: emission volume (Tokyo, 1977,1978, 1983, 1994, 1995, 1996, 2000, 2004, 2005, 2009; Kawasaki , 2008, 2009; Kita Kyushu, 1974; Matsuura; Osaka, 1971; 1976, 1991, 2001, 2004, 2009; Amagasaki, 1976).
air quality data (Kawasaki , 2008, 2009; Kita Kyushu, 1974; Osaka, 1971; 1976, 1991, 2001, 2004, 2009).

For air-quality data, the annual average values of the pollutants (NO2, SPM) in residential areas in Tokyo (23 wards) (NIES, 2008), (Kawasaki, 2008, 2009). and (Osaka,1975, 1976, 1991, 2001, 2004). during 1960-2006 were used. Air quality in terms of NOx and SPM levels seemed to deteriorate with an increase in the NOx and PM emission volumes.

Air Quality Influences the Number of Certified Patients (Hypothesis 2). This research focused on causal relationship analysis between air quality (NOx, SPM, SO2, CO) and health effects (number of certified, recovered, and new patients, and total mortality) in the areas that were certified as Class 1 under the Pollution-related Health Damage Compensation System: Tokyo (19 wards[1]) (ERCA; Tokyo, 1973, 1978, 1985, 1995). Kawasaki (Kanagawa, 1973, 1982, 1985, 1988, 1994; Kawasaki, 2008), (Osaka, 1974), (Amagasaki, 2008, 2009), and Kita-Kyushu (Kyushu, 2008; Osaka, 2007). To verify the values for health effects, Eq. 1 was used. The difference betwee,n the estimated values and the values obtained from the track record was less than 5%. However, some of the track record values for these areas were not obtained because the local governments monitored either stationary or mobile sources according to air pollution reduction policies.

For air quality data, the annual averages of pollutants (NO2, SPM) in the residential areas of Tokyo (23 wards) (Matsuura), Kawasaki (Kawasaki, 2008,2009; NIES, 2008). (Osaka,1975, 1976, 1991, 2001, 2004), (Amagasaki, 1970, 1076), and (Kitakyushu, 2009) during 1960-2006 were used. The air quality data were converted, and the maximum air quality index (AQI) value was selected for assessments. AQI (London, 2006). is an index for reporting air quality with regard to public health. A lower AQI indicates better air quality[2]. Figure 5 shows the AQI values and the numbers of certified patients. An increase in the AQI value corresponded to an

[1] Class 1 in the Tokyo area included the following wards: Chiyoda, Chuo, Minato, Shinjuku, Bunkyo, Taitou, Sumida, Koutou, Shinagawa, Meguro, Ohta, Shibuya, Toshima, Kita, Arakawa, Itabashi, Adachi, Katsushika, Edogawa.

[2] http://www.airnow.gov/index.cfm?action=aqibasics.aqi

Break point				AQI	Category
PM10 (microgram/m3)	CO ppm	SO2 ppm	NO2 ppm		
0-54	0-4.4	0-0.034	0-0.02	0-50	Good
55-154	4.5-9.4	0.035-0.144	0.021-0.04	51-100	Moderate
155-254	9.4-12.4	0.145-0.224	0.041-0.06	101-150	Unhealthy for sensitive groups
255-354	12.5-15.4	0.225-0.304	0.061-0.64	151-200	Unhealthy
355-424	15.5-30.4	0.305-0.604	0.65-1.24	201-300	Very unhealthy
425-504	30.5-40.4	0.605-0.804	1.25-1.64	301-400	Hazardous
505-604	40.5-50.4	0.805-1.004	1.65-2.04	401-500	Hazardous

NO2 has no short term NAAQS. It can generate an AQI only above a value of 200
For analysis purpose, the values below 200 is created
according to health effect trend in Tokyo 1094-2006.

increase in the number of certified patients. Figure 6 shows the AQI values and mortality rates. An increase in the AQI value corresponded to an increase in mortality. The AQI values for Osaka from 1974 to 1984 were higher than the corresponding values for the rest of the years. The data for Osaka seems to be distributed into 2 groups. The AQI value in the right group in Figure 6 seems to be higher than that of the left group.

The mortality in Amagasaki from 1974 to 1975 was relatively lower than those in the rest of the years. Figure 7 shows the AQI and recovery values. An increase in the AQI value corresponded to a decrease in recovery. Figure 8 shows AQI values and the numbers of new patients. When the AQI value increased, new patient numbers seemed to increase. After 1989, the number of new patients was zero. There were 2 outliners from Amagasaki and 1 from Osaka. These data are from 1974 and 1975; in the early 1970s, the methodology was not well established. In fact, many experts have pointed out that the methodology for air-quality measurements was not well-established around 1974, when the permissible limits were introduced. Due to variable measurements, different results have been recorded in different governmental documents. Further, due to the introduction of the compensation system in 1974, the data for certified patients, new patients, mortality, and recovery for that year are also not stable. Because of the data instabilities caused by the introduction of the compensation system and air-quality measurements, some values are outside the general trend.

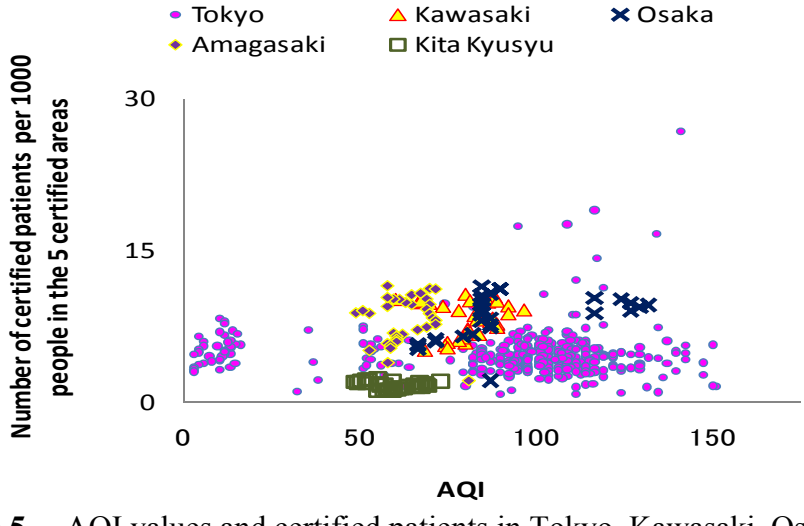


Figure 5. AQI values and certified patients in Tokyo, Kawasaki, Osaka, Amagasaki, and Kita Kyushu.

Data sources: air quality data: (Amagasaki,2008; ERCA; Kanagawa, 1973, 1982, 1985, 1988, 1993, 1994, 2008, 2009; Kawasaki, 2008; Kita Kyushu, 2008; Kita Kyushu, 1982, 1987, 2008; Matsuura; NIES,2008; Osaka,1974, 1975, 1976, 1991, 2001, 2004; Tokyo, 1973, 1978, 1985,1995)

health effects: (Amagasaki, 1970, 1976, 1985, 1998, 2001, 2006, 2008; ERCA; Fukuoka, 1970; Kanagawa, 1973, 1982, 1985, 1988, 1994; Kawasaki, 1993, 2009; Kita Kyushu, 1982, 1987; 2008, 2009; Osaka,1974; Tokyo, 1973, 1978, 1985,1995).

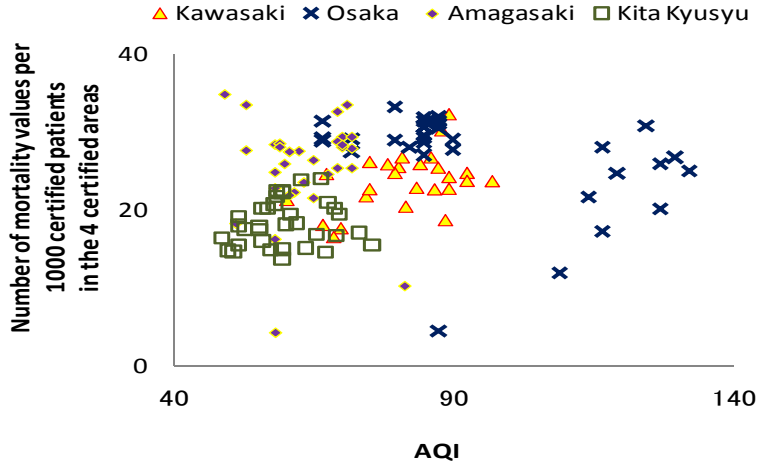


Figure 6. AQI values and mortality values of certified patients in Kawasaki, Osaka, Amagasaki, and Kita Kyushu.

Data sources: air quality data: (Amagasaki,2008; ERCA; Kanagawa, 1973, 1982, 1985, 1988, 1993, 1994, 2008, 2009; Kawasaki, 2008; Kita Kyushu, 2008; Kita Kyushu, 1982, 1987, 2008; Matsuura; NIES,2008; Osaka,1974, 1975, 1976, 1991, 2001, 2004; Tokyo, 1973, 1978, 1985,1995)
health effects: (; Amagasaki, 1970, 1976, 1985, 1998, 2001, 2006, 2008; ERCA; Fukuoka, 1970; Kanagawa, 1973, 1982, 1985, 1988, 1994; Kawasaki, 1993, 2009; Kita Kyushu, 1982, 1987; 2008, 2009; Osaka,1974; Tokyo, 1973, 1978, 1985,1995).

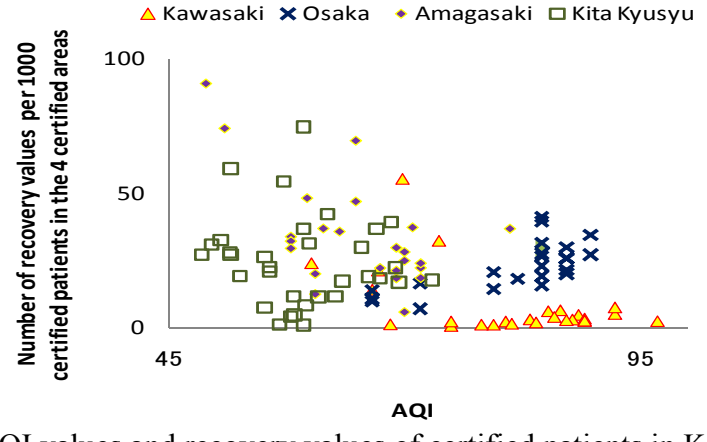


Figure 7. AQI values and recovery values of certified patients in Kawasaki, Osaka, Amagasaki, and Kita Kyushu.

Data sources: air quality data: (Amagasaki,2008; ERCA; Kanagawa, 1973, 1982, 1985, 1988, 1993, 1994, 2008, 2009; Kawasaki, 2008; Kita Kyushu, 2008; Kita Kyushu, 1982, 1987, 2008; Matsuura; NIES,2008; Osaka,1974, 1975, 1976, 1991, 2001, 2004; Tokyo, 1973, 1978, 1985,1995)
health effects: (; Amagasaki, 1970, 1976, 1985, 1998, 2001, 2006, 2008; ERCA; Fukuoka, 1970; Kanagawa, 1973, 1982, 1985, 1988, 1994; Kawasaki, 1993, 2009; Kita Kyushu, 1982, 1987; 2008, 2009; Osaka,1974; Tokyo, 1973, 1978, 1985,1995)

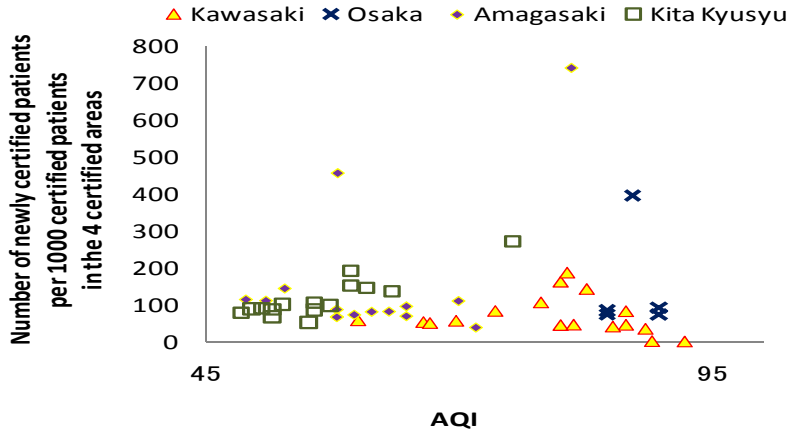


Figure 8. AQI values and newly certified patients in Kawasaki, Osaka, Amagasaki, and Kita Kyushu.

Data sources: air quality data: (Amagasaki,2008; ERCA; Kanagawa, 1973, 1982, 1985, 1988, 1993, 1994, 2008, 2009; Kawasaki, 2008; Kita Kyushu, 2008; Kita Kyushu, 1982, 1987, 2008; Matsuura; NIES,2008; Osaka,1974, 1975, 1976, 1991, 2001, 2004; Tokyo, 1973, 1978, 1985,1995)
health effects: (; Amagasaki, 1970, 1976, 1985, 1998, 2001, 2006, 2008; ERCA; Fukuoka, 1970; Kanagawa, 1973, 1982, 1985, 1988, 1994; Kawasaki, 1993, 2009; Kita Kyushu, 1982, 1987; 2008, 2009; Osaka,1974; Tokyo, 1973, 1978, 1985,1995)

Scenario Analysis

This section describes a scenario analysis to determine the accuracy of the statistics. To confirm the causal relationship presented in chapter 2, this research focused on estimating numbers of certified, recovering, and new patients and the mortality values. Mortality, recovery, and new-patient data are available for Kawasaki (1986-2006), Osaka (1974-2005), Amagasaki (1974-2006), and Kita Kyushu (1974-2006). Because the number of new patients in the track records is very limited, it was only estimated without comparison with its track record. The following scenarios are extracted from these 2 causal relationship analyses.

- The number of certified patients, mortality, and new patients per 1000 patients will increase with worsening of air quality.
- The number of recoveries per 1000 patients will decrease with worsening of air quality.

The US EPA (71) estimates mortality and the incidence of respiratory diseases from the disease per population values by using an exponential model. WHO estimates these values from the population with risk factors (USEPA, 1999). The UK Department of Health (69) prepares a list of estimated results from the literature for North America and Europe. The models are simplified from literature reports from major institutions, US EPA, WHO, and UK Department of Health and include studies on the relationship between mortality and respiratory diseases. Additionally, comparison of the main models from the US EPA, WHO, and UK Department of

Health suggest that a linear function seems to be better for fitness. For this analysis, we used 4 types of air concentrations because direct cause effects can be seen by type of air concentration.

$$\text{Certified patients}_t = \text{Certified patients}_{t-1} + \text{New patients}_t - \text{Recovery} - \text{Mortality} \quad \text{Eq. 1}$$

$$\text{Certified patients} = Ax + b \quad \text{Eq. 2}$$

$$\text{Recovery} = Bx + b \quad \text{Eq. 3}$$

$$\text{Mortality} = Cx + b \quad \text{Eq. 4}$$

$$\text{New patients} = Dx + b \quad \text{Eq. 5}$$

t: year

Certified patients: Certified patients per 1000 people

Mortality: Mortality per 1000 certified patients

Recovery: Recovery per 1000 certified patients

x: Air concentration in terms of annual average (PM10, NO_2, SO_2, and CO).

Table 1 shows results of estimation using Eq. 2, Eq. 3, Eq. 4, and Eq. 5

Table 1. Results of linear estimation between air pollutants by category of patient for Eq. 2, Eq. 3, Eq. 4 and Eq. 5 in Kawasaki, Osaka, Amagasaki, and Kita Kyushu.

Inverse correlation

Area	Category	NOx	PM10	SO2	CO
Kawasaki	New	-3584.2	2853.9	7877.4	111.2
	Mortality	336.0	-251.1	-795.0	-3.3
	Recovery	-1700.0	479.5	2717.7	29.7
	Certified	66.5	46.3	1.9	1.7
Osaka	New	-399.9	7455.4	19104.6	98.7
	Mortality	47.4	462.5	843.5	4.6
	Recovery	555.2	-4.3	-11.1	0.0
	Certified	132.7	14.4	-104.9	0.5
Amagasaki	New	1367.3	9231.8	30723.7	411.9
	Mortality	285.5	-155.1	1180.3	11.9
	Recovery	-350.6	1167.2	2814.1	46.1
	Certified	84.4	47.8	-78.2	-1.1
Kita Kyushu	New	9021.5	2795.5	9862.4	24.3
	Mortality	-15.5	-46.2	-36.1	-1.1
	Recovery	-1088.9	256.3	-349.1	1.1
	Certified	-79.4	4.7	-24.7	0.2

The relationship between air pollutants and new patients should have a positive coefficient. Amagasaki and Kita Kyushu had positive coefficients for all the pollutants. This result indicates that the number of new patients increases with a worsening of air quality. The rest of the areas had negative coefficients, indicating an inverse relationship.

The relationship between air pollutants and mortality should have a positive coefficient. Osaka had positive coefficients for all the pollutants. This result indicates that mortality usually increases with a worsening of air quality. The rest of the areas showed inverse relationships for certain pollutants.

The relationship between air pollutants and recovery should have a negative coefficient. This would indicate that the number of recovered patients decreases when the air quality worsens. Each area had positive (inverse relationship) and negative signs.

The relationship between air pollutants and certified patients should have a positive coefficient. Only Kawasaki had positive coefficients for all the pollutants. This result indicates that the number of certified patients usually increased when the air quality worsened. The rest of areas had negative coefficients, indicating inverse relationships.

These inverse relationships may have occurred because of the effect of the introduction of and amendments to the Compensation System since 1989. After 1989, applications by new patients ended, because air quality had improved. Only renewals of certifications were accepted.

Table 2. Results of linear estimation between air pollutants by category of patient for Eq. 2, Eq. 3 and Eq. 4 by period (Amagasaki).

Pollutant	Category	Period	Coefficient	Intercept	RSQ
NOx	Certificate	1974-1988	-26.24	9.74	0.00
		1989-2006	600.09	-8.33	0.65
	Mortality	1974-1988	105.49	19.82	0.00
		1989-2006	316.29	16.92	0.01
	Recovery	1974-1988	-529.69	51.92	0.01
		1989-2006	4014.68	-93.14	0.53
PM10	Certificate	1974-1988	-275.58	24.17	0.70
		1989-2006	217.15	-0.63	0.90
	Mortality	1974-1988	-630.17	56.88	0.42
		1989-2006	305.39	13.91	0.06
	Recovery	1974-1988	-54.59	41.88	0.00
		1989-2006	1361.63	-38.20	0.65
SO2	Certificate	1974-1988	-602.81	16.24	0.80
		1989-2006	885.48	1.61	0.83
	Mortality	1974-1988	-1006.28	34.33	0.26
		1989-2006	664.14	20.86	0.02
	Recovery	1974-1988	-361.10	43.17	0.00
		1989-2006	5642.24	-24.79	0.62
CO	Certificate	1974-1988	-9.15	17.85	0.54
		1989-2006	11.00	1.22	0.30
	Mortality	1974-1988	-14.37	36.16	0.15
		1989-2006	-9.82	30.75	0.01
	Recovery	1974-1988	-1.15	40.00	0.00
		1989-2006	3.59	14.47	0.00

The values used for the scenario analysis

With respect to the standard deviations of the numbers of certified patients, recovery, mortality, and new patients in these areas, the variance of the values for 1974-1975 was relatively large. Because the system started in 1974-1975, the number of new patients for this year would be unstable. Thus, during 1974-1989, the number of patients increased in spite of an improvement in the air quality.

The inverse relationships may cause over- or underestimation in the scenario analysis. To address this shortcoming, the causal relationship was estimated by separating the assessments for the first period (1974-1988) and the second period (1989-2006). Table 2 shows the example of Amagasaki. The values in shade were used for the scenario analysis for Amagasaki in the next chapter because they yielded relatively better results. Figure 9 shows the track record and estimation results obtained by using the coefficients and intercepts in Table 2 (Amagasaki). The results for the other cities (Kawasaki, Osaka, Kita Kyushu) are shown in the Appendix.

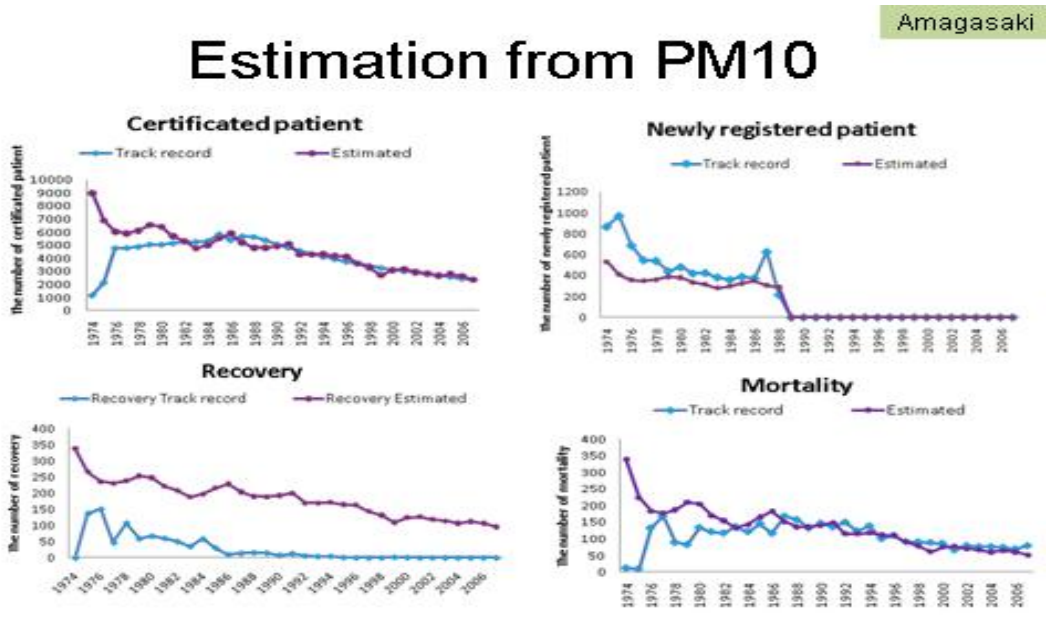


Figure 9. Track record and estimation results obtained using coefficients and the intercept in Table 2 (Amagasaki).

Fitness of the Results to the Track Record

The estimated results (Fig. 9 and Appendix) were compared with the track record to assess accuracy (Eq. 6). All t-value results were more than 2, except those for the relation CO and recovery.

Track recordn = Estimation n X + c.....Eq. 6

Estimation: Estimation results

Track record: Track record of certified patients, mortality, and recovery

n: certified patients, mortality, and recovery

X: coefficient

Table 3. Estimation results obtained using Eq. 6.

Track record = Estimated result X + b

Upper line: coefficient
Below line: t-value

	Certificated patient	R2	Mortality	R2	Recovery	R2	Newly registered patient	R2
NO2	0.657 21.469	0.881	0.197 5.594	0.442	0.845 7.994	0.576	−0.113 0.014	−0.160
PM10	0.624 15.457	0.801	0.289 5.533	0.438	−0.990 −3.173	0.269	1.750 18.742	0.852
SO2	−0.900 −3.629	0.300	−0.329 −1.402	0.123	−1.244 −4.249	0.350	1.296 4.512	0.364
CO	−0.737 −2.912	0.245	−0.836 −2.555	0.219	0.005 0.018	0.002	1.276 5.374	0.422

Inverse relationship

Table 3 shows the results obtained using Eq. 6. Most of coefficients are supposed to be positive. However, there are some inverse relationships. Possible causes for these inverse relationships may include data quality. Estimation with only air quality data may cause over- or underestimation of certified, new, or recovered patients, or mortality.

Conclusions

This paper sought to assess the causal relationships among emission volumes, air quality, and health effects in Japan during 1974-2006. First, the relationship between emission volumes and air quality was clarified using data on emission volumes and air quality in Tokyo, Kawasaki, and Osaka. Air quality (SPM, NOx) tended to worsen with an increase in emission volumes (stationary and mobile sources).

Second, the relationship between air quality and health effect was clarified using the health effect data from the Compensation System in Tokyo, Kawasaki, Amagasaki, Osaka, and Kita-Kyushu. The numbers of certified and new patients and mortality tended to increase when the air quality worsened. Recovery tended to decrease when the air quality worsened. However, inverse relationships were also observed in the linear estimation results for the period between 1974 and 2006. To increase accuracy, the coefficients were reexamined by separating the periods 1974-1988 and 1989-2006.

Finally, to support hypothesis 2, the numbers of certified, new, and recovered patients and mortality were estimated by a simple general equilibrium model. These estimated results were compared with the track records. Most of coefficients would be expected to be positive. However, there were also some inverse relationships. Estimations using only air quality data may cause over- or underestimation of certified, new, or recovered patients, or mortality. Possible causes of these inverse relationships may include data quality and missing data.

Acknowledgements

This research was conducted as a part of the research for "Economic Analysis for Emission Reduction Measures from automobile sector in Asian Large cities". I am grateful to Mr. Hirokazu Sugizaki (ERCA), Mr. Toshiro Kokaji (Safety and Environmental Bureau, Osaka Prefecture Government), Mr. Yamada (Environmental Protection Division, Kita Kyushu City Government), Mr. Satoshi Kakurai (NEXCO), Mr. Michiya Kousaka (Library of Health and Safety Center, Tokyo Metropolitan Government) for data, librarians Tokyo Metropolitan Institute for Environmental Protection for advice of data collection, many colleagues of Kashima Seminar (Chuo University), Dr. Tazuko Marikawa (JARI), Dr. Naomi Tsukue (JARI), Dr. Koichiro Ishii (Tokyo Metropolitan Institute for Environmental Protection) for discussion. I am solely responsible for all the errors in this paper.

Reference

Air Pollution Division Pollution Regulation Bureau Tokyo Metropolitan Government. Air Pollution Emission Survey (in Japanese); Tokyo Metropolitan Government, Tokyo. (1977), 41-44.

Amagasaki City Government. Amagasaki Statistical Yearbook (in Japanese); *Amagasaki City Government*, Amagasaki. (2008), 129-130.

Amagasaki City Government. Environment in Amagasaki (in Japanese); *Amagasaki City Government*,Amagasaki.(2008), 23.
http://www.city.amagasaki.hyogo.jp/sogo_annai/toukei/033amakan.html (Accessed on 20 June 2009).
Amagasaki City Government. Environment in Amagasaki. (in Japanese); A*magasaki City Government,* Amagasaki.(1998), 125-128.
Amagasaki City Government. Environment in Amagasaki. (in Japanese); *Amagasaki City Government,* Amagasaki.(2001), 197.

Amagasaki City Government. Environment in Amagasaki. (in Japanese); *Amagasaki City Government*, Amagasaki.(2006), 129-130.

Amagasaki City Government. Pollution in Amagasaki (Amagasaki Environmental White Paper) (in Japanese); *Amagasaki City Government*, Amagasaki.(1974), 192-195.

Amagasaki City Government. Pollution in Amagasaki. (in Japanese); *Amagasaki City Government*, Amagasaki.(1970), 221-228.

Amagasaki City Government. Pollution in Amagasaki. (in Japanese); Amagasaki City Government, Amagasaki.(1976), 9.

Amagasaki City Government. Pollution in Amagasaki. (in Japanese); *Amagasaki City Government,* Amagasaki.(1985), 164-165.

Environmental Bureau of Tokyo Metropolitan Government. Total Automotive Emission Reduction Projection in the Future (in Japanese); *Tokyo Metropolitan Government*, Tokyo.(2000).

Environmental Bureau of Tokyo Metropolitan Government. Total Emission Reduction Plan Promotion Management Survey (in Japanese); Tokyo Metropolitan Government, Tokyo.(1994), 129.

Environmental Bureau of Tokyo Metropolitan Government. Total Emission Reduction Plan Promotion Management Survey (in Japanese); *Tokyo Metropolitan Government,* Tokyo.(1996), 128.

Environmental Bureau of Tokyo Metropolitan Government. Total Emission Reduction Plan Promotion Management Survey (in Japanese); T*okyo Metropolitan Government*, Tokyo.(2000), 10.

Environmental Bureau of Tokyo Metropolitan Government. Total Emission Reduction Plan Promotion Management Survey (in Japanese); *Tokyo Metropolitan Government*, Tokyo.(2004), 6.

Environmental Bureau of Tokyo Metropolitan Government. Total Emission Reduction Plan Promotion Management Survey (in Japanese); *Tokyo Metropolitan Government,* Tokyo.(2005), 8.

Environmental Bureau of Tokyo Metropolitan Government. Total Emission Reduction Plan Promotion Management Survey (in Japanese); *Tokyo Metropolitan Government*, Tokyo.(2009), 10.

Environmental Bureau of Tokyo Metropolitan Government. Total PM Emission Survey 2003 (in Japanese). *Tokyo Metropolitan Government*, Tokyo.(2000), 6-12.

Environmental Protection Bureau of Tokyo Metropolitan Government. Trend of Pollution from Automobile (in Japanese); *Tokyo Metropolitan Government*, Tokyo.(1984). 6.

Environmental Restoration and Conservation Agency of Japan (ERCA). The Pollution-related Health Damage Prevention Program.
http://www.erca.go.jp/english/activities/ac_02.html (accessed on 23 July 2008).

ERCA. Pollution-related Health Damage Compensation; ERCA annual report. Chapter 3.

Fukuoka Prefecture Government. Fukuoka Environmental White Paper (in Japanese); Fukuoka Prefecture, Fukuoka.(1974), 26.

Fukuoka Prefecture Government. Fukuoka Environmental White Paper (in Japanese); Fukuoka Prefecture, Fukuoka. (1978), 17.

Fukuoka Prefecture Government. Fukuoka Environmental White Paper(in Japanese); Fukuoka Prefecture, Fukuoka.(1970), 15.

Hyougo Prefecture Government (2003). Automobile NOx and PM Emission Reduction Plan; *Hyougo Prefecture Government*. Hyougo, 1.

Kanagawa Prefecture Government. Kanagawa Environmental White Paper (in Japanese); *Kanagawa Prefecture Government,* Kanagawa.(1973), 24.

Kanagawa Prefecture Government. Kanagawa Environmental White Paper (in Japanese); *Kanagawa Prefecture Government*, Kanagawa.(1982), 209.

Kanagawa Prefecture Government. Kanagawa Environmental White Paper (in Japanese); *Kanagawa Prefecture Governmen*t, Kanagawa.(1983), 398.

Kanagawa Prefecture Government. Kanagawa Environmental White Paper (in Japanese); *Kanagawa Prefecture Government*, Kanagawa.(1985), 187.

Kanagawa Prefecture Government. Kanagawa Environmental White Paper (in Japanese); *Kanagawa Prefecture Government*, Kanagawa.(1988), 105.

Kanagawa Prefecture Government. Kanagawa Environmental White Paper (in Japanese); *Kanagawa Prefecture Government*, Kanagawa.(1994),129.

Kawasaki City Government. Environment in Kawasaki (in Japanese).(2008), 12.

Kawasaki City Government. Kawasaki City Environmental White Paper (in Japanese); *Kawasaki City Governmen*t, Kawasaki.(2009), 12.

Kawasaki City Government. Pollution in Kawasaki (Kawasaki Environmental White Paper) (in Japanese); *Kawasaki City Government*, Kawasaki.(1993), 26.

Kawasaki City Government. Pollution in Kawasaki (Kawasaki Environmental White Paper) (in Japanese); *Kawasaki City Government*, Kawasaki.(1994), 103.

Kawasaki City Government. Statistics in Kawasaki (in Japanese); *Kawasaki City Government,* Kawasaki.(2008).

http://www.city.kawasaki.jp/20/20tokei/home/tokeisyo/tokeisyo20/html/mokuji.htm (Accessed on 2 May 2009).

Kita Kyushu City Government. (2007) "Kita Kyushu Air Pollution Prevention Plan (in Japanese).", http://www.city.kitakyushu.jp/pcp_portal/PortalServlet?DISPLAY_ID=DIRECT&NEXT_DISPLAY_ID=U000004&CONTENTS_ID=1964 (Access on 20 June 2009)

Kita Kyushu City Government. Environment in Kita-Kyushu. Appendix, *Kita Kyushu City Government*, Kita Kyushu.(2008), 11.

Kita Kyushu City Government. Environment in Kita-Kyushu; Kita *Kyushu City Government*, Kita Kyushu.(2008), Appendix 11-73.

Kita Kyushu City Government. Pollution in Kita-Kyushu. Appendix, *Kita Kyushu City Governmen*t, Kita Kyushu.(1987), 215.

Kita Kyushu City Government. Pollution in Kita-Kyushu; *Kita Kyushu City Government*, Kita Kyushu.(1981), Appendix 243.

Kita Kyushu City Government. Pollution in Kita-Kyushu (Kita Kyushu Environmental White Paper) (in Japanese); *Kitakyushu City Government*, Kita Kyushu.(1974), 246-247.

Kita Kyushu City Government. Pollution in Kita-Kyushu; *Kita Kyushu City Government*, Kita Kyushu.(1982), Appendix 19.

Kita Kyushu City Government. Pollutiont in Kita-Kyushu; *Kita Kyushu City Government*, Kita Kyushu, Appendix .(1974), 87.

Matsuura, I. Process of establishment of Pollution-related Health Damage Compensation Law, Jurist.(1984), No.821, 29-35.

Matsuura, I. Process of establishment of Pollution-related Health Damage Compensation Law, Jurist; No.822, 80-85.

Matsuura, I. Process of establishment of Pollution-related Health Damage Compensation Law, Jurist; No.824, 91-97.

National Institute for Environmental Studies. http://www-gis.nies.go.jp/ (Accessed 27 August 2008).

Osaka City Government. Osaka City Statistical Yearbook; *Osaka, Osaka City Government*, Chapter 20, Table 25.

Osaka City Government. Osaka City White Paper; Osaka, *Osaka City Government.*(1975), 48-49.

Osaka City Government. Osaka City White Paper; Osaka, *Osaka City Government*.(1976), 76.

Osaka City Government. Osaka City White Paper; Osaka, *Osaka City Government*.(1991), 25.

Osaka City Government. Osaka City White Paper; Osaka, *Osaka City Government*.(2001), 36.

Osaka City Government. Osaka City White Paper; Osaka, *Osaka City Government*.(2004), 31.

Osaka City Government. Osaka City White Paper; Osaka, *Osaka City Government*.(2007), 36.

Osaka City Government. Pollution in Osaka (Osaka Environmental White Paper) (in Japanese); *Osaka City Government*, Osaka.(1974), 169.

Osaka City Government. Statistical data of certificated Patients and cancelation. (Personal contact: 04 July 2009).

Pollution Bureau of Tokyo Metropolitan Government. Air Pollutants Emission Estimation Survey (in Japanese); *Tokyo Metropolitan Government*, March. (1971), 54-55.

Pollution Regulation Bureau of Tokyo Metropolitan Government. Air Pollutant Emission Factor Survey (in Japanese); *Tokyo Metropolitan Government*, Tokyo,July.(1973), 149.

Tokyo Metropolitan Government. Number of Certificated Patients and Mortality, Tokyo Health Annual Report; *Tokyo Metropolitan Government*, Tokyo. (1985), No.37, 247-262.

Tokyo Metropolitan Government. Number of Certificated Patients and Mortality. Tokyo Health Annual Report; *Tokyo Metropolitan Government*, Tokyo. (1995), No.47, 276-277.

Tokyo Metropolitan Government. Number of Certificated Patients and Mortality. Tokyo Health Annual Report; *Tokyo Metropolitan Government*, Tokyo. (1999), No.51, 272-273.

Tokyo Metropolitan government. Pollution Prevention Plan in Tokyo Region; *Tokyo Metropolitan Government*, Tokyo.(1978).

Tokyo Metropolitan Government. Tokyo Environmental White Paper; *Tokyo Metropolitan Government*, Tokyo.(1983), 79

Tokyo Metropolitan Government. Tokyo Environmental White Paper; *Tokyo Metropolitan Government*, Tokyo. (1995), 75.

Tokyo Metropolitan Government. Tokyo Environmental White Paper; *Tokyo Metropolitan Government*, Tokyo. (1998), 32.

Tokyo Metropolitan Government. Tokyo Regional Pollution Prevention Plan (in Japanese); *Tokyo Metropolitan Government*, Tokyo. (1978), 6.

Tokyo Metropolitan Government. Tokyo Statistical Yearbook; *Tokyo Metropolitan Government*.(1998), 32.

UK Department of Health (UK DH). Cardiovascular Disease and Air Pollution. A report by the Committee on the Medical Effects of Air Pollutants; UK DH, London,(2006), 21-137.

US EPA, Air Quality Index; USEPA, Washington D.C..
http://www.airnow.gov/index.cfm?action=aqibasics.aqi (Access:ed on 30 June 2009).

USEPA. The Benefits and Costs of the Clean Air Act 1990 to 2010; USEPA, Washington D.C.(1999), D-58.

World Health Organization. *The World Health Report*; WHO, Geneva.(2002), 19.

Optimization of Mode Share for Low-carbon-oriented Urban Passenger Transportation System

Liang X. PI[1,4,5], Yamamoto Toshiyuki[2] and Morikawa Takayuki[3]

[1] Doctorate Candidate, [2]Associate Professor, Department of Civil Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Japan. 464-8603. Email: pxl_iii@hotmail.com.
[3] Professor, Graduate School of Environmental Studies, Nagoya University, Japan.
[4] Doctorate Candidate, Transportation Engineering School, Tongji University, Shanghai, China
[5] Lecturer, Shanghai Maritime University, Shanghai, China

Abstract

To develop the public transportation system and change the passenger transport mode share, as well as to control the resorting to private cars have been considered one of the most effective methods to reduce carbon emission. This paper aims at determining an optimal mode share scheme for the urban passenger transportation system, which meets the carbon emission limitation and the government investment constraint at the lowest social cost. This optimal mode share scheme could work as a suggested mode split target for a low-carbon city. Stackelberg game theory based bi-level modeling is used as the modeling approach because there are multiple decision makers involved. The objective of the upper level model will work as decision making assistance for the government in determining an optimal mode share scheme of the urban passenger transportation system, which meets the low carbon emission, the government investment and some other constraints, at the lowest social cost, including pollution treatment fee for CO_2 and construction fee for different traffic systems in different district. Classic UE equilibrium based automobile traffic flow assignment model and Non-UE transit flow assignment method on the network is the second level; it is built on behalf of users' consideration, and to reflect the performance of network with the changing of mode share, as well as to build the loop between two levels with the changing of average speed of vehicles. Algorithm for bi-level modeling is also studied and a sample network is used to test the workability of the proposed modeling approach.

Key Words: low-carbon; bi-level programming problem; multi-mode passenger transportation; mode share

Introduction

During the past decades, transport emission and traffic congestion became the crucial problems of many cities, especially it has been highlighted to how to reduce the carbon emission in the urban transportation system during the rapid urbanization in developing countries, because in most cases it is more difficult to meet the limitation of emission than to meet that of capacity of the network. On the other hand, to develop the public transportation system and change the passenger transport mode share, as well as to control the resorting to private cars have been considered one of the most effective methods to reduce carbon emission (Kenworthy et al. 1992).

Studies on transportation environment conducted by Carslaw and Beevers (2002) and Shrestha (2005) have only focused on the supply side regardless of the users, and they also did not consider changes in travel demand that could occur due to changes in transport mode share schemes. Furthermore, they usually do not consider the cost of travel time and congestion; in fact, there could be significant difference in travel time and congestion level between different mode share schemes, and the flow variations on the network could also influence the emission factors. Eco-city designing principles have been brought forward in the studies of Register (1987), Roseland (1997) etc. but they did not provide the specific application methods for urban transportation system.

This paper aims at determining an optimal mode share scheme for the urban passenger transportation system, which meets the carbon emission limitation and the government investment constraint at the lowest social cost (SC). And this optimal mode share scheme could work as a suggested mode split target for a low-carbon city.

Methodology

Stackelberg game theory is used as the modeling approach because there are multiple decision makers involved (Benson 1989; Vicente et al. 1996). Moreover, they are in different layers in the whole decision making system. The upper layer has guide rights just like the government or transportation system management agencies, who can make the decision of what kind of mode share should be achieved under the low-carbon-constraint and investment constraint. And the lower layer, the users of transportation system, can also implement his decision making right under its selection range to choose the travel route they prefer to, according to the generalized travel cost.

Optimal Low-Carbon-Oriented Passenger Transportation Mode Share

Different mode has a definitely different carbon emission contribution. Thus we can calculate the primal mode split under the consideration of environment disutility of different traffic mode. Meanwhile, different traffic mode costs different levels of construction fee. Usually, urban transit system is the lowest carbon emission

passenger transport system, including bus system and subway system; however, they perhaps are more expensive system simultaneously. The government should consider these two aspects, carbon emission performance and investment performance, when they plan an optimal low-carbon-oriented mode share in the planning year.

The objective of the model will work as decision making assistance for the government in determining an optimal mode share scheme of the urban passenger transportation system, which meets the low carbon emission, the government investment and some other constraints, at the lowest social cost. Total cost includes pollution treatment fee for different carbon emission, e.g., CO2 and CO, and construction fee for different traffic modes in different district. The objective function can be expressed as:

$$\min \sum_i \sum_b \sum_k P_{ki}^{G*} x_{ki}^b E_{ki}\left(\overline{V_i^b}\right) + x_i^b d_i^b \tag{1}$$

where $P_{ki}^{G*} x_{ki}^b E_{ki}\left(\overline{V_i^b}\right)$ is the costs of carbon emission price according to emission market, $P^G$is the cost of carbon emission price per passenger kilometer for mode i emission type k , x_i^b is passenger kilometer of mode i emission type k in zone b, $E_{ki}\left(\overline{V_i^b}\right)$ is carbon emission factor for passenger kilometer of mode i emission type k in zone b, it is a function of the average speed $\overline{V_i^b}$ in zone b (Pattas 1994, Nagurney 2000). The emission factor for every mode is calculated under the consideration of average vehicle sharing coefficient of this mode. $x_i^b d_i^b$ is construction fee for different traffic mode in different district. x_i^b is passenger kilometer of mode i in zone b, d_i^b is construction fee of per passenger kilometer of mode i in zone b. The function of carbon emission factor is given as:

$$E_{ki}\left(\overline{V_i^b}\right) = \frac{\alpha_i^k \exp\left(\beta_i^k \overline{V_i^b}\right)}{\gamma_i^k \overline{V_i^b}} \tag{2}$$

where α_i^k, β_i^k, γ_i^k are parameters for carbon emission factor of mode i type k.

Carbon Emission Constraint. Total passenger transport emission of type k in the planning year must be less than or equal to the carbon emission upper limit G_k, which is decided by the environmental consideration or by national/local environmental regulations.

$$\sum_i \sum_b x_i^b E_{ki}\left(\overline{V_i^b}\right) \le G_k \tag{3}$$

where $\overline{V_i^b}$ is average speed in zone b of mode i. The average speed can be obtained from the lower level of modeling structure, which is introduced in the next section of this paper.

Investment Constraint. Total passenger transport facilities land use and investment of different modes should be less or equal to the uper limit *I*, which is decided by a specific situation of a given city.

$$\sum_i x_i d_i \le I \tag{4}$$

Other Constraints. Total supply of all modes of urban passenger transportation should be equal to the demand forecasting for the planning year D; and supply of mode i should be greater than or equal to the lower limit (S_{iMin}) of passenger in the planning year. At the same time, to avoid the waste of existing equipment and recourses, it should be less than or equal to upper limit (S_{iMax}) of passenger in the planning year. All of them are with demission of person kilometer.

$$\begin{aligned} &\sum_i \sum_b x_i^b = D \\ &S_{iMin}^b \le x_i^b \le S_{iMax}^b \end{aligned} \tag{5}$$

Joint Traffic Flow Assignment on the Network

Classic UE equilibrium based automobile traffic flow assignment model and Non-UE transit flow assignment method, that means the flow assignment of transit line is according to the transit vehicle departure frequency, are jointly used to obtain traffic flow on every link (Sheffi 1984). Objective function of UE equilibrium can be expressed as:

$$\min_y \sum_a \int_0^{y_a} t_a(y)dy$$

$$s.t. \begin{cases} \sum_o f_o^{rs} = \dfrac{x_3}{\sum_i x_i} D^{rs}, \quad \forall r,s \\ f_o^{rs} \ge 0, \quad \forall o,r,s \\ y_a = \sum_r \sum_s \sum_o f_o^{rs} \delta_{a,o}^{rs} \end{cases} \tag{6}$$

where, for the mode automobile, y_a is traffic flow of automobile on link a, t_a is traffic time on link a, f_o^{rs} is traffic flow on route o. Mean while, we can obtain the average speed $\overline{V_3^b}$ of zone b as an output by:

$$\overline{V_3^b} = \frac{VMT_3^b}{VHT_3^b} = \frac{\sum_{a \in b} l_a y_a}{\sum_{a \in b} t_a y_a} \tag{6}$$

where VMT_i^b is vehicle miles traveled in zone b of mode i, VHT_i^b is vehicle hours traveled in zone b of mode i, l_a is length of link a. And it is regarded that the average speed of mode bus $\overline{V_5^b}$ equal to $0.65\overline{V_3^b}$ in the same zone. Here the

parameter of 0.65 is just an assumed parameter for research job, it could be adjusted according to specific situation of a certain city.

Framework of Bi-Level Modeling

Optimal low-carbon-oriented passenger transportation mode share is the first level, a modeling for government decision making on what kind of mode share scheme should be a suggested mode share for a low-carbon city, so this part is built under the benefit of government consideration. Meanwhile, it directly influences the mode choice by users.

Joint traffic flow assignment on the network is the second level; it is built on behalf of users' consideration. Furthermore, the calculation should begin with this part with the actual data for the actual situation of a city, and there is a loop between the two levels. The change of mode share would influence the traffic flow assignment and the flow assignment result would produce new speed on every link, consequently produce new emission factors, and then cause a new optimal mode share scheme. The difference between sequential iterations should gradually decrease before a convergence criterion is satisfied. Algorithm for bi-level modeling is as follows:

Step 1. According to the assignment results of the super net which is based on OD data and existing mode split of present situation of certain city, calculate the initial values for carbon emission factor.
Step 2. With emission factor and energy consumption factor, solve the first level model and receive the initial primal mode split as the new input to the second level model.
Step 3. With the mode split results of step 2, compute the UE assignment again on the super net and compute new carbon emission factor for next loop.
Step 4. Go to step 5 if objective function of the second level is not improved; otherwise repeat from step 2.
Step 5. Receive the suggested mode split scheme form step 2 and investigate the network performance from the assignment results of step 3.

A Numerical Example

An artificial simple network is used to test the workability of the proposed modeling approach. To simplify, there are only two modes, automobile and bus, and two links in this network as Figure 1. And only CO_2 is considered as the carbon emission.

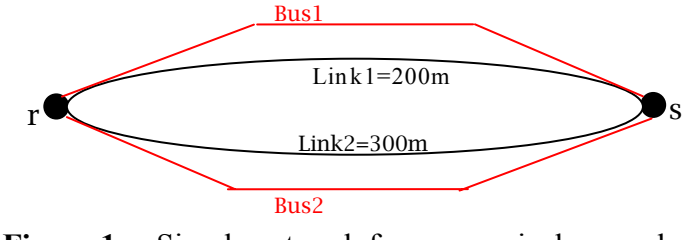


Figure 1. Simple network for a numerical example.

Calculation Input. In this network, the traffic flow data is assumed as known for input of calculation, and assumed that the flow assignment of bus line is according to the bus departure frequency. Assume first OD flow between r and s, Q_{rs}^1 is equal to 100, and existing mode split is $(Auto:Bus)^1$ equal to 60:40, link performance function is $t = 10 + 0.2y_a^2$ for both links, bus lines departure frequency ratio is $f_{bus1}:f_{bus2}=1:3,\quad d_5=1.7, d_3=1700$. For the planning year, OD flow between r and s, $Q_{rs}=120$, and passenger kilometer of passenger overturn $D=30$. The threshold for the ending of the loop between two levels is set as that when the improvement of the second level is less than or equal to 2%. Emission factor of automobile and bus in CO_2 emission (k = 1) is $a_{1,3}=8a_{1,5}=3.3963\dfrac{e^{0.01456\overline{V}_3}}{1000\overline{V}_3}$ (Penic and Upchruch, 1992). The planning year's G_1 is equal to $0.8G_1^0$, which is calculated from the existed situation.

Calculation Result. The interative calculation between the two levels got convergence after three loops, and the optimized mode split meeting with the eco-constrains at the planning year is $(Auto:Bus)^4$=44:56. Output at each interation is shown in Table 1. The results indicate that the mode split of auto and bus in the planning year should shift from current 60:40 to 44:56 with the CO_2 mitigation to 80% of current level. Note that in this example, no investment and construction is included, the objective function, SC, directly represents CO_2 emission cost.

Conclusion and Future Research

Stackelberg game theory based bi-level modeling has following advantages in the application to optimize the mode share for low-carbon-oriented urban passenger transportation system:

Table 1. Output of the optimization at each interation.

	Q_{rs}	$\overline{V_3}$	$Auto:Bus$	SC	SC improve rate
Existing situation	100	48.61	60:40	529432	
Loop 1	120	50.84	50:50	487607	7.9%
Loop 2	120	51.25	45:55	472004	3.2%
Loop 3	120	51.37	44:56	463508	1.8%

- ✧ Different objectives can be analyzed simultaneously during the decision-making process
- ✧ Bi-level criteria for transport management and planning section, usually the government, and the users, can be treated as more realistic, and the interaction between them can be described properly.
- ✧ Modular structure of the modeling achieves a good flexibility for every module, and makes the whole structure have a nice portability in different cities with varying contexts.

More complicated and realistic example would be investigated in the future research, and the sensitivity analysis would be carried out. And the study has considered only the passenger transport modes and has not included freight transportation options yet. One of the key factors of the mode share problem is the values of parameters of the model. It would be given more descriptions in the future research. Actually, the performance of a mode share strategy is very sensitive to variations in specification, and particularly to changes in service fares and charges for car use. And not surprisingly the optimal mode share strategy differs considerably between cities, and is dependent in particular on the inherited levels of infrastructure provision, subsidy for public transport, and congestion. All these merit further research.

Acknowledgments

The authors would like to thank to Prof. Yoshitsugu Hayashi for his comments on the earlier version of this paper.

References

Anthony D. May, Simon P. Shepherd, and Paul M. Timms. (2000). "Optimal Transport Strategies for European cities." Transportation, 27:285-315, 2000.

Benson H. P.(1989). "On the structure and properties of a linear multilevel programming problem." *Journal of Optimization Theory and Applications*. 60(3), 353-373.

Carslaw C. D. and Beevers D. S. (2002). "The efficacy of low emission zones in center London as a means of reducing nitrogen dioxide concentration." Transportation Research Part D, 7, 49-64.

Kenworthy J. R., Newman, P., and Lyons T. J. (1992). "The ecology of urban driving. I-Methodology." *Transportation Research.* Part A, 26, 263-272.

Nagurney A. (2000). "Congested urban transportation netwoks and emission paradoxes." *Transportation Research* Part D, 5, 145–151.

Pattas, K. N., Kyriakis N. A., and Samaras Z. C. (1994). "Actual emissions of vehicles of the NI category." *Science of The Total Environment*, 146(7), 191-199.

Penic M.A. and Upchruch J. (1992). "transyt-7f, Enhancement for fuel consumption, pollution emissions and user costs." *Transportation Research* Record 1360.

.

Register, R. (1987). "Eco-city Berkeley: Building Cities for A Healthier Future." CA: North Atlantic Books.

Roseland, M. (1997). "Dimensions of the future: an eco-city overview." *Eco-city Dimension, Edited by Roseland M*., 1-12. New Society Publishers.

Sheffi, Y. (1984). "Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods." Englewood Cliffs, N.J., Prentice-Hall.

Shrestha, R.M., Anandarajah, G, Adhikari, S., Jiang, K., and Songli, Z. (2005). "Energy and environmental implications of NOx emission reduction from the transport sector of Beijing: a least-cost planning analysis." *Transportation Research* Part D, 10, 1-11.

Vicente, L., Savard, G., and Judice J. (1996). "Discrete linear bilevel programming problem." *Journal of Optimization Theory and Applications.* 89(3), 597-614

A Comparative Study on Chinese and Foreign Road Transport Energy Consumption Statistical Indicators

Qipeng Sun[1 2], Baohua Mao[2] and Shunping Jia[3]

[1] Associate Professor, School of Economics and Management, Chang'an University, Cuihua Road, School of Economics and Management of Chang'an University, Xi'an, Shaanxi Province Xi'an 710064, China; Email: sunqip2003@163.com
[2] Prof. [3] Prof., Integrated Transport Research Center of China, Beijing Jiaotong University, Beijing 100044, China

Abstract: To improve the statistical indicators for road transport energy consumption and accurately analyze its energy consumption level of China, the article analyzes energy statistics scope, indicators, vehicle classification and energy consumption proportion of road transport in China, the United States, Britain and Japan. The horizontal and vertical comparison methods are used to propose differences in energy consumption statistics between China and developed countries and to put forward the trends of China's energy consumption for road transport. Studies have shown as follows: the scope of China's road transport energy consumption statistics has differences from the international accepted patterns; at present, energy consumption statistical indicators for China lack the conveying efficiency of transport vecicle; the proportion of China's energy consumption for road transport to the total society is far lower than that of foreign countries, but it can be rapidly increased in the future.
Key Words: transportation management; energy consumption; road transport; statistical indicators; China and Foreign comparison

Introduction

Energy stress has been the focus around the world, while the energy consumption statistical data of transportation are the highly concerned information at home and abroad. Road transportation is taking up the biggest part of energy consumption.

Accurate and objective statistical data of road transport energy consumption can provide the government with information support and decision-making basis on energy management. But compared with developed countries, current energy consumption statistics for China's road transport still has many shortages, which has brought about tremendous obstacles to accurately grasp the energy level and the development trend. For this, many scholars have conducted a lot of researches, such as Li Lian-Cheng (2008),Song Min-hua (2009), Li Xian-sheng (2007), Peng Jia-hui (2008), Shi Jing-ya (2008) and so on. They have done useful exploration about the transport energy consumption statistical indicators from different angles. This article conducts comparative study of energy consumption indicators on the basis of the results of these studies.

Comparison of the Statistical Content

Currently, the energy balance system established by the International Energy Agency (IEA) and the Organization for Economic Cooperation and Development (OECD) in 1990s is the prevailing model in the world, and the United Nations, the European Union, the Asia-Pacific Economic Cooperation (APEC) are using this model as well. The model categorizes the terminal energy sector into four groups as industrial sector, transport sector, other sectors and non-energy products consumption, and identifies five major transport models: road transport, railway transport, air transport, domestic water transport, pipeline transport. Road transport energy consumption includes all transport fuel used for road transportation activities, but the fuel consumption of transport enterprises in other areas is included in the "Business and public utilities" of other sectors, and the data from surveys is only related with the fuels consumed by the transport activity itself.

For the energy consumption statistical data by industry, National Bureau of Statistics of China sees transportation, storage and postal industry as a whole， including the following areas: railway transport, road transport, urban public transport industry, water transport, air transport, pipeline transport, handling transportation, storage, postal and other transportation services. Among them, road transport industry includes:

(1) The public passenger transport (referring to the road passenger transport outside the city, excluding the city passenger transport, which is included in the urban public transportation).

(2) The road freight transport (excluding Auto Freight Station, Packaged services of road freight transport, which are included in Supporting activities of road transport; to the exclusion of road freight transport agency services, which is included in the transport agency services; operations and management activities of cargo loading and unloading stations, the freight yard, and warehouse are contained in other storage);

(3) The road transport and supporting activities (ie transport service activities related with road).
(4) Passenger Bus Station (to the exclusion of management activities of transfer station for freight cars，which is included in other Road transport and supporting activities; excluding operation and management activities of the city bus station, which is included in urban public transportation).
(5) Road management and maintenance (excluding the management and maintenance activities of city roads, bridges and tunnels, which are included in the municipal public facilities management; to the exclusion of roads, bridges, tunnels, repairs and reconstruction, which are included in the construction of rail , roads, tunnels and bridges engineering).
(6) Other road transport and supporting activities (excluding management and maintenance for the city's road , which is included in the municipal public facilities management; to the exclusion of bus terminal, which is listed separately; exclusive of freight transfer station, which is included in other warehousing).

Table 1. Comparison on statistical indicators energy consumption between some developed countries and current China's.

Country	The United States	The United Kingdom	Japan	China (existing)
Indicators	Number of vehicles Vehicle miles Fuel consumption Person-mile Ton mile	Number of vehicles Vehicle miles Fuel consumption Person-mile Ton mile	Number of vehicles Vehicle miles Fuel consumption Person-km Ton km	Number of vehicles ------- Fuel consumption Volume of passenger transportation freight turnover
Projection index	The average mile per vehicle	The average mile per vehicle	The average km per vehicle	-
	The average miles per gallon	The average miles per gallon	-	Fuel consumption of 100 km
	the average fuel consumption per vehicle	-	-	-
	Energy intensity Btu / person-mile)	-	-	100 tons (in thousands) km fuel consumption

Note: "-" said in the table means there is not that indicator.

Comparison of key statistical indicators

The energy statistics survey indicators of Road transport in developed countries are basically the same. For example, the energy consumption survey indicators of the United States, Britain and Japan are as follows: number of vehicles, vehicle miles (vehicle km), fuel consumption, person-mile (one km) and ton miles (ton-km)，which can be seen in Table 1(Liu Li,Dong.et.al., 2008). The United States has relatively more projection indexes, such as The average mile per vehicle, The average miles per gallon, the average fuel consumption per vehicle and energy intensity.

At present, China's road transport energy consumption statistical indicators are mainly fuel consumption for business-oriented vehicle, fuel (gasoline, diesel) consumption per 100 kilometers for business-oriented bus, fuels (gasoline, diesel oil) consumption per 100 kilometers for commercial freight, fuel (gasoline, diesel) consumption per 100 ton-km. These indicators are all obtained from a sample survey, according to fuel consumption of the average daily in survey period, and then multiplied by the corresponding number of calendar days to calculate the total amount of fuel consumption; 100 tons (in thousands km) fuel consumption can be derived by the total distance fuel consumption dividing fuel consumption for 100 cars-km; fuel consumption of 100 ton-km is equal to the total fuel consumption divided by the total conversion turnover （Sun Qipeng .et.al., 2009）.

Compare the Statistics indicators and survey indicators of the energy consumption between China and foreign countries, some of China's current statistical indicators of road transport can be converted with the developed countries', such as the "average miles per gallon" into "100 vehicle km fuel consumption". However, there are certain differences in some indicators, such as "person miles" in the energy intensity indicators, which is projected by the average number of passengers through surveys, then multiplied by the vehicle miles, while China's "volume of passenger transportation" refers to each passenger actually conveyed multiplying by their corresponding distance, the unit is "person-km". The units of the two indicators are both "person-km", but with a different calculation method at home and abroad. Surely, there is a big difference in results, so the comparability will be affected. In addition, Current statistical indicators in China are not comprehensive enough to reflect key indicators of "vehicle travel mile", that is "vehicle miles".

Vehicle Classifications in the Research

A distinguishing characteristic of vehicle classification in U.S. road transportation about energy consumption is consistent in the classification modes of vehicle about statistical indicators, so a series of average indicators can be calculated through the

survey indicators to analyze the trends of energy consumption changes. The vehicle classification in British for energy consumption indicators can be shown in Table 2. Since the data are from different sources so that the vehicle classification modes in British are various.

Table 2. Vehicle Classification of energy consumption statistical indicators in British road transport.

Statistical indicators	Vehicle Classification						Data sources
Number of vehicles	Private cars and light trucks, motorcycles, public transport vehicles, trucks, licensed vehicles, specialized vehicles, the Royal Vehicle						
Vehicle miles	Cars and taxis, motorcycles, buses and long-distance passenger cars,light trucks, trucks (also divided into seven categories according to axle)				Total		
Volume of fuel consumption	Gasoline	Cars and taxis, motorcycles, light trucks			Total		Trade and industry
	Diesel	Cars and taxis, light trucks, heavy trucks, buses and long-distance buses					
Average miles per gallon	Buses	Gasoline	Vehicle age less than 2 years	Complex	Complex		National travel survey of national statistics bureau
			Vehicle age more than 2 years less than 6 years				
			Vehicle age more than 6 years less than 10 years				
			Vehicle age more than 10 years				
		Diesel	Vehicle age less than 2 years	Complex			
			Vehicle age more than 2 years				
	Trucks	Monomer trucks, hinge trucks					
Ton-mile (total mass> 3.5t)	Operation mode of		Public vehicles				
			Private Vehicles				
	Total Vehicles		Monomer trucks	Total vehicles 3.5t-17t	Total		Road Freight Survey of State Transportation Department
				Total vehicles 17t-25t			
				Total vehicles >25t			
			Hinge trucks	Total vehicles 3.5t-33t			
				Total vehicles >33t			
	Freight types		15 categories of goods				

For example, “average miles per gallon” in buses is classified by the age of vehicles, while the same indicator in trucks is classified by the form of vehicles. “The volume of fuel consumption” is subdivision of fuel types based on “vehicle miles”. Vehicle Classification of various statistical indicators is inconsistent, which directly affects the mutual use and projection among indicators.

Vehicle classification in Japan is quite coarse, with less in vehicle types. Vehicle classification in statistical indicators is basically the same, except “person-miles” in buses, and the other indicators can be basically projected and used each other, which can be seen in Table 3.

At present, the fuel consumption statistical indicators for commercial vehicles set by “Comprehensive Statistical Reporting System for Transportation” in China’s transportation industry can be seen in Table 4. Form the view of statistical indicators designing, it lacks of some data statistics about passenger capacity and tonnage, so it is difficult to calculate person-km and fuel consumption for ton-km.

Table 3. Vehicle classification of energy consumption statistical indicators in road transport of Japan.

Statistical indicators	Buses					Trucks			
	Operating buses			Private buses		Operating trucks		Private trucks	
	Public buses		Other buses	Private cars	Other cars	Ordinary trucks	Light trucks	Ordinary trucks	Light trucks
	Ordinary	franchise							
Number of vehicles									
Vehicle- km									
Ton-km		Public buses	buses	Other buses					
Person-km						trucks			
Fuel consumption		gasoline		diesel		Gasoline		Diesel	

Table 4. Statistical indicators for fuel consumption.

Indicators	Conversion turnover		Fuel consumption		Gasoline consumption for 100 kilometers	Diesel consumption for 100 kilometers
	Gasoline vehicle	Diesel	Gasoline vehicle	Gasoline		
Truck						
passenger						
service						
vehicle						

Table 5. Classification of two different vehicle types.

Classification standard	Categories	
According to vehicle length	Passenger service vehicle	Extra Large, Large, Medium, Small
	Truck	Heavy, Large, Medium, Small
According to radial truck	Passenger service vehicle	Medium And Small Buses, Buses
	Truck	Mini Truck, Light truck, Medium truck, Heavy truck

In addition, different statistics have different classification methods in vehicle type classification, which causes confusion of statistical results and non-uniform of related data to some extent. Such classification shown in Table 5 is carried out by vehicle length and radial truck.

It can be seen from further analysis that the existing classification principles of operating vehicles by load capacity and by energy statistics statistics are basically the same, which can further calculate the total energy consumption. But there are also some problems. First, the number of vehicles is classified in accordance with load capability without subdivision, including fuel types, passenger capacity and tonnage etc. Second, statistical indicators of energy consumption is too macro, with incomplete statistics objects, it is difficult to provide a basis to make further analysis.

By comparative analysis, we found that there are a few differences in road energy consumption statistics between Chinese and the international standards : (1) industry classification, our country sees transport, storage and the postal as the same industry to count, but storage and postal services are excluded with transport in foreign countries; (2) In China, operating vehicles' energy consumption in the transportation sector is only counted in energy consumption , not including vehicles' energy consumption in any other sectors and private vehicles; (3) Some vehicles used in tourism, whose energy consumption isn't contained in energy for transportation when counting, but is in tourism industry; (4) Energy consumption for transport buildings is also contained in energy consumption of transport industry. The second is big in number and involved with a wide range of energy values large, which has a great impact on calculating the proportion of energy consumption in transport, making our energy consumption data much less than international statistical standards in transport. Other differences related to energy values are also small, and they have less impact on the proportion of energy consumption for transportation.

Transport Energy Consumption Ratio Comparison at Home and Abroadfor the Same Aperture

There are certain differences in statistical aperture at home and abroad, so statistical

data does not have comparable relationship strictly and correspondingly. However, in order to further compare the energy consumption level of domestic and international road transport, this article has undergone certain amendments to try to select similar aperture to compare them, mainly comparing from the proportion of different transportation modes in the total energy consumption. After amending, China's energy consumption for road transport, including all modes of transport (except for military vehicles) is as follows:

In 2007, the United States had a proportion of 79.9% for energy consumption of all transport modes on highway in the total transport energy consumption, and was on the decrease from 82.2% in 2003 to 79.9% in 2007. Light vehicles (including cars, light trucks) and motorcycles accounted for 60.4% of the total energy consumption in transport, public buses with 0.70%, and heavy trucks with 18.8%. British had a proportion of 71.9% for energy consumption of all transport modes on highway in the total transport energy consumption, and showed a downward trend from 79.8% in 1990 to 71.9% in 2007. In 2006, Japan had a proportion of 86.9% for energy consumption of all transport modes on highway in the total transport energy consumption, and basically the same level has been maintained since 1990. Private motor vehicles and operating motor vehicles accounted for 54.1% and 1.7% respectively, while bus station and goods vehicles accounted for 1.7% and 29.4% respectively (Jia, et.al., 2009; Mao. et.al., 2009).

However, China's energy consumption ratio for road transport was 73.3% in 2007, which was higher than that of the Britain, but less than that of the United States and Japan. From the trends of total energy consumption in transport sector, the developed countries changed little, while China's growth was quite fast in recent years, and growth speed was much higher than other developed countries'. The total energy consumption of transportation in China was 323.4 million tons of standard coal in 2007, which increased by 85.7% compared with the ratio in 2002, while an increase of 7.7% in United States, 4.2% in Britain and 2.8% in Japan in the same period.

What should be noted is that China's total energy consumption of transport accounted for only 12.7% of energy consumption of the entire community in 2007, and the proportion was far below the United States with 28.6%, Britain with 28.6% and Japan with25.9% (2006). Through analysis, it is related to different countries with different development stages and industry. At present, China's industrial structure is mainly based on industry, which has the largest proportion in energy consumption. Worldwide development shows that the more developed the economy, the lager proportion of energy consumption for transport in the total energy consumption, while China will enter into a rapid development and industrial structure optimization stage in the future, therefore, the proportion of transport

energy consumption is bound to increase rapidly, while energy consumption for road transport which accounts for the largest proportion will be increasing rapidly.

Conclusions

From comparison of energy consumption for road transport at home and board, we conclude as follows: First, China's statistical indicators for transport energy consumption have the following differences with the international norms: (1)In industry classification, our country sees transport, storage and the postal as the same industry to count, but storage and postal services are excluded in foreign countries; (2) In China, operating vehicles' energy consumption in the transportation sector is only counted in energy consumption , not including vehicles' energy consumption in any other sectors and private vehicles; (3) Some vehicles used in tourism, whose energy consumption isn't contained in energy for transportation, but is in tourism industry; (4) Energy consumption for transport building is also contained in energy consumption of transport industry. Second, compared with developed countries, energy consumption for road transport has the largest proportion in energy consumption for transportation industry, while China's proportion of energy consumption for road transport accounting for the entire society is still far lower than that of developed countries. However, the proportion of transport energy consumption is bound to increase rapidly with China's economic development and industrial structure optimization in the future.

References:

Jia Shunping, Peng Hongqin, Liu Shuang, and Zhang Xiaojie. (2009). "Review of Transportation and Energy Consumption Related Research." *Journal of Transportation Systems Engineering and Information Technology*, (9)3,6-16.

Li Liancheng, and Wu Wenhua. (2008). "Energy Utilization Efficiency and Development Tendency of China's Transportation." *Comprehensive Transportation*, (3), 16-20.

Li Xiansheng,Wang Yunlong, and Cai Fengtian,et al. (2007). "Contrast Research on Fuel Consumption Statistics Model for Road Transportation." *Computer and Communications*, 25(6), 49-55

Liu Li,Dong Guoliang, and Wang Yunlong,et al. (2008). "Research on Energy-consumption Statistical Indicators of Highway Transportation." *Transportation Energy-saving and Environmental Protection*, (3),16-22.

Mao Baohua, Jia Shunping, and Sunqipeng,et al. (2009). "Energy Consumption,Emissions and their Comparison among Different Transport Modes." *Energy Foundation*, Integrated Transport Research Center of China Beijing Jiaotong University, (12), 99-102.

Peng Jiahui, Lin Xueshan, and Jiang Han. (2008). "Calculation and Analysis on Terminal Energy Consumption of Chongqing." *Construction Science and Technology*, (19), 109-111.

Shi Jingya, Su Yongqing, and Yue Jiguang. (2008). "Analysis of the Influencing Factors of Energy Consumption for Rail Transit and Establishment of Energy Consumption Evaluation System." *Railway Transport and Economy*, 31 (2), 12-15.

Song Minhua. (2009). "Research on Development Tendency of Urban Rail Transit Energy-saving Technologies." *Construction & Design for Project*, (1), 15-19.

Sun Qipeng ,Wang Qingyun, and Mao Baohua. (2009). "Framework Design of Different Transportation Modes'Energy Consumption Factors and Comparabilities Study." *Journal of Transportation Systems Engineering and Information Technology*, (9)4, 10-14.

Comparison Analysis of Idle Emissions Measurement Program in JABODETABEK Indonesia

S.B. Nugroho, A. Fujiwara and Junyi Zhang

Transportation Engineering Laboratory, Graduate School for International Development and Cooperation, Hiroshima University Japan

Abstract

This paper describes an initial analysis of vehicle inspections and maintenance program for private cars in Jakarta and surrounding cities namely JABODETABEK, as regulated in the 2005 bylaw. The 2005 bylaw on air pollution control, which took effect on February 2006, stipulated that all private car owners must get their vehicles' emission tested twice a year based on in-use vehicle emission standard. We apply bivariate probit model for the likelihood of CO and HC emission violations given a set of vehicle characteristics. Data of passenger car emission measured randomly during on-road emission measurements at Jakarta city and surrounding cities Bogor, Depok, Tangerang and Bekasi in 1999, 2000, 2004 and 2005. The engine size, fuel system, vehicle maintenance quality and passenger car travelled kilometer per year play a significant role in determining the probability of emission test failure. Vehicle kilometer travelled per year is different based on the vehicle-type and its usage. To analyze household vehicle-type ownership in Jakarta city, we use person trip data in 2002 and apply Multinomial Logit Model. Household income, family sizes are play significant in determining vehicle-type owned by households. Vehicle type Jeep and SUV are less to use by the households compare to Sedan or passenger car. In order to pass CO emission test, empirical study confirm that a combination of lower use of high fuel consumption car is better than higher use of low fuel consumption car.

Key Words: Bivariate Probit; vehicle-type; dimensionless analysis; multinomial logit

Introduction

A study by Japan International Cooperation Agency (JICA) and Environmental Impact Management Agency (Japan International Cooperation Agency, 1997)

revealed that more than 50% of Carbon Monoxide (CO) was emitted from private motor vehicles and approximately 20 percent from motorcycles. Private motor vehicles and motorcycles are responsible for approximately 40 percent of Hydrocarbon (HC) emissions(SEI, UNEP, KEI, 2002). In 2005, the Local Government of Jakarta issued The 2005 Jakarta's bylaw on air pollution control, which took effect on February 2006, mentioning about air pollution control from mobile sources. It was stipulated that all private car owners must get their vehicles' emission tested biennially (Adianto P. Simamora, 2006). This testing process is commonly referred to as an inspection and maintenance (I/M) program. In order to pass the test, the vehicles emission should be lower than the in-use vehicle standard. The vehicle emission test is performed by the authorized automobile workshops and technicians. If a vehicle passes the emission test, the owners will be given a certificate and a sticker to be attached on the upper left side of the windshield. In contrast, if a vehicle fails in the emission test, it should be repaired or perform an appropriate maintenance to achieve the allowable emission standard. The bylaw No 2 Year 2005 states that the emissions certificate will be required for extending the vehicle's registration (Badan Pengelolaan Lingkungan Hidup Daerah, 2005).

There have been ongoing debates on the effectiveness of the implementation of this program to reduce the mobile emissions. In several studies by Hubbard, T. (1997), Wasburn, et. al. (2001), and O, Bin (2003), the I/M programs have been heavily criticized in their three main points since their inceptions. First, it has been argued that I/M programs are an inefficient use of resources to achieve air quality objectives. It is also inconvenient to the vast majority of the driving population. Second, the I/M programs are not the most effective way to identify gross polluting vehicles. The I/M test procedure for Jakarta city has been based on the idle-mode test, so it does not account for the real world driving conditions such as acceleration and deceleration cycles, and thus vehicles passing the emissions test may still be gross polluters in the real world driving conditions. Third, the programs have failed to provide drivers with incentives to minimize their vehicle emissions (Wasburn, et. al., 2001).Additionally, many drivers tamper with their engine and emission control in order to past the emission test and the repair of high emitting vehicles is waived because of arbitrary cost. It is not difficult to 'cheat on' in the idle-mode emission test. Implementation of in-use passenger car emission standards and also passenger car emission testing does not usually result in additional direct costs for governments. Usually, implementation costs transferred to passenger car owners. In case of Jakarta city, the local government endorse the private passenger car mechanic workshops to perform the I/M program. The private passenger car owner can do the emission test and passenger car maintenance in order to meet the allowable standard.

In this study, first, we attempt to analyze the existing condition of in-use vehicle emissions based on the observed data measured by an onsite idle emission testing at several major roads in Jakarta city and surrounding cities (Bogor, Depok, Tangerang, and Bekasi) which are popular as the JABODETABEK area in 1999, 2000, 2004 and 2005. Road traffic flow in Jakarta city is characterized very frequently during the day by elevated congestion level. This implies a great number of vehicles operating in

idle and stop-and-go driving mode during a significant period. It is known that vehicle exhaust emission of NOx, CO, Hydrocarbons and particles are different for these two operating modes, being highest during acceleration (K.S Chen.et.al, 2003). Nevertheless, the contribution of emission during idle mode can be similar if idling time is prolonged like what occurs in congested traffic. By using the data collected from the sampled vehicles drawn randomly at 4 different time points, we can compare the actual emission conditions of in-use passenger car with the Indonesian's standards. The Bivariate Probit analysis is conducted for the likelihood of carbon monoxide and hydrocarbon emission violations given a set of passenger car characteristics. This methodology finds the effects of characteristics such as carburetor/injection system, engine size, passenger car travelled kilometer per year and dummy variable of passenger car maintenance which determine by air-fuel ratio on the likelihood of emissions test failure. Furthermore, we analyze the influence of social-economic factor, residence location and family type on vehicle-type ownership in Jabodetabek. Using person trip data in 2002, Multinomial Logit Model apply to examine household characteristics on vehicle-type owneship in JABODETABEK area. We assume the difference probability to pass the emission test of observed passenger car have been influenced by vehicle-type and its usage. Output of this study can also be used as significant input to the policy makers to evaluate vehicle-type ownership and strategies to reduce emission from vehicles.

Methodology

Idle Emission Measurement of In-use Vehicles. To perform idle-emissions tests, we refer to Indonesian Standard SNI 09-3678-1995 which was already revised to be SNI 19-7118.3.2005. These Indonesian standards developed based on International Organization for Standardization (ISO) 3930/OIML R99-instrument for measuring vehicle exhaust emission 2000- and United Nation for Economic Commission for Europe (UN-ECE). To prepare idle-emission test, passenger car exhaust pipe shouldn't have any leakage, under normal temperature of engine and the ambient temperature in between 20^{o}C-35^{o}C. Idle conditions means the passenger car engine working without any acceleration of fuel system, neutral transmission position for manual type of passenger car, neutral transmission or parking position for automatic passenger car. At the same time, other passenger car accessories which influence to engine rotation were shut down. Vehicles exhaust gas measured by a gas analyzer to obtain the concentration or emissions levels of CO (%) and HC (ppm).

As well, emissions level is also influenced by the driving cycles. Many factors may influence on-road driving cycle and levels of passenger car exhaust pollutants. Driving cycles of urban areas differed significantly from rural areas (K.S Chen.et.al, 2003). Urban cycles has shorter travel distance, lower travel speed, more idle time, and lower acceleration/deceleration time than rural ones. Urban areas generally have more congested traffic, intersections and traffic signal than do rural areas. Accordingly, urban driving cycles generally consumes more fuel (approx. 30% more per km) than does rural driving but the emission levels in urban driving differ insignificantly from those in rural driving (K.S Chen.et.al, 2003).

Road traffic in big cities in Indonesia is characterized with high congestion levels during day times. Thus, there is great number of vehicles operating at idle or stop-and-go driving conditions. It is known that vehicle exhaust emissions of NOx, CO, HC and particles are different at driving conditions, being the highest during acceleration (M. Pujadas,.et.al., 2004). Nevertheless, we can assume that emission test results can be used to represent the real world conditions at sites over-congested traffic that result in long idling times. A vehicle may have over 25% of its time spent in the idle mode (Tong, H.Y. et .al., 1999).

Bivariate Probit Model for In-Use Vehicles Exhausts Emissions. The bivariate probit regression analysis used to examine the likelihood of CO and HC emission violations. In this study we refer to our national standard in state Ministry of Environment Decree No 13 Year 1993 which was already revised in 2006, the maximum allowable idle emissions limit value of CO is 4.5% volume and maximum HC value is 1200 ppm. Other secondary data contain information on the various passenger car characteristics of tested passenger car, such as vehicle registration number, manufacturer, make and model as well as model year, engine fuel system, odometer reading, lambda, and passenger car types. In the model, emission test violation is defined as the effect of passenger car characteristics such as engine size, engine fuel system and running kilometer of passenger car. Other dummy variable of maintenance quality was determined by air-fuel mixture ratio (AFR) and its deviation to the normal standard value of AFR which assume represent the car maintenance quality. Using all independent variables, we propose a Bivariate binary Probit regression model of the emission test failure. Bivariate binary Probit regression model depends on simultaneous observation of two discrete binary observed-dependent variables, i.e., y_{i1} *and* $y_{i2,}$ that indicates the emissions test failures of CO and HC. Based on the observed dependent variables that take binary discrete values, underlying continuous dependent variables, z_{i1} and z_{i2}, can be expressed as:

$$\begin{aligned} z_{i1} &= \boldsymbol{\beta}_1' \mathbf{x}_{i1} + \varepsilon_{i1} \\ z_{i2} &= \boldsymbol{\beta}_2' \mathbf{x}_{i2} + \varepsilon_{i2} \end{aligned}, \; y_{ij} = 1 \text{ if } z_{ij} > 0, \; y_{ij} = 0 \text{ otherwise}, \; j = \{1, 2\} \qquad (1)$$

where i denotes an observation; $\boldsymbol{\beta}$ and $\mathbf{x}$ stand for the vectors of parameters and the independent variables respectively; ε_{i1} and ε_{i2} are random variates distributed jointly as standard Bivariate Normal and a free correlation parameter, , i.e., BNV [0,0,1,1,]. Based on the equation given above, the log- likelihood function of the sample can be given as:

$$\log L = \sum_i \log \Phi_2 \left[q_{i1} \boldsymbol{\beta}_1' \mathbf{x}_{i1}, q_{i2} \boldsymbol{\beta}_2' \mathbf{x}_{i2}, q_{i1} q_{i2} \rho \right] \qquad (2)$$

where Φ_2 stands for the standard Bivariate Normal distribution; q is an indicator variable such that $q_{im} = 2y_{im}-1$, $m = \{1, 2\}$. Based on data collected by the on-road measurement at Jakarta, in 1999, 2000, 2004, and 2005 the model is estimated by using LIMDEP Version 8.0 econometric software (Greene, W., 2002).

Table 1. Variables Definition in the Study.

No	Indicator	Driver
A	Emission Indicator	
	1. Carbon Monoxide (CO)	Carbon monoxide measured by the percent of total volume of emission gas. Compare to National Standard (1 if passed the emission test, other=0)
	2. Hydrocarbon (HC)	Hydrocarbon measured by parts per million. Compare to National Standard (1 if passed the emission test, other=0).
B	Private Vehicle Characteristic	
	1. CARB	Carburetor Cars=1, Other=0
	2. RKTYRCLS	Private Passenger Car Running Kilometer per year. Calculate by observed odometer reading and vehicle age. Divided into six class: 1<10000 km/yr; 10001<2<20000 km/yr; 20001<3<30000 km/yr; 30001<4<40000 km/yr; 40001<5<50000 km/yr; 6>50001 km/yr
	3. ENGSZCLS	Engine Size Class (cc). 1 < 1000 cc; 1001<2<1500 cc; 1501<3<2000 cc; 4>2001 cc
	4. IMCODE	Dummy Variable Maintenance Quality. Calculate by using measured Air to Fuel Ratio. Very good ($0.95 \leq \lambda < 1.05$); Good ($0.9 \leq \lambda < 0.95$ or $1.05 \leq \lambda < 1.1$); Moderate ($0.85 \leq \lambda < 0.9$ or $1.1 \leq \lambda < 1.15$); Bad ($0.8 \leq \lambda < 0.85$ or $1.15 \leq \lambda < 1.2$); and Very bad ($\lambda < 0.8$ or $1.2 \leq \lambda$)
C	Attributes of Car owner	
	1. Socio-Economic	Household Income (1) Low= less than 1 million rupiah (100 US$, 1US$=Rp 10000); (2) Medium to low (1 million < income < 3 millions); (3) Medium to high (3 < income < 5 millions); (4) High (5 < income < 7.5 millions); (5) Very high (> 7.5 millions)
	2. Family Type	Family type: (1) Single; (2) Couple; (3) Small family: couple with children (max 2), (4) Big family: Household members more than 4 persons.
	3. Residence Location	Housing location (1) CBD: city center of Jakarta, (2) Urban non-cbd: non city center of Jakarta city ; (3) sub-urban: Bogor, Depok, Tangerang and Bekasi, namely BODETABEK.
	4. Period of residence	(1) Short: less than 1 year; (2) Medium= 1 to 5 years; (3) Long= 5 to 10 Years; (4) Very long= more than 10 years
	5. Vehicle-type	(1) Sedan/Passenger Car; (2) Sport Utility Van (SUV),Van and Jeep; (3) Pickup and Light truck

Passenger car running kilometer per year calculated by divide the actual odometer reading for each samples and vehicle age. According to vehicle model year and passenger car running kilometer per year samples distribution, we classify all data into six groups (Table 1). Based on the engine fuel system, we categorized two groups of carburetor and injection cars. For dummy variable of maintenance quality, we use air-to-fuel ratio measured from each samples and classified them into five categories: very good, good, moderate, bad and very bad based on their deviation if compare to the reference value (one).

Vehicle-Type Ownership.Vehicle ownership by household is a critical demographic characteristic influencing many aspects of travel demand and its impacts. It is well recognized that different vehicle types impose different emissions intensity. Vehicle-type ownership is a critical variable in policy analysis regarding on the Inspection and Maintenance Program in Jakarta city to reduce emissions from mobile sources. A recent study expects sharp increase in passenger car ownership level when per capita income level reaches a level between US$ 3000 and US$ 5000 (Dargay, J, 1999).

Due to rapid development and economic growth of developing Indonesian country, household income level increase gradually. GDP per capita of Jakarta city, the capital of Indonesia, reaches US$ 4992 in 2006 (Siadari, M., 2007) followed by Surabaya city, second largest city in Indonesia, which GDP per capita around US$ 3481 in the same period. Other study, Ingram and Liu (1999) estimate that passenger car ownership in developing countries is expected determinism-a sociological explanation-which associates car ownership in developing countries exclusively with the middle class life styles, and stresses the social forces on the middle class to sustain a mobility level tied to car ownership (Vanconcellos, E.A.1997). Indonesian automaker association (Gaikindo) reported that the annual vehicles sales in 2008 reached 600 thousand unit which almost 30-40% sold in Jabodetabek area(Domestic Market and Export-Import, 2004-2008).

Car ownership model found in the academic literature are classified into a number of types depend on a number of criteria and variety of purpose (de Jong, G.et al.,2004). The majority of models are developed in a discrete choice framework. Automobile ownership studies have typically been modeled by using well established Multinomial Model (MNL) (Mohammadian, A., Miller and E.J, 2003). These models identify two major components of automobile ownership: the number of vehicles owned and vehicle type. Interest in vehicle ownership modeling appears to have slowed, perhaps due to many modeling techniques have matured and improvements that can be made to existing techniques are generally marginal (Zhao, Y.et.al, 2001). Recent study in Hamilton, Canada(Potoglou, D. and P.S. Kanaroglou.,2008) found that family structure, socio-economic characteristics and accessibility at the residential location influence households' decision on the number of cars to own. By using both explanatory analysis and discrete choice modeling, the authors also reported the empirical findings on the relationship between vehicle-type choice and neighborhood characteristic in the same city, Hamilton-Canada (Potoglou, D.,2008). Another study by Bhat *(*2006), a multiple-discrete continuous extreme value model

applied to examines the important effects of household demographics, residence location and vehicle attributes on household multiple vehicle-type holding and use. Senbil (2007) use bivariate ordered probit model to analyzed household car and motorcycle ownership which are independent of each other in JABODETABEK.

In this study, we focus on vehicle-type holding by household single car owner and its indirect cause-effect relationship with emission test performance under I/M program in Jakarta city. Looking at the quantity, this group of households holding one car without having motorcycle are 6.3% of total household samples (> 100000 households) in JABODETABEK. We explore the influence of household characteristics shows in table 1 such as income level, family type, housing type and residence location on vehicle-type (car/passenger car, SUV/Jeep, pickup/light truck) holding by households in Jabodetabek. A discrete choice model, Multinomial Logit Model (MNL) used to examine the relationships between households characteristic and the vehicle-type own. Vehicle-type *j* holding by a single car household owner (*Yi*) is classified as unordered response. In an unordered model, the polytomous response variable does not have an ordered structure. The generalized logit model focuses on the household as the unit of analysis and uses household characteristics (*Xi*) as explanatory variables. The explanatory variables, being characteristics of households (*Xi*) are constant over the alternatives. Based on data collected by JICA person trip survey in 2002 the probability of vehicle-type *j* owned by household *i* is estimated by using econometric software LIMDEP Version 8.0(Greene, W., 2002).

Data

Vehicle Populations in JABODETABEK. According to The Jakarta Metropolitan Area (Metrojaya) State Police, data reveals that most of motor vehicle fleet in Jabodetabek was motorcycles around 58,9 % followed by Passenger Car around 28.7% respectively. In 2002, total motor vehicle population in Jakarta area was approximately 5 million unit or 20.3% of total motor vehicle population in Indonesia (Nugroho, S.B.et.al., 2004) Based on annual motor vehicle sales data reported by Gaikindo (Indonesian Association of Automobile) from 1976 to 2003, the highest vehicle sales accounted in 1997 around 397 thousand units per year. Due to the economic crisis, vehicle sales from 1998 drop significantly. However, the vehicles sales recently have rebound with the total annual sales were up to 354 thousand units in 2003. During the period from 2000 to 2003, the percentage of vehicle sales in Jabotabek was around 34.3 – 37.4% of total vehicle sales in Indonesia. Most of vehicles sold during period of year 2000 to 2003 were gasoline vehicles and around 92% of total sales were accounted for passenger cars and the remaining were for both trucks or and buses. Based on person trip survey by JICA in 2002, 47.5% samples (1,08 Millions samples) households in JABODETABEK have vehicles in their house. It is consists of motorcycle holders 33%, car holders 8 % and multiple car and motorcycle holders 6.3% (table 2). In this study, we focus on the group of single car holders.

Table 2. Vehicle Ownership in JABODETABEK.

No	Households car and motorcycle ownership	Percents
A	Households without car and mc (no Car and no MC)	**52.53**
B	Motorcycle Owner (Motorcycle only)	
	1. One motorcycle	30.10
	2. Two motorcycles	2.71
	3. Three or more motorcycles	0.44
C	Car owner (car only)	
	1. One car	6.31
	2. Two cars	1.23
	3. More than 2 cars	0.31
D	One-vehicle owner and motorcycle owner	
	1. One car and one motorcycle	4.52
	2. One car and two motorcycles	0.68
	3. One car and 3 or more motorcycles	0.15
E	Two-vehicles owner and motorcycle owner	
	1. Two cars and 1 motorcycle	0.59
	2. Two cars and 2 motorcycles	0.15
	3. Two cars and more than 2 motorcycles	0.04
F	More than two vehicles owner and motorcycle owner	
	1. More than 2 cars and 1 motorcycle	0.18
	2. More than 2 cars and 2 motorcycles	0.04
	3. More than 2 cars and more than 2 motorcycles	0.01
	Total	**100**

Source: Person trip survey, 2002.

On-Road Emission Measurement for In-Use Vehicles.Idle Emission testing was held in several major roads in Jabodetabek (Figure1). In this study, sample sizes obtained from 4 different times periods were 2784 data sets, which distributed as 230(8.3%) samples measured at 1999, 1386 (49.8%) samples from year 2000, 727 (26.1%) samples taken at 2004, and 441 (15.8%) samples from year 2005. Nine locations were selected in 2004 as onsite vehicle measurement spots (check points) which consist of 5 locations in Jakarta city and four sampling points which located in Bogor, Depok, Tanegrang and Bekasi (BODETABEK) (figure 1). A brief face-to-face interview was performed simultaneously during onsite idle emission measurement.

Thirty percent of samples were failed emission test failure which mean their emission level of Carbon monoxide nor hydrocarbon were exceed allowable limit which stated in our national emission standard. Car with engine size capacity in between 1500 and 2000 was most famous (51.5 %) among car holder in JABODETABEK. The annual personal mobility car in Jakarta city is less than 10000 km per year (44.5%) followed by the mobility in between 10000 km and 20000 km per year around 33.8%. Total number of passenger car owner or user in Indonesia

who perform good car maintenance is around 33.6 %. Detail descriptive statistics of samples passenger car in this paper are shown in Table 3.

The Bivariate Probit regression analysis is conducted using independent variables of engine size class, running kilometer per year of passenger car, engine fuel system (Carburetor), dummy variable maintenance quality, and vehicle-type. In order to ascertain the influence of vehicles characteristics (engine size, running kilometer per year, carburetor, lambda) and the impact of households vehicle-type ownerships on passenger car emissions (i.e., CO and HC), two alternative specifications are estimated to test for the significance of inclusion of the factors. The first model includes the passenger car characteristics. By applied the model, we obtained the variables which determined significantly the emission test results and these variables differed for each vehicle-type. Then, we compare the probability of emission test among vehicle-types.

Table 3 Means and standard deviations of vehicle characteristics.

Variable	Definitions	Total	Emission Test Failure		
Name			Fail_Test	Fail_CO	Fail_HC
	Samples (N)	2784	836	801	68
CO	Carbon monoxide measured by the percent of total volume of emission gas	3.584 (13.687)	7.880 (24.329)	8.147 (24.831)	4.649 (3.303)
HC	Hydrocarbon measured by parts per million	400.563 (394.761)	641.202 (596.229)	590.575 (544.835)	1939.580 (1396.175)
ENGSCL	Engine Size classification by cubic centimeter displacement				
	(1) Engine Size ≤ 1000 cc	5.24%	43.15%	40.41%	4.11%
	(2) 1000< Engine Size ≤ 1500 cc	29.71%	34.70%	32.65%	3.75%
	(3) 1500< Engine Size ≤2000 cc	51.47%	30.50%	29.73%	1.61%
	(4) 2000 cc < Engine Size	13.58%	12.96%	12.17%	2.12%
RKTYR	Running Kilometer per year				
	(1) RKTYR ≤ 10000 km/yr	44.50%	32.53%	30.67%	3.39%
	(2) 10000< RKTYR ≤ 20000	33.80%	29.01%	28.06%	1.28%
	(3) 20000< RKTYR ≤ 30000	14.55%	25.93%	24.69%	2.47%
	(4) 30000< RKTYR ≤ 40000	3.05%	34.12%	32.94%	1.18%
	(5) 40000< RKTYR ≤ 50000	1.22%	26.47%	26.47%	2.94%
	(6) 50000< RKTYR	2.87%	21.25%	20.00%	2.50%
I/M Quality	Air to fuel ratio(Lambda)				
Very Good	$0.95 \leq \lambda < 1.05$	33.58%	7.38%	6.31%	1.50%
Good	$0.9 \leq \lambda < 0.95$ & $1.05 \leq \lambda < 1.1$	22.23%	11.47%	10.34%	1.45%
Moderate	$0.85 \leq \lambda < 0.9$ & $1.1 \leq \lambda < 1.15$	17.06%	49.05%	48.63%	2.11%
Bad	$0.8 \leq \lambda < 0.85$ & $1.15 \leq \lambda < 1.2$	12.86%	71.23%	70.39%	3.07%
Very bad	$\lambda < 0.8$ & $1.2 \leq \lambda$	14.26%	52.39%	49.12%	6.05%

Notes: in each cell, the first row indicates the mean and the second row (inside the brackets) indicates the standard deviations.

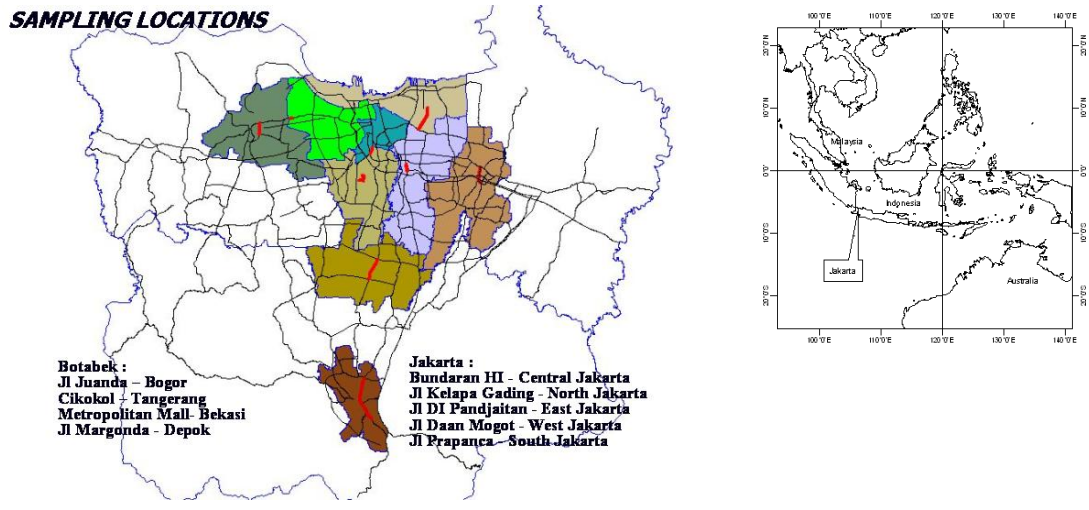


Figure 1. Locations of on-site emission measurement in 2004.

Model Estimations and Results Discussion

Table 4 report the estimation results of the regression analyses the influence of vehicle characteristics. Looking at table 4, running Kilometer per year and engine size plays significant role in determining emission test failure for both CO and HC at the estimation results for 2000, 2004 and 2005. All coefficients estimated for annual VKT and engine size were positively significant in determining CO and HC emission test performance in the same three years. In contrast, Engine fuel system (Carburetor) only for play significant role in determining the probability of CO observed in 1999 and 2004. Carburetor system of car is positively play significant role in determining CO in 1999 but it was negative in 2004. Vehicle maintenance quality always gives significant impact negatively on CO exhaust performance. In contrast, it was significant for HC observed in 2000. Basically, vehicle engine size (capacity) is directly related to emission and energy consumption. Vehicle with larger engine consume more fuel and emit more pollutions. There is often however a large difference between the most and least efficient vehicle within each engine size range (D. Stead, 1999). Also, a large difference of engine pollution control device between engine size classes. In order to comply with the regulation, small size engine cars are worse compare to medium and large engine size cars. Observing all estimation results, dummy variable of maintenance quality keep remain stable as one of significant factor at 1% level which influence to CO emission test failure. One key parameter very effective in the formation of CO and HC is the air-to-fuel ratio when the fuel is burned (Rubin, E.S., 2001*)*. In this study, air–to-fuel ratio (lambda) is taken as the *stoichiometric amount*. When air–to-fuel ratio is less than 1, it is impossible to burn all fuel as there is insufficient air. In this case, CO and HC should

be emitted as products of incomplete combustion. Generally, air–to-fuel ratio is subject to passenger car maintenance conditions. Routine maintenance keeps air-to-fuel ratio near the *stoichiometric amount*, which automatically sustains the passenger car emissions at the desired levels. It was justified from observed samples that reducing the maintenance quality consequently increase the probability of emission test failure (Table 4).

Vehicle running kilometer per year tends to decrease gradually based on their vehicles age (Figure 2). Observed from samples, the highest average individual car mobility per year observed for passenger car/sedan type (17723 km/year) and the lowest one observed for SUV/Jeep (14579 km/year). Passenger car annual running kilometer per year slightly decrease compare to other two types of vehicles. Considering annual vehicle kilometer travelled by passenger car which most populated among others, average running kilometer per year for new vehicles (1-3 years) is almost similar to the average of passenger car/sedan annual running kilometer per year.

Table 4 The bivariate probit estimation for vehicle characteristics

Variables	1999		2000		2004		2005	
	Beta	t-score	Beta	t-score	Beta	t-score	Beta	t-score
Parameter CO								
Annual Running Kilometer	0.0565	0.108	0.0862	2.409	0.1137	2.347	0.3052	5.035
Engine Size Capacity	0.3298	0.6	0.5134	14.951	0.4845	9.855	0.6447	8.419
Carburettor	0.0245	3.536			-0.2088	-1.97	-0.1207	-0.538
Maintenance Quality	-0.1683	-3.154	-0.3995	-16.197	-0.3292	-10.228	-0.5289	-10.424
Parameter HC								
Annual Running Kilometer	-0.0642	1.029	0.166	2.502	0.326	3.394	0.4382	3.617
Engine Size Capacity	0.6431	-0.396	0.5662	6.781	0.6596	4.982	0.5163	4.723
Carburettor	0.5457	3.228			0.0421	0.101	-0.2225	-0.705
Maintenance Quality	-0.0843	-1.015	0.0939	1.655	-0.0802	-1.077	-0.1115	-1.193
Disturbance correlation (rho)	0.3818	1.717	0.0466	0.447	0.388	3.17	0.2417	1.723
Log likelihood function	-166.0899		-818.917		-470.1381		-228.7673	
Sample size	230		1386		727		441	

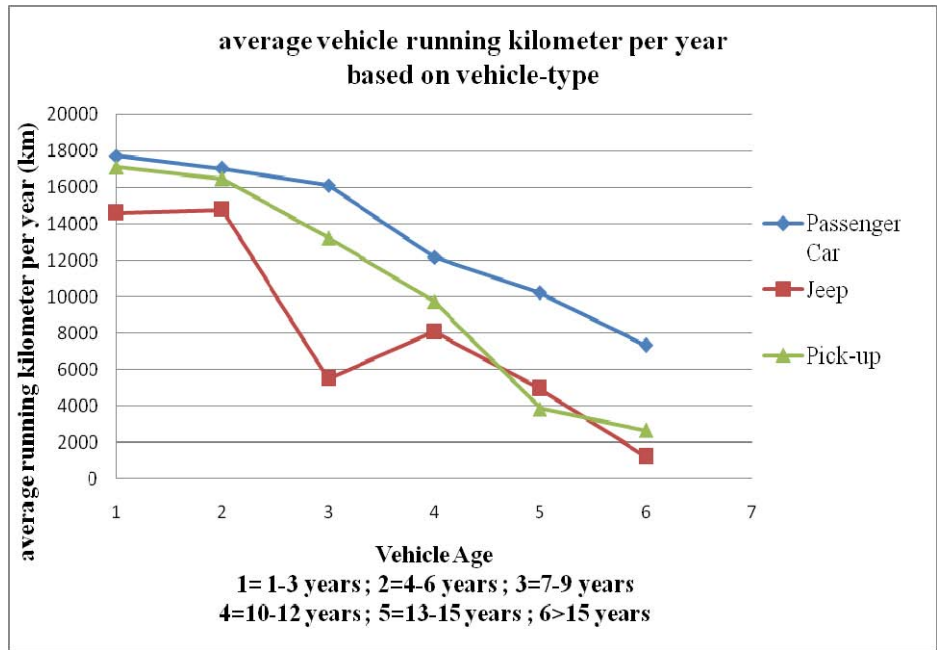


Figure 2 Average Running Kilometers Per Year Of Vehicles In JABODETABEK.

Data distribution of measured vehicles consists of Sedan/Passenger car type around 74.9 % followed by Pickup/Light Truck 15.7 % and Jeep/SUV around 9.4 % respectively. Based on person trip survey data about single-car owner households in JABODETABEK in 2002, 86.3% samples have sedan/passenger car, 7.9% own sport utility vehicle (SUV)/Jeep and 5.8 % has Pickup/Light truck. Looking inside the owner characteristics influence on vehicle-type based on Multinomial Logit Model, family structure always negative significant for all vehicle-types (Table 5). Household income also negative and significant, but in some case are insignificant. In case of Sedan's ownership, very high income group which have more than 7.5 million rupiah (equals to 750 US$/month, 1US$=Rp 10.000) is insignificant. In contrast, low income level became insignificant for SUV-VAN-Jeep vehicle type and medium to low income level also insignificant for Pickup-Light truck vehicle type.

About the residential period, it was only significant for Sedan type. Interesting phenomena was found for residential location consists of urban CBD, urban non-CBD and Sub-urban areas. The coefficient estimated always negative for ownership of sedan in all locations. In contrast, the sign were positive for other two vehicle types. Sport utility vehicle-van-jeep is more to be owned by household live in sub-urban rather than households live in CBD and urban non-CBD areas.

By using data from person trip survey in 2002, more than 68% of Sedan/Passenger car owner use their car for the daily primary transport mode, 20 % used bus and only 4.3% use motorcycle. Tend to use as daily primary transport mode decrease from 53.7% for SUV/Jeep owner and 39.8% for Pickup or light truck owner. In contrast, tend to use bus a daily primary transport mode increase from 30.3% for Sedan/Passenger car owner to 39% for SUV/Jeep and 40 % for pickup/light truck

owner (Figure 3). About daily travel time, average travel time of sedan owner's around 42.2 minutes followed by SUV/Jeep owner around 35.8 minutes and 33.3 minutes for Pickup/light truck owners.

Table 5 Multinomial Logit Model of Vehicle-Type Ownership

Household Characteristics	Sedan/Passenger Car		SUV,JEEP		PICKUP-LIGHTRUCK	
	Beta	t-score	Beta	t-score	Beta	t-score
Residence Location						
a. CBD	-0.314	-3.24	-	-	-	-
b. Urban Non-CBD	-0.406	-3.894	-	-	0.539	3.13
c. Sub-Urban	-0.921	-8.644	0.956	6.563	0.764	4.357
Family Type						
a. Single	-6.556	-13.671	-3.318	-3.672	-3.782	-3.972
b. Couple	-6.593	-14.18	-3.568	-3.989	-3.509	-3.757
c. Small Family	-6.749	-14.548	-3.324	-3.724	-3.343	-3.59
d. Big Family	-7.014	-15.119	-3.219	-3.606	-2.94	-3.156
Household Income						
a. Low Income	-2.45	-20.607	-	-	1.584	9.746
b. Medium to Low Income	-1.222	-10.383	-0.549	-4.829	-	-
c. Medium to High Income	-0.482	-4.042	-1.005	-8.535	-1.176	-6.862
d. High Income	-0.369	-2.84	-1.219	-8.926	-1.124	-5.568
e. Veryhigh Income	-	-	-1.762	-10.351	-1.779	-6.654
Residence Period						
a. Short	-2.808	-7.122	-	-	-	-
b. Medium	-2.306	-5.835	-	-	-	-
c. Long	-2.614	-6.641	-	-	-	-
d. Very long	-2.749	-6.988	-	-	-	-
Constanta	13.283	16.636	-	-	-	-
Log likelihood function	-25359.3		-18043.98		-13689.04	
Sample size	68365		68365		68365	

All written in 1 % significant level

In fact, we have two empirical situations which combine vehicle-type and its usage. First, sedan/passenger car which assume less in fuel consumption compare to other type (SUV/Jeep and Pickup/Light truck) but use it more intense and longer. On the other hand, SUV/Jeep and Pickup/light truck assume consumes more fuel but it less to use and shorten. To evaluate the influence of this empirical situation on the emission test performance, we use and modify our previous bivariate model to incorporate vehicle-types. Sedan/Passenger car's data used as the reference to

compare the probability of emission test result with other vehicle-type. We found that probability of pickup/light truck is differ significantly compare to Sedan/Passenger car (Table 6) especially in order to pass CO emission test observed in 1999, 2000 and 2004. The sign always positive and statistically significant which show that the probability of Pickup or light truck to pass emission test is higher compare to Sedan/passenger car type. As for SUV/Jeep, the sign was not consistent, which is positive at 2000 and 2004 but it was negative in 2005 and also significant in 1999. We could preliminary conclude that the probability of CO emission test failure much more depending on the vehicles usage rather than internal vehicles characteristics (fuel consumption and device designated for emission control). It is necessary to explore more detail in future about individual characteristics (fuel consumption, emission control device in car, car mobility) and its impact on vehicle emission test performance based on vehicle-type.

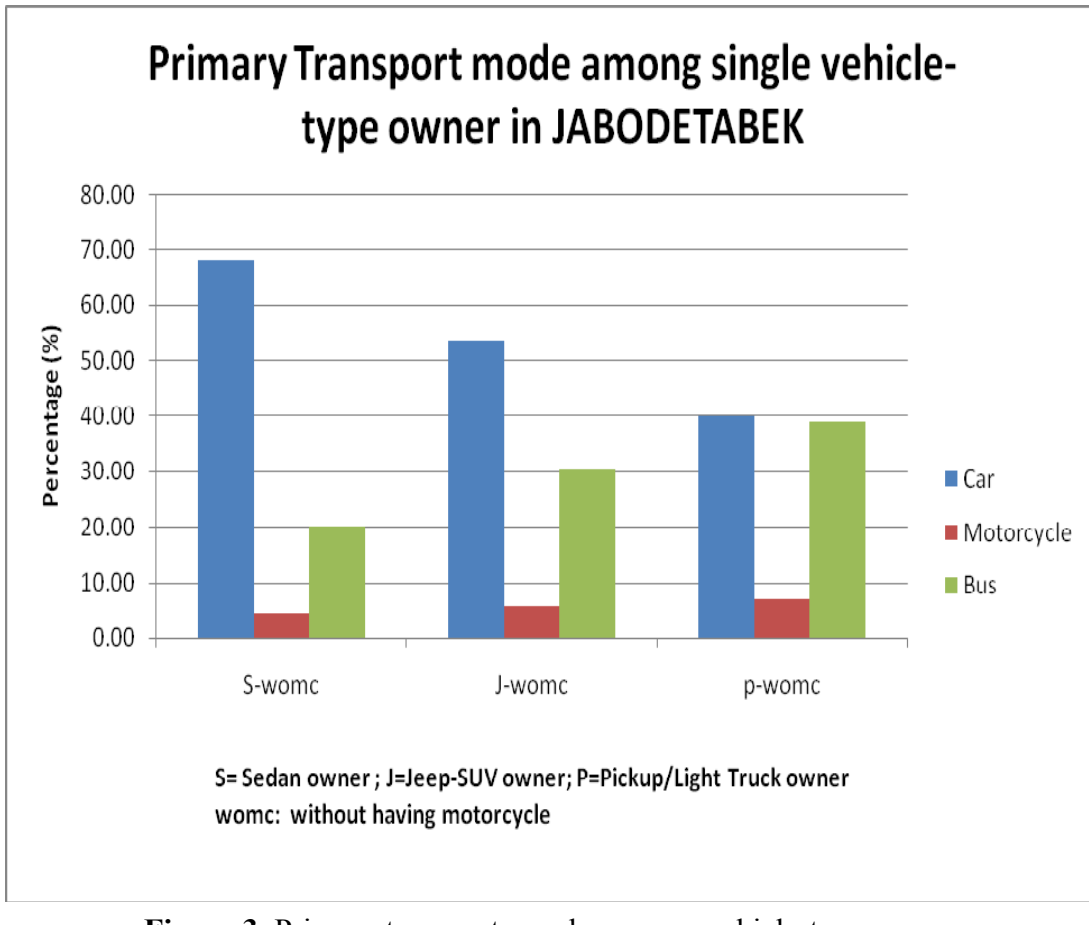


Figure 3 Primary transports modes among vehicle-type owner.

Table 6 Bivariate Probit Model based on Vehicle-Type

Variables	1999		2000		2004		2005	
	Beta	t-score	Beta	t-score	Beta	t-score	Beta	t-score
Parameter CO								
Annual Running Kilometer	0.0404	0.414	0.096	2.635	0.1063	2.139	0.2934	4.531
Engine Size Capacity	0.3477	3.826	0.5111	14.699	0.4896	8.877	0.6639	8.153
Carburettor	-0.0279	-0.115			-0.3011	-2.52	-0.0861	-0.323
Maintenance Quality	-0.1737	-2.914	-0.4388	-16.368	-0.3352	-9.574	-0.5275	-9.804
Comparing among vehicle Type								
Passenger Car /Sedan(ref)								
- Jeep,Van,SUV	-0.419	-1.914	0.195	1.385	0.2379	1.103	-0.366	-1.256
- Pickup, Light Truck	0.6086	2.439	0.6546	5.694	0.2529	1.885	0.2095	0.343
Parameter HC								
Annual Running Kilometer	-0.1097	-0.345	0.1593	2.339	0.3396	3.202	0.448	3.321
Engine Size Capacity	0.6611	1.777	0.5756	6.739	0.6788	5.009	0.5233	4.065
Carburettor	0.3318	0.482			-0.1682	-0.366	-0.2836	-0.853
Maintenance Quality	-0.0782	-0.547	0.0611	1.004	-0.0933	-1.181	-0.1226	-1.296
Comparing among vehicle Type								
Passenger Car /Sedan(ref)								
- Jeep,Van,SUV	0.1697	0.259	0.8647	1.815	4.7523	0.000	-0.0317	-0.09
- Pickup, Light Truck	0.8934	0.727	0.1866	0.859	0.3555	1.439	5.806	0.00
Disturbance correlation (rho)	0.3687	1.478	0.0198	0.19	0.3675	2.731	0.2288	1.61
Log likelihood function	-157.328		-798.204		-464.006		226.721	
Sample size	230		1386		727		441	

Conclusions and Future Reseach Issues

In this study, bivariate probit regression model applied to examine the probability emission test failure of passenger car as part of the inspection and maintenance program at Jakarta city and its surrounding city as known as Bogor, Tangerang,

Depok and Bekasi (BODETABEK) in 1999,2000,2004 and 2005. We successfully identified the characteristic of passenger car which affect significantly on emission test failure of Carbon Monoxide and Hydrocarbon. As a preliminary assessment of

the inspection and maintenance program for in-use vehicles in the capital city of Indonesia, this study indicates that vehicles running kilometer per year, fuel system injection, engine size capacity and air-to-fuel ratio play a significant role in determining I/M test results. In terms of citizen responsibility or their urban lifestyle toward low carbon society, passenger car owners should be fully responsible to do regular and routine maintenance and repair in order to reduce their passenger car emissions.

Related to vehicle-type ownership among single car household owner, family type and income level always play a significant role in determining vehicle-type own by household. Household live in sub-urban area tend to have SUV/Jeep rather than household live in CBD and urban non-CBD area. Residential period determine significantly for Sedan/Passenger car ownership by household. The probability of CO emission test empirically found much more depending on their vehicle usage rather than vehicle characteristics. Empirical study confirms that the probability of CO emission test failure of Pickup is higher than sedan/passenger car type. In order to perform cross vehicle-type evaluation for Inspection and Maintenance Program, it is necessary to explore more detail in future about individual characteristics more specific especially on fuel consumption, emission control devices, and driving behavior.

Acknowledgements

This research is partially supported by The Global Environmental Leadership Education/Training Program at Graduate School of International Development and Cooperation, Hiroshima University, Japan.

References

Adianto P. Simamora, (2006). “Garages ask for speedier emission testing approval.” The Jakarta Post-The Journal of Indoneia Today, City News. http://www.thejakartapost.com May 01.

Badan Pengelolaan Lingkungan Hidup Daerah, Propinsi DKI Jakarta, *Kumpulan Peraturan Tentang Pengendalian Pencemaran Udara di Propinsi DKI Jakarta*, BPLHD, 2005.

Bin, O.(2003). “A logit Analysis of vehicles emissions using inspection and maintenance testing data.” *Tranportation Research Part D*,(8), 215-227.

D. Stead.(1999). “Relationships between transport emissions and travel patterns in Britain.” *Transport Policy* 6, 247-258.

Bhat, C.R., S. Sen.(2006). "Household vehicle type holdings and usage: an application of the multiple discrete-continuous extreme value (MDCEV) model." *Transportation Research Part B* (40), 35-53.

Japan International Cooperation Agency (JICA) (1997). *The Study on the Integrated Air Quality Management for Jakarta Metropolitan Area*, Final Report collaborated with the Environmental Impact Management Agency (BAPEDAL) the Republic of Indonesia, Nippon Koei Co., Ltd., and Suuri Keikaku Co., Ltd.

SEI, UNEP, KEI (2002). Benchmarking Urban Air Quality Management and Practice in Major and Mega Cities of Asia, Stage 1.

Hubbard, T.(1997). "Using inspection and maintenance program to regulate vehicle emissions." *Contemporary Economic Policy 15*, 52-62.

Washburn, S. Seet, J. Mannering, F.(2001). "Statistical modeling of vehicle emission from inspection/maintenance testing data: an explanatory analysis." *Transportation Research Part D*, (6), 21-36.

K.S Chen, Wang, W.C., H.M. Chen, C.F. Lin, H.C. Hsu, J.H. Kao, M.T. Hu.(2003). "Passenger car emissions and fuel consumption in urban and rural driving conditions." *The science of the Total Environment*, (312), 113-122.

M. Pujadas, Nunez, L. Plaza, J. Bezares, J.C. Fernandez, J.M.(2004). "Comparison between experimental and calculated vehicle idle emission factors for Madrid fleet." *Science of the Total Environment*, (334-335), 134-140.

Tong, H.Y. Hung, W.T. Cheung, C.S.(1999). "Development of a driving cycle for Hongkong," *Atmospheric Environment* (33), 2323-2335.

Greene, W. (2002). *Econometric Analysis*, Prentice-Hall, New Jersey.

Dargay, J. Gately, D.(1999). "Income's effect on car and vehicle ownership, worldwide: 1960-2015." *Transportation Research* A (33), 101-138.

Siadari, M. Linda, K. Rahmat, U. Sari, P.A. (2007). *Evaluasi Kualitas Udara Perkotaan 2007*, Asdep Pengendalian Pencemaran Sumber Emisi Bergerak, Deputi Bidang Pengendalian Pencemaran Lingkungan, Ministry of Environment, Republic of Indonesia.

Ingram, G.K. Liu, Z.(1999), "Vehicles, road and road Use: alternative empirical specifications." *Policy Research Working Paper* 2036. World Bank, Washington DC, 39-40.

Vanconcellos, E.A.(1997)." The demand for cars in developing countries." *Transportation Research A* 31, 245-248.

Domestic Market, and Export-Import (2004-2008), *The association of Indonesia Automotive Industries*,http://www.gaikindo.or.id/download/statistic/02-past/01-by-category/ 05_bycat_ market_exim_04_08.pdf

de Jong, G. Fox, J. Daly, A., Pieters, M. Smit, R.(2004). "Comparison of car ownership models." *Transport Reviews* (24), 379-408.

Mohammadian, A. Miller, E.J.(2003). "An empirical investigation of household vehicle type choice decisions." *Transportation Research Record* (1854), 99-106.

Potoglou, D., P.S.(2008). "Kanaroglou. Modeling car ownership in urban areas: a case study of Hamilton, Canada." *Journal of Transport Geography* (16), 42-54.

Potoglou, D.(2008). "Vehicle-type choice and neighbourhood characteristics: An empirical study of Hamilton Canada." *Transportation Research Part D* (13), 177-186.

Senbil, M., J. Zhang. A. Fujiwara.(2007). "Motorization in Asia-14 countries and three metropolitan areas." *Journal of IATSS Research (31)*, 46-58.

Nugroho, S.B, Setyono, B. Mappabeta, A.M. Abbas, S. Pramudhianto, A., and Dewata, I. (2004). "Study on the Air Quality Improvement Analysis by New Emission Regulation of Vehicle." JICA MOE and CRHRE-UI, Jakarta.

Rubin, E.S. (2001). "Introduction to Engineering and the Environment." McGraw-Hill International Edition, Singapore.

Zhao, Y. Kockelman, K.(2001). "Household vehicle ownership by vehicle-type: application of a multivariate negative binomial model." CD-ROM *Transportation Research Board*-.2001.

Estimation of CO_2 Mitigation Potential from Freight Transport Based on System Dynamics Model—A Case Study of Beijing, China

Yunjing Wang[1], Jinling Fei[2], Xuesong Feng[3], Hidefumi Imura[4], Yoshitsugu Hayashi[5]

[1] PhD student, [2] PhD student, 3 Researcher, 4 Prof., 5 Prof., HAYASHI & KATO LAB, Graduate School of Environmental Studies, Nagoya University. Furo-cho, Chikusa-ku, Nagoya City, 464-8603, JAPAN. E-mail: yunjingw@gmail.com, imac.fei@gmail.com, charles.x.feng@gmail.com, imura@ genv.nagoya-u.ac.jp, yhayashi@genv.nagoya-u.ac.jp

Abstract

Freight transport has increased rapidly with the raid economic growth, leading to rising energy consumption and CO_2 emission. In this paper, taking the urban freight transport of Beijing as a case, three scenarios are established to evaluate the potential of CO_2 emission reduction based on system dynamics model. Scenario results show that freight turnover of Beijing will touch 80 billion t-km at the end of 2020, with an annual average growth rate around 5%, and the CO_2 emission mitigation potential range from 13-30% by 2020.

Key Words: CO_2 mitigation potential; freight transport; system dynamics model; scenario analysis

Introduction

Transport accounts for 23% of world energy-related GHG emissions, especially for CO_2 emission (IEA, 2006b) and is one of the few industrial sectors where emissions are still growing (WBCSD, 2001). Growing urbanization in developing countries has resulted in a rapid increase in urban transport services and associated energy demand. Freight transport has been growing even more rapidly than passenger transport and is expected to continue to do so in the future (IPCC, 2007). As the second largest urbanization megacity of China, the average annual growth rate of CO_2 emission from the urban transport system of Beijing during 1985-1998 was 14.4% and the CO_2 emission from urban freight transport accounted for more than 40% of the total

CO_2 emissions from the urban transport system in 1998, (SHRESTHA, 2002). The rapid growth of freight transport has been one of the major causes of CO_2 emissions in Beijing.

The analyses of energy consumption and CO_2 emissions from transport sector have been done in uncountable published literatures with particular emphasis on passenger travel. Poudenx and Merida (2007) analyzed the energy consumption and CO_2 emissions of all private and transit vehicles by the introduction of renewable energy, such as hydrogen and electricity; Zervas and Lazarou (2008) investigated the ways to weight control of new passenger cars and its impact on CO_2 emissions; Jeong et.al (2009) and Al-Hinti et.al (2009) analyzed the effects of the introduction of diesel passenger cars against CO_2 emissions; Hankey and Marshall (2009) explored how urban form affected CO_2 emissions from passenger-vehicles; Nicolas and David (2009) analyzed how individual travel can improve the understanding of CO_2 emissions by identifying upstream socio-economic factors. At the same time, the freight transport within individual countries has increased rapidly with the economic growth, leading to rising energy use and CO_2 emission. However, the literatures on urban freight transport mainly focused on the road freight transport. For example, Zanni and Bristow (2010) and McKinnon (2009) measured the CO_2 emissions from road freight transport; Léonardi (2004) surveyed CO_2 efficiency in road freight transportation linked fuel consumption to transport performance parameters; Piecyk and McKinnon (2009) assessed the CO_2 emission levels from road freight transport in 2020. Covering all urban freight modes comprehensively, however, such analyses are limited, especially for city scale.

Taking the urban freight transport of Beijing as a case, the goal of this paper is to evaluate the potential of CO_2 emission reduction by 2020 under different policy package scenarios using system dynamics model. Section 2 starts with an introduction of study area and data and the application of system dynamics model. Section 3 describes the transport and energy policies and scenario design. Finally, section 4 discusses the result and verifies the model, while section 5 presents the conclusions.

Study Area, Data and System Dynamics Model

Study Area. Beijing, as the capital and a municipality of the People's Republic of China (PRC), is a transportation hub, with a sophisticated network of roads, railways and a major airport. Four completed ring roads (with the 6th ring nearing completion) encircle the Beijing city with nine expressways heading in virtually all compass directions, supplemented by eleven China National Highways.

According to the Code for transport planning on urban road (GB50220-95), urban freight transport includes three general types: freight pass through, freight moved in and out of a city and intra-city freight transport. At city scale, here this study investigates the first two types with railway, highway, civil aviation, and pipeline

modes. When considering the CO_2 emission of each mode, however, pipeline is omitted because of data limitation on emission factors.

Data. The data used in this study comes from "Beijing Statistical Yearbook" (Beijing Municipal Bureau Statistics and NBS Survey Office in Beijing, 1982-2007). Data include annual freight turnover (t-km), freight volume (t), and network length by four main modes (highway, railway, civil aviation, and pipeline), distance carried (km), main technical and economic indicators of highway, railway, and civil aviation, as well as economic data such as annual gross domestic product (GDP) and the municipal infrastructure investment on transportation , and the conversion factors from physical unit to coal equivalent in Chinese Energy Statistical Yearbook.

System Dynamics Model. Figure 1 is the detailed physical flow diagram of the system dynamics model for this study. The general goals are to reduce fuel consumption and to mitigate CO_2 emissions of the urban freight transport in Beijing. The goals are set in terms of specific parameters which strongly impacted the freight CO_2 emissions. Freight CO_2 emissions is determined mainly by overall level of freight activity, the modal structure used, and the energy consuming characteristic of each mode. There are some other factors affecting the final freight CO_2 emissions, like the improvement of vehicle technology or fuel. These components are not completely independent, but interactional. Through various policies and investments in each mode's network expansion, environmental impacts of the system can be changed.

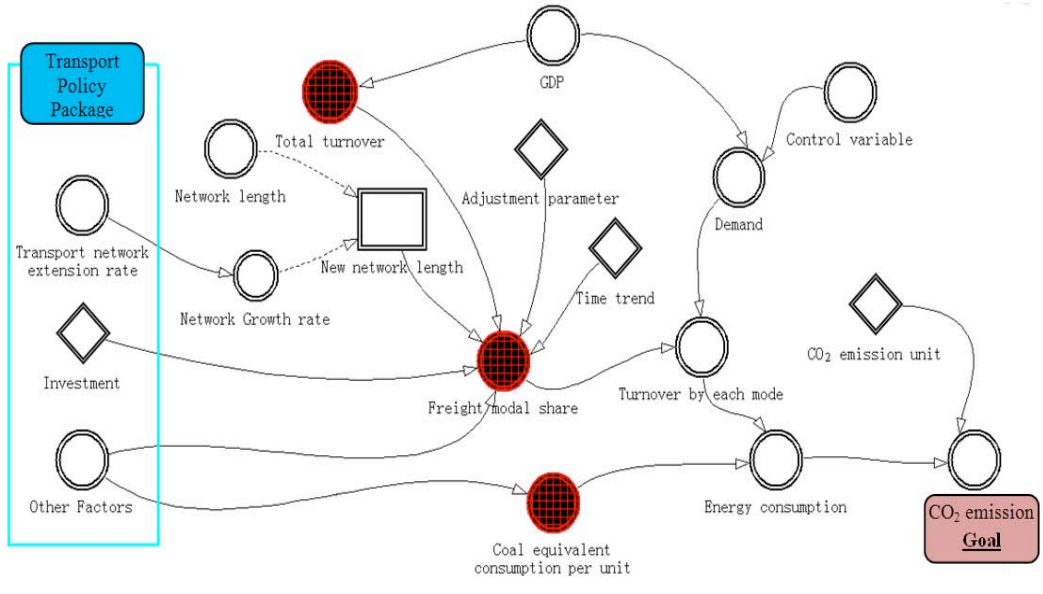


Figure 1. Flow diagram of the system dynamics model

Freight Activity. The usual indicator of freight activity is freight turnover (t-km), which measures the weight of freight and the distance it is moved for each mode. Figure 2 shows the GDP and freight turnover from 1986 to 2004 of Beijing. The total

freight activity has increased from 25.7 billion t-km in 1986 to 40.2 billion t-km in 2004, with an annual growth rate of around 3.1%, while the GDP increased more than 20 times in 18 years.

The GDP and CO_2 emissions have directly correlation since economic growth increases the demand of goods and the freight activity, which will increase the fossil fuel consumption and bring more CO_2 emission. The correlation coefficient between freight turnover and GDP is equal to 0.907, and the significantly correlation indicates that the estimation model between freight activity and GDP is reliable, and their functional line is shown in Figure 3.

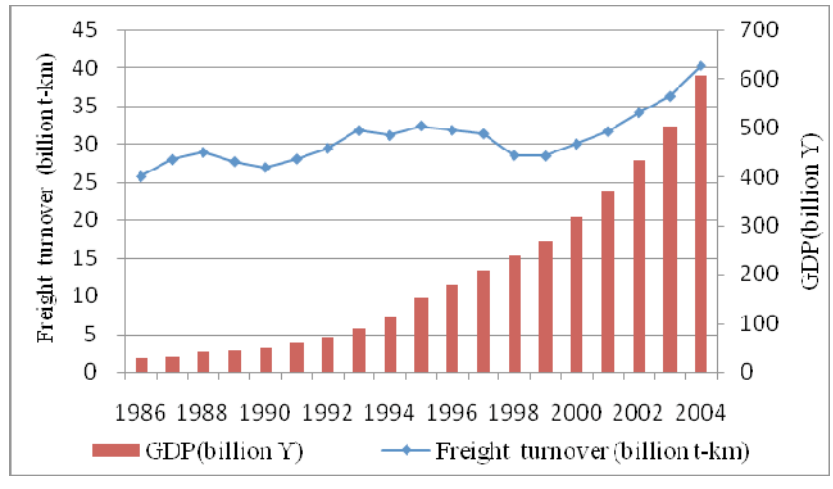


Figure 2. GDP and freight turnover of Beijing, China (1986-2004)

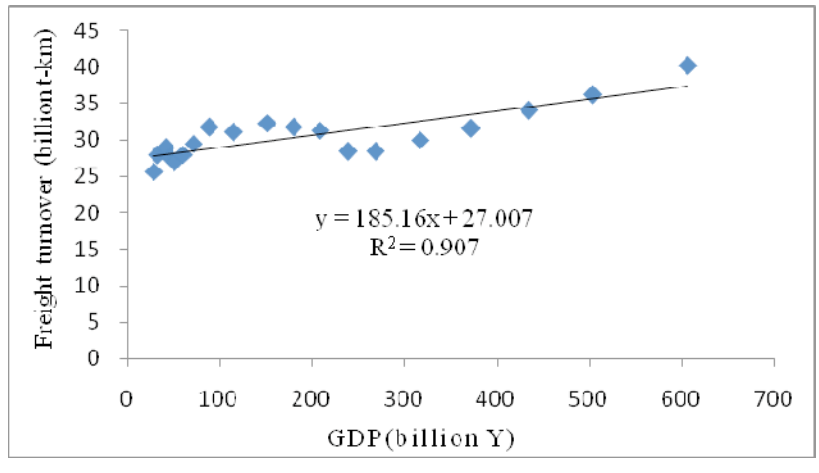


Figure 3. Estimation model for freight activity, a case of Beijing, China.

Freight Modal Structure. Freight modal structure is measured by the share of the freight activity transported by each mode: highway, railway, civil aviation, and pipeline. In our study, modal structure is assumed to be determined by the total

freight turnover, network length and its growth rate, the municipal infrastructure investment of each mode.

From the variation trend of the modal share of each mode in Figure 4, it is clearly shown that between 1986 and 2004 there was a significant shift towards highway and away from railway, although railway freight transport always had the largest share of more than 50%. More recently, with the different priority given to the various modes, road freight transport has played a more and more important role in freight movement. For example, in 2000, the investment on railway, highway, and civil aviation were about 0.03, 3.22 and 0.18 billion yuan respectively. With large amounts of investment on highway, the network length of highway increased 3.4% per year during the Ninth 5-Year period, and achieved a highest modal share of 28% in 2000. The Eleventh 5-Year Plan goes further to stress the investment and development of the highway. Furthermore, the modal share of the civil aviation and pipeline are all increasing gradually.

To assess the role of possible determinants of modal share, a multiple linear model is employed. During the estimation, a time trend variable is introduced to reflect longer-term trends in modal share not explained by the included variables.

$$MS_{i,t} = \alpha_i + \alpha_{1,i} TFT_t + \alpha_{2,i} NET_{i,t} + \alpha_{3,i} INV_{i,t} + \alpha_{4,i} T \quad (1)$$

where i and t denotes transport mode and year; MS is the percentage of modal share; TFT is the total freight turnover; NET is the network length; INV is municipal infrastructure investment of each mode; and T is a time trend variable. Table 1 shows the results of determinants of the modal share by multiple linear regressions. The adjusted R^2 in each case is close to 1.0, which indicates the linear combination of explanatory variables is sufficient and reliable for modal share estimation.

An adjustment parameter is introduced to ensure that the annual summation of all projected modal share is equal to 100%.

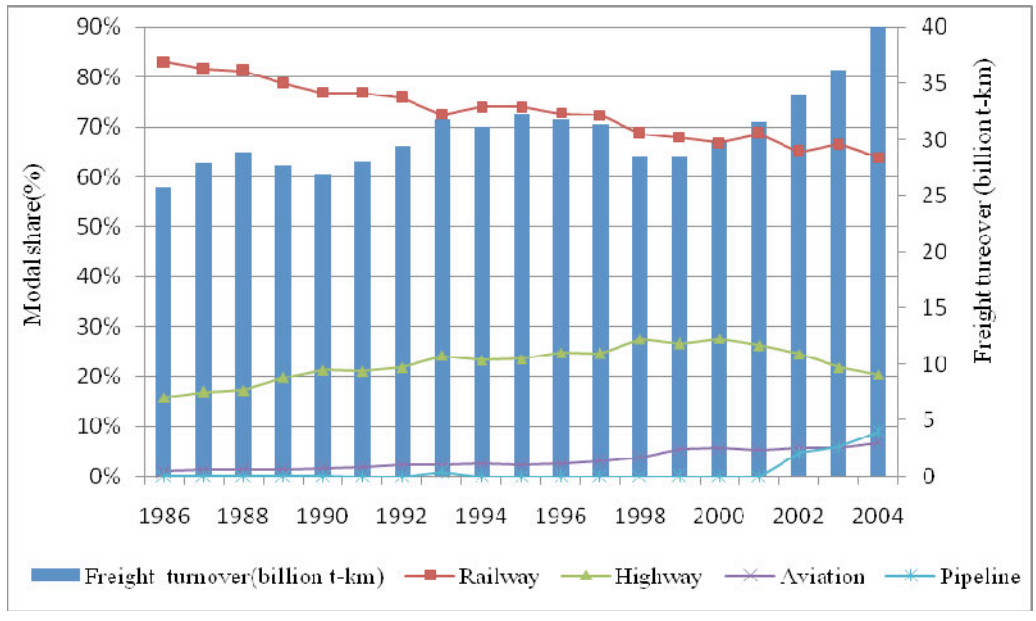


Figure 4. Freight turnover and modal share of Beijing, China (1986-2004).

Table 1. Determinants of modal share in freight transport of Beijing, China.

	Railway		Highway		Civil aviation		Pipeline	
	Coefficient	*t*-Value	Coefficient	*t*-Value	Coefficient	*t*-Value	Coefficient	*t*-Value
Total freight turnover(*TFT*)	.074	1.406	-.259	-3.688	-.022	-.942	.077	-.705
Network length(*NET*)	.189	1.085	.536	2.845	.003	.385	.000	4.749
Investment(*INV*)	.002	.726	-.004	-.497	-.003	-1.104	--	--
Time trend variable(*T*)	-.013	-6.144	-.007	-1.119	.003	4.352	.000	-.172
Constant (*a*)	24.213	9.291	13.790	1.248	-6.545	-4.079	-.873	-.705
Adjusted R^2	0.976		0.988		0.976		0.987	

Note: The data of municipal infrastructure investment on pipeline is not available.

Energy Consuming Characteristic of Each Mode. Although the system dynamics model allows consideration of different types of fuel used by each mode, here, to simplify calculation, fuel consumption is all converted to the same type of coal equivalent according to the conversion factors from the physical unit to coal equivalent in Chinese energy statistical yearbook. Figure 5 shows the coal equivalent consumption per thousand t-km for freight by each mode. The energy consumption per thousand t-km by railway and civil aviation clearly decreased from 1986 to 2004, although civil aviation was always significantly more energy consumption than any other modes.

CO_2 Emission. Since the fuels consumed in freight transport are converted to coal equivalent according to their equivalent heat volume, Eq. (2) is used for estimating CO_2 emissions based on the revised 1996 guide lines for national GHG inventories provided by the International Panel on Climate Change (1997).

$$\mathrm{EM_t} = \{\sum_i (TFT_t \times MS_{i,t} \times CEC_{i,t})\} \times \sigma \times \theta \times \varphi \times \frac{44}{12} \quad (2)$$

where *i* and *t* denote transport mode and year; *EM* is CO_2 emissions from urban freight transport; *TFT* is the total freight turnover; *MS* is the percentage of modal share; *CEC* is the coal equivalent consumption per thousand t-km by each mode; σ is heat conversion factor taking values for 2.1 * 1014J/unit; θ is a carbon emission factor per unit, which is 26 kg carbon/billion-J; φ is the fraction of carbon oxidized, estimated at 98%.

Transport Policy and Scenario Design

Since 1980, China government has implemented a series of policies and regulations to control energy consumption and CO_2 emissions in transport sector. From a national perspective, the current basic principle of China's energy strategy is to focus

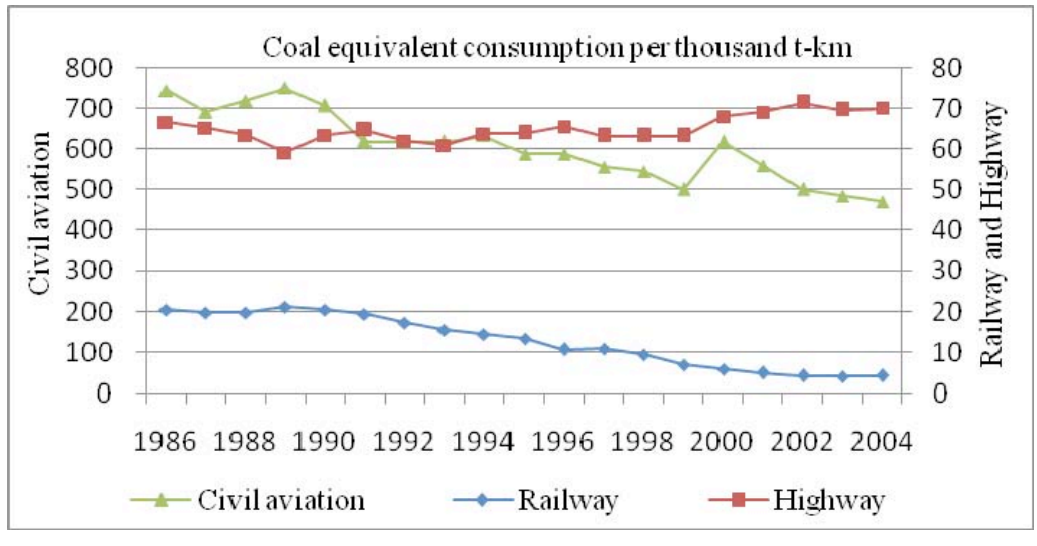


Figure 5. Coal equivalent consumption for freight of Beijing, China (1986-2004)

Table 2. Transport policies on energy conservation and CO_2 emission (1986-2004)

Policy/Regulation	Issuer	Issued year
Law on Fuel Saving Management in Transport Sector	Ministry of Transportation	1986
Energy Conservation Law of the People's Republic of China	Standing Committee of the National People's Congress etc.	1997

Detailed Rules on the Implementation of the Energy Conservation Law in the Transport Industries	Ministry of Transportation	2000
Medium and Long Term Special Plan on Energy Conservation	National Development and Reform Commission	2004
Medium and Long Term Railway Network Plan	Ministry of Railways, etc.	2004
Beijing City Master Plan(1991-2010)	Beijing Municipal Commission of Development and Reform	1993
The Tenth 5-Year Plan for Transport of Beijing	Beijing Municipal Commission of Development and Reform	2001
The Tenth 5-Year Plan for Logistics of Beijing	Beijing Municipal Commission of Development and Reform	2001

on sustainable development, which means working on energy conservation and economic development concurrently, but with an emphasis on increasing energy efficiency (Table 2). At city scale, Beijing government carried out a series of specific plan policies or regulations on freight transport to mitigate CO_2 emission. For example, one of the goals in the Tenth 5-Year Plan for logistics of Beijing is to increase the operation proportion of professionalized freight transport to decrease the energy consumption and CO_2 emission.There are three primary ways to reduce CO_2 emissions from transport: Avoid (avoid travel or avoid travel by motorized modes) , Shift (shift to more environmentally friendly modes) and Improve (improve the energy efficiency of transport modes and vehicle technology) (Dalkmann and Brannigan, 2007). Although technology improvement can help to reduce CO_2 emission by a certain extent (DfT, 2005b), a combination of avoid, shift and improve, such as a policies package of regulation, tax, better technology and demand restraint, can eventually stabilize the CO_2 emission (Geurs, 1998). Based on the analysis of previous transport policies and studies, three scenarios are built to evaluate the potential for CO_2 mitigation from freight transport. In all the three scenarios, the base year is 2000, and 2020 is set as the target year. On a macroeconomic level, all scenarios have the same hypothesis: a constant GDP growth of 9% per year up to 2020; The differences between each scenario are linked to the municipal infrastructure investment on different modes, network growth rate, related modal share, energy efficiency and different policies package to mitigate CO_2 emission.

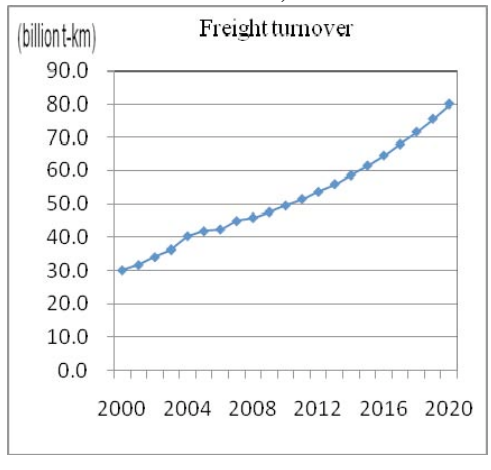
Business-As-Usual Scenario (BAU). The business-as-usual scenario (BAU) is a reference scenario showing the continuation of present trends, and there is no new regulations are implemented for urban freight transport system. In general, for the period up to 2020, the BAU scenario is based on the Tenth 5-Year Plan of Beijing and the historical trend. The extension of historical trend in transport network growth and investment are railway by 0.02% and 10%, highway by 4% and 15%, civil aviation by 12% and 4% respectively, with continuing to give priority to highway.

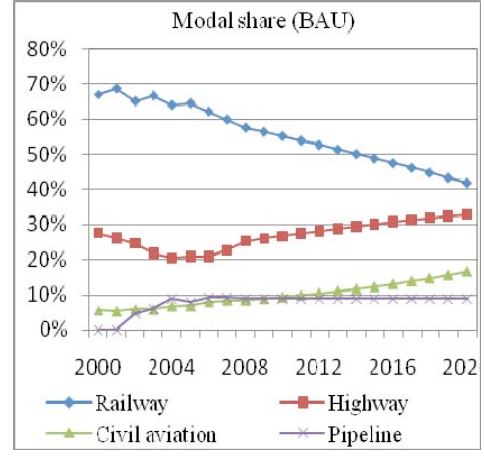
Middle Mitigation Scenario (MM). Comparing to the BAU scenario, a policy package is designed to control the rapid growth of highway and civil aviation freight in the middle mitigation scenario (MM). It is assumed that the Beijing government invests more money on the transport infrastructure construction, especially for railway and pipeline ; Electrification proportion of railway will increase to 50% and the proportion of transportation capacity professionalized vehicle will increase to 20%.

High Mitigation Scenario (HM). In the high mitigation scenario (HM), policy package for high level control and a possible instrument implementation time path are designed to achieve a sharp emission reduction goal. Based on the MM scenario, to displace part of the freight transport from highway and civil aviation to railway and pipeline, the municipal infrastructure investment increase the railway and pipeline to insure that the modal share of railway will not decline to 50% and the modal share of pipeline will grow to 11% by 2015 and 13% by 2020. Additional, the energy efficiency will increase by 5% to contribute the CO_2 mitigation.

Result and Model Verification

Result. Figure 6 is the estimation result of the total freight turnover and the modal share of the three scenarios. It shows that freight turnover will reach up to 80 billion t-km at the end of 2020, with an annual average growth rate around 5%. Under the





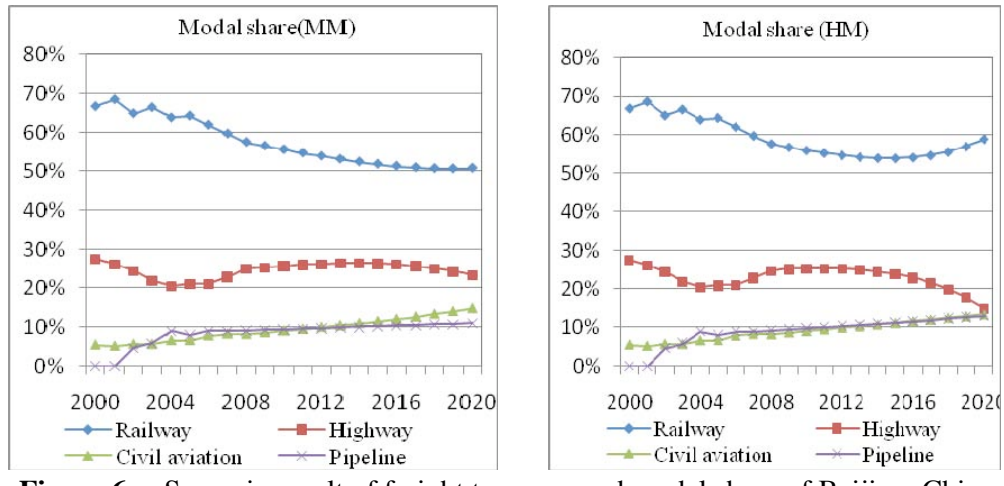


Figure 6. Scenario result of freight turnover and modal share of Beijing, China (2000-2020)

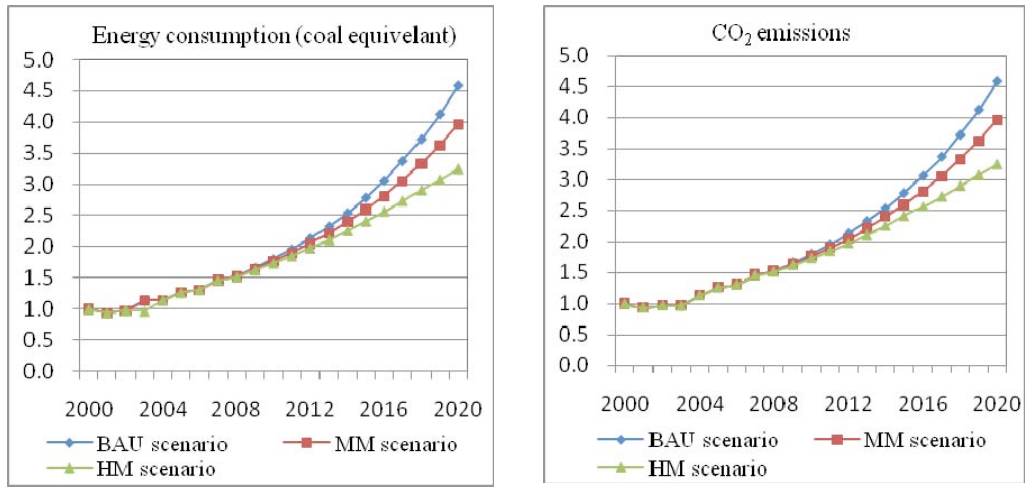


Figure 7. Scenario result of energy consumption and CO_2 emission of Beijing, China (2000-2020).

Table 3. Error verification of model result

Year	Total freight turnover	Modal share				CO_2 emission
		Railway	Highway	Civil aviation	Pipeline	
2001	7.2%	-1.4%	-6.5%	-9.8%	-6.7%	-3.9%
2002	2.8%	2.9%	-11.3%	-9.9%	-2.8%	-8.3%
2003	0.3%	-1.7%	10.8%	-6.3%	-0.3%	-0.1%
2004	-5.0%	1.8%	8.8%	5.9%	5.2%	4.0%
2005	-4.8%	-5.0%	5.6%	10.5%	5.1%	5.3%
2006	-1.8%	0.5%	5.0%	1.0%	1.8%	0.2%
2007	-1.3%	-1.9%	2.1%	-5.6%	1.3%	-4.8%

Absolute average value	3.3%	2.2%	7.1%	7.0%	3.3%	3.8%

BAU scenario, civil aviation contribute the most to transport growth with its share increasing by 10% in 20 years, while the share of railway sharply decreased from 67% to 42% .Under the middle and high mitigation scenarios, the introduction of specific policies package aiming at energy conservation and CO_2 mitigation supports environmentally friendly modes such as railway and pipeline, while the highway's share is reduced because its relatively high energy consumption and CO_2 emissions. For example, in the high mitigation scenario, the share of railways is close to 59% by 2020 with the share of other modes are all less than 15%.

Figure 7 is the scenario result of energy consumption and CO_2 emission. Based on the year 2000, the energy consumption (coal equivalent) and CO_2 emission in 2020 will decrease by 13% and 30% under the middle mitigation and high mitigation scenario respectively.

Model Verification. To verify the system dynamics model in this study in projecting total freight turnover, modal share and CO_2 emissions, historical statistical data in selected years are used to compare with the BAU scenario. Table 4 shows the results, the errors of all variables are under10% except two top values of 11.3%, 10.5%, and the absolute average value of error are less than 8%, which illustrates the model and the estimation result are reasonable. the errors of all variables are under10% except two top values of 11.3%, 10.5%, and the absolute average value of error are less than 8%, which illustrates the model and the estimation result are reasonable.

Conclusion

Generally speaking, CO_2 emissions in the freight transport of Beijing till 2020 will continue increasing, even under the high mitigation scenario. Concluding from the above analysis, a huge emission reduction potential exists in the current freight transport system of Beijing. By 2020, the emission reduction potential range from 13-30% based on the level of 2000. Accelerating the development of railway network and improving the energy efficiency will make great contributes to the CO_2 mitigation.

Acknowledgements

The authors express great appreciation to the anonymous referees who provided valuable suggestions that led to an improved manuscript.

References

A.C. McKinnon, M.I. Piecyk. (2009). "Measurement of CO_2 emissions from road freight transport: A review of UK experience." *Energy Policy,* 3733-3742.

Alberto M. Zann, Abigail L. Bristow. (2010). "Emissions of CO2 from road freight transport in London: Trends and policies for long run reductions." *Energy Policy,* 1774-1786.

DfT.2005a). "Visioning and back casting for uk transport policy (vibat). stage 1 baseline report." *Department for Transport*, Crown Copyright, 81.

Efthimios Zervas, Christos Lazarou. (2008). "Influence of European passenger cars weight to exhaust CO_2 emissions." *Energy Policy*, 248-257.

Holger Dalkmann, Charlotte Brannigan, (2007). "Transport and climate change, Module 5e: sustainable transport: a sourcebook for policy-maker in developing cities." GTZ, 4-5.

I. Al-Hinti, A. Al-Ghandoor, B. Akash, and E. Abu-Nada. (2007). "Energy saving and CO_2 mitigation through restructuring Jordan's transportation sector: The diesel passenger cars scenario." *Energy Policy*, 5003-5011.
IEA, 2006b: World Energy Outlook. (2006). International Energy Agency, Paris, 596 -597.

International Panel on Climate Change, 2007. IPCC Fourth Assessment Report, Working Group III Report ''Mitigation of Climate Change", 325.

Jacques Léonardi, Michael Baumgartner. (2004). "CO_2 efficiency in road freight transportation: Status quo, measures and potential." *Transportation Research Part D: Transport and Environment*, 451-464.
Karst Geurs, Bert van Wee. (1997). "Sustainable transport: consequences of a 89-90%emission reduction". http://www.etcproceedings.org/paper/download/2284.

Maja I. Piecyk, Alan C. McKinnon. (2009). "Forecasting the carbon footprint of road freight transport in 2020." *International Journal of Production Economics*, 1-10.

Pascal Poudenx, Walter Merida. (2007). "Energy demand and greenhouse gas emissions from urban passenger transportation versus availability of renewable energy: The example of the Canadian Lower Fraser Valley." *Energy*, 1-9.

Ram M. SHRESTHA, Rabin SHRESTHA, Gabrial ANANDARAJAH. (2002) "Mitigating environmental emissions from the transport system of selected Asian cities: analyses of technical options." *Proceedings of Workshop of IGEC/APN Mega-City Project*, 1-12.

Ram M. SHRESTHA, S.C. Bhattacharya, Nazrul lslam, N.T. Kim Oanh. (2002) "Mitigating environmental emissions from the urban transport system." *COP8*, New Delhi, India.

Steve Hankey, Julian D. Marshall. (2009). "Impacts of urban form on future US passenger-vehicle greenhouse gas emissions." *Energy Policy*, doi:10.1016/j.enpol.

Suk Jae Jeong, Kyung Sup Kim, and Jin-Won Park. (2009). "CO_2 emissions change from the sales authorization of diesel passenger cars: Korean case study." *Energy Policy*, 2630-2638.

WBCSD, 2001. World Business Council for Sustainable Development. Mobility 2001: World Mobility at the End of the Twentieth Century and Its Sustainability. Onlinehttp://www.sustainable-finance.org/web/projects/mobility/english_full_report.pdf.

An Analytical Method for Intersection Entering Traffic Flow Estimation

Rui Wang[1], Ryota Horiguchi[2], Shinji Tanaka[3] and Masao Kuwahara[4]

[1] Dr. Eng., Research Associate, Dept of Civil Eng., Yokohama National Univ., 79-5 Tokiwadai, Hodogaya, Yokohama, 240-8501 Japan; Email: wang-rui@ynu.ac.jp
[2] Dr. Eng., President, i-Transport Lab. Co., Ltd., 1-4 Jinbo-cho, Chiyoda-ku, Tokyo, Japan
[3] Dr. Eng., Lecturer, [4] Ph.D., Prof., Institute of Industrial Science, Univ. of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo, Japan

Introduction

In this paper, we try to provide a simple analytical approach to describe distribution of intersection entering traffics on a grid road network. The approach is a continuous model with explicit structure and free of the need of a detailed OD (Origin-Destination) table.

For the purpose of proper road network planning and controls, it is very important to quantify the traffic distribution on a road network. Currently the most widely used approaches can be roughly divided into two types: the static traffic assignment models and the dynamic traffic simulation models, either macroscopic or microscopic ones. However, the traditional models demands very detailed OD information, with inevitable complicated, black-box type of algorithms. As the result, the users often reach to the outputs without knowing the underlying logic behind. The traditional models are all link oriented, except the microscopic analysis, intersections play a trivial role (if any) in the traditional models, despite the fact that intersections rather than basic sections of road links, decide the performance of urban road networks in modern cities.

In this paper we assume the passing probability of an intersection equals to the utilization of the intersection in the routes between OD pairs within a road network. Under this assumption, an explicit analytical model is developed to describe the distribution of traffic within a grid network by a single pair of OD, firstly a discrete model and followed by a continuous model that can be subjected to the method of integral in the future. Finally we

use a proved commercial simulation model to validate the proposed approach. Our approach can be regarded as a simplified estimation method of traffic distribution and this work is also helpful for the researches on the road network density. Through our research we do not consider about the influence of congestion since our objective is limited to the planning stage only.

An Discrete Intersection Utilizing Model of a Single OD Pair

A road network consists of both links (roads) and nodes (intersections). For most modern urban road networks, the common knowledge is that the smoothness of traffic movement is more likely decided by intersections, rather than basic road sections. This research concentrates on the intersections by assuming all OD trips are all between them instead of the traditional zones. Similarly, routes between any OD*s* are also described by a sequence of intersections, rather than links.

A Road Network Model. A road network consists of intersections and roads connected them is shown in Figure 1. In this paper, we assume the roads (solid lines) as of high-hierarchical with a uniform background of low-hierarchical roads (grey shade), as suggested by H. Lehmanh[1)].

A Discrete Intersection Choice Model for Estimating Directional Intersection Entering Traffics. As stated earlier, in this paper, we assume the percentage of traffic crossing a certain intersection equals to the utilizing probability of the intersection in the routes between *O* and *D*.

Number of routes from *O*. Route choice behavior is defined by choices of intersections in this paper to estimate directional intersection entering traffics. For a network of *IJ* with *O* and *D* at two opposite corners, the route numbers from *O* to all the intersections are shown in Figure 2(a). If we define $R(O, I)$ is the number of routes from *O* to the intersection $I_{i,j}$, we can have the Eq. (1).

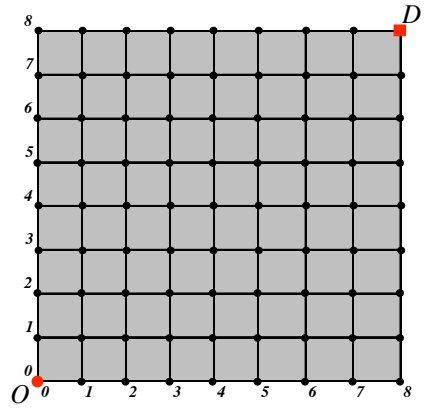
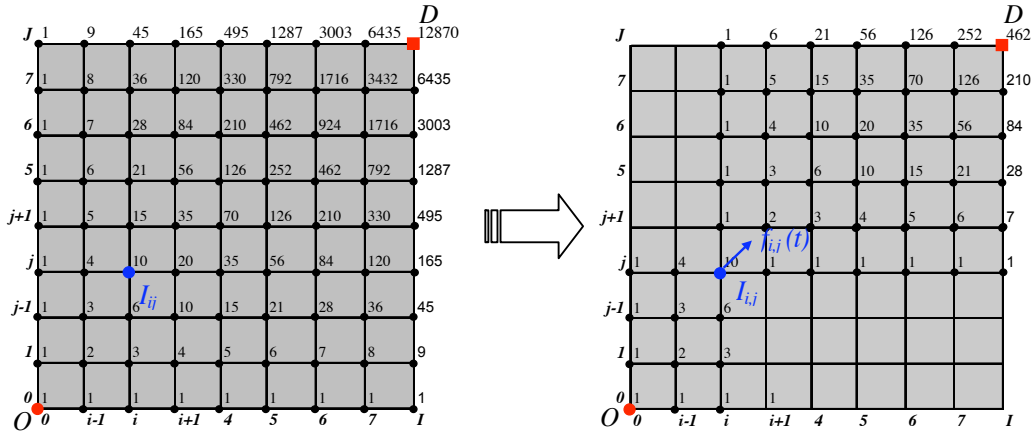
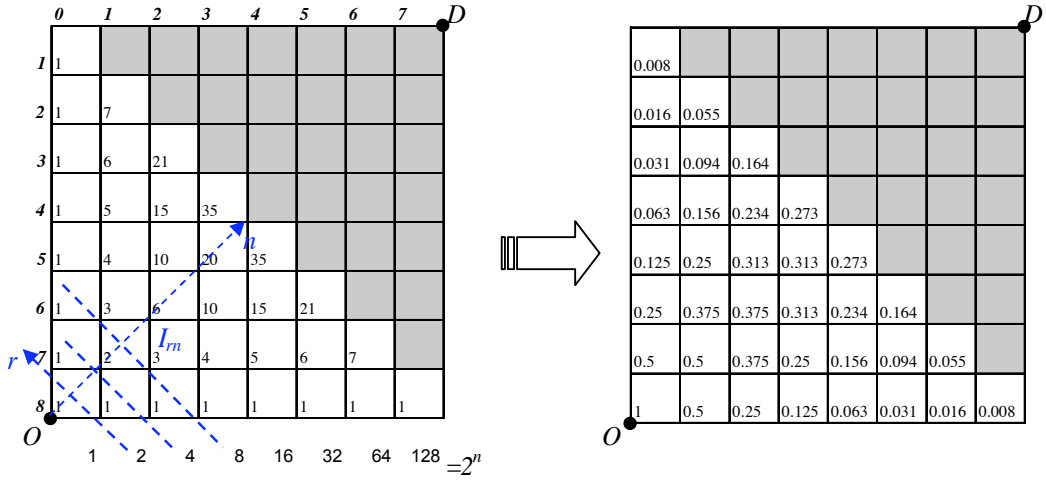


Figure 1. A road network model.



(a) $R_o(i,j)$, route numbers from O (b) Estimating directional traffics

Figure 2. An Intersection choice model.



(a) Number of routes (b) Utilizing probability

Figure 3. Intersection utilizing probability of a symmetrical network

$$R(O,I_{ij}) = \frac{[(i-0)+(j-0))]!}{(i-0)!(j-0)!} = \frac{(i+j)!}{i!\,j!} \tag{1}$$

Estimating directional traffics. As shown in Figure 2(b), at the intersection $I_{i,j}$, the total entering traffic at time t, $f_{ij}(t)$, can be estimated by the following Eq. (2).

$$f_{ij}(t) = \rho_{OD}(t)\frac{R(O,I_{ij})\cdot R(I_{ij},D)}{R(O,D)} = \rho_{OD}(t)[\frac{(i+j)!}{i!\,j!}\cdot\frac{(I-i+J-j)!}{(I-i)!(J-j)!}/\frac{(I+J)!}{I!J!}] = \rho_{OD}(t)\cdot p_{ij} \tag{2}$$

where $R(Iij, D)$ is number of routes from Iij to D, $R(O, D)$ is the total number of routes from O to D, and $\rho_{OD}(t)$ is the trip generating/attracting rate between O and D.

Definition of Intersection Utilizing Probability (p_{ij}),: here we define p_{ij} as the crossing probability from O to D and it can be calculated by the Eq. (3). Obviously, pij depends on the network itself thus independent of the demand on the network or the OD pattern.

$$p_{ij} = \frac{R(O, I_{ij}) \cdot R(I_{ij}, D)}{R(O, D)} = \frac{(i+j)!}{i!\,j!} \cdot \frac{(I-i+J-j)!}{(I-i)!(J-j)!} / \frac{(I+J)!}{I!J!} \tag{3}$$

Development of a Continuous Intersection Utilizing Model

In this section, we try to develop a continuous model on the basis of the discrete model by assuming a symmetrical road network with infinite road density. A symmetrical grid network is shown in Figure 4, the number of routes and the utilizing probability can be easily estimated by using the earlier model.

As for the new *r-n* coordinate system, when the traffic from O to D crossing the blue lines, p_{rn} the utilizing probability of an intersection I_{rn} on these blue lines can be calculated by the following Eq. (4) since it consist of a perfect binomial distribution.

$$p_{rn} = \frac{R(O, I_{rn})}{\sum_{r<=n} R(O, I_{rn})} = \frac{\binom{n}{r}}{\sum_{r<=n} \binom{n}{r}} = \binom{n}{r} / 2^n = \frac{n!}{r!(n-r)!} (\frac{1}{2})^n \tag{4}$$

where $R(O, I_{rn})$ is the number of routes from O to I_{rn}.

As $n \to \infty$, by assuming $x = (r - n/2)/(\sqrt{n}/2)$, applying the Stirling Equation of $n! \cong \sqrt{2\pi} e^{-n} n^{n+1/2}$, an appropriate continuous equation can be obtained[2].

$$p(r) \cong \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2}} \tag{5}$$

where $\sigma = \sqrt{n}/2$.

Validation of the Proposed Model

In order to validate the proposed models, a 9 by 9 grid road network is constructed under a simulation platform of SOUND/4U©. The simulation platform is based on a macroscopic model developed under the cooperation between the Industrial Science Institute, University of Tokyo and i-Transport Lab. Co., Ltd (ITL).

Estimation Results.

For the purpose of comparison, by assuming $\rho_{OD}(t)$=10000, intersection entering traffics of the grid network are estimated by both equation (4) and the simulation model. The results can be found in the Tables 1 and 2.

Table 1. Estimation results of the proposed models.

n \ *r*	0	1	2	3	4	5	6	7	8	Sum
0	10000	-	-	-	-	-	-	-	-	10000
1	5000	5000	-	-	-	-	-	-	-	10000
2	2500	5000	2500	-	-	-	-	-	-	10000
3	1250	3750	3750	1250	-	-	-	-	-	10000
4	625	2500	3750	2500	625	-	-	-	-	10000
5	313	1563	3125	3125	1563	313	-	-	-	10000
6	156	938	2344	3125	2344	938	156	-	-	10000
7	78	547	1641	2734	2734	1641	547	78	-	10000
8	39	313	1094	2188	2734	2188	1094	313	39	10000

Table 2. Estimation results of SOUND.

n \ *r*	0	1	2	3	4	5	6	7	8	Sum
0	9906	-	-	-	-	-	-	-	-	9906
1	4324	5491	-	-	-	-	-	-	-	9815
2	2553	3961	3209	-	-	-	-	-	-	9723
3	1386	3119	3344	1782	-	-	-	-	-	9631
4	735	2164	3295	2442	904	-	-	-	-	9540
5	343	1399	2778	1893	1627	408	-	-	-	8448
6	122	793	2128	2937	2301	919	156	-	-	9356
7	32	385	1381	2693	2727	1538	462	47	-	9265
8	6	121	811	2099	2893	2146	939	151	7	9166

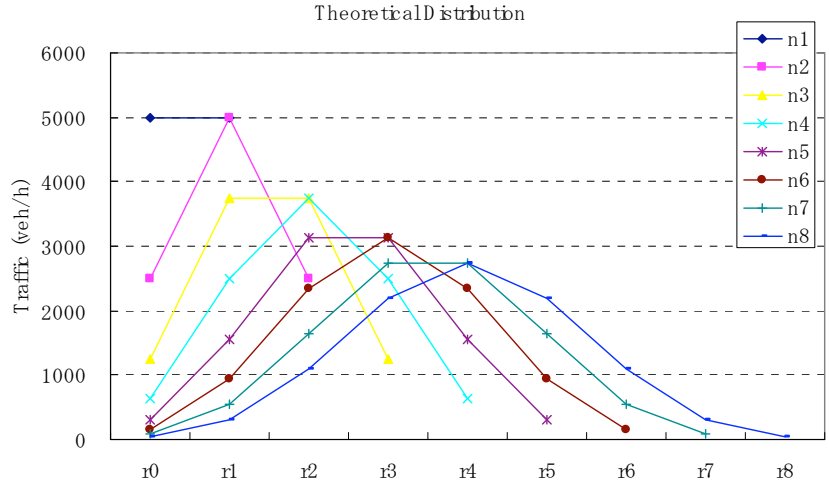
Traffic Distribution.

Comparison of distributions of intersection entering traffics is illustrated in Figure 4. For simulation results, since many vehicles stay on the roads when the calculation finished, the further the intersections, the fewer the intersection entering traffics are estimated. General speaking, the proposed model can give out an acceptable estimation result.

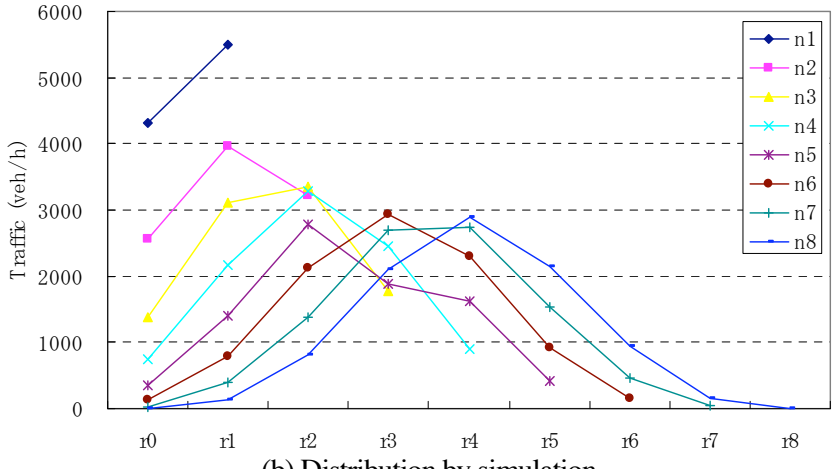
Concluding Remarks

In this research, an analytical approach is proposed to describe intersection entering traffic distributions on a grid network. Both a discrete and a continuous model is developed and the estimation results are validated by a mature commercial simulation platform. The proposed approach is full analytical with explicit structure; its influence on both the academic and practical fields is expected.

In future, the author will concentrate on two parts: 1) the case of a full OD table and 2) the case of non grid networks. The major purpose of the future research will also be insist on, to describe a network using the network parameters only.



(a) Distribution by proposed model



(b) Distribution by simulation

Figure 4. Comparison of distributions of intersection entering traffics.

Acknowledgements

The SOUND/4U Simulation platform kindly provided by the i-Transport Lab. Co., Ltd (ITL) is highly appreciated. Thank you very much, Mr. Hanabusa, for all the kindly support! The kindly suggestions and encouragement from Prof. Nakamura, Prof. Okamura and other members of the transportation lab at Yokohama National University is also highly appreciated.

References

Lehmann, H. (2003). "Directed and diffusive contributions to urban traffic flow patterns." *EuroPhysics Letters*, 64 (2), 288-294,

Yamamoto, K.. (1986). *Permutation, Combination and Probability*, Iwanami Shoten Publishers, Tokyo,.234-239.

Cycle Length Optimization at Isolated Signalized Intersections from the Viewpoint of Emission

Danpeng Ma[1] and Hideki Nakamura[2]

[1] Doctoral student, [2] Prof., Dept. of Civil Eng., Nagoya Univ., Furo-cho, Chikusa-ku, Nagoya, 464-8603, Japan; Email: ma@genv.nagoya-u.ac.jp

Abstract

Emission is one of the important measures of effectiveness (MOEs) when developing traffic signal control setting. Alleviating emission burden through traffic signal control is cost-effective compared with other environment-friendly measures. This paper aims to investigate cycle length optimization at isolated signalized intersections from the viewpoint of emission. First, emission can be divided into four parts in terms of contributions of vehicle's cruise, idle, deceleration and acceleration modes. Second, emission from the whole intersection is modeled as a function of cycle length, and the cycle length minimizing emission is identified as $C_{emission}$. Then its characteristics is investigated and compared with the delay based cycle length. Finally $C_{emission}$ is recommended to be applied on conditions with high cruise speed or high heavy vehicle percentage.

Key Words: Emission; cycle length; delay; stops

Introduction

Many countermeasures have been proposed and implemented to alleviate emission burden. For example, more trails for freight are recommended to be built, and environment-friendly vehicles are promoted. However, most of these measures are costly, and the wide-scale implementation may be many years away. Traffic signal control is more cost-effective to reduce emission, e.g. NO_x, PM, and it realizes quick effect.

In practice, minimizing delay and increasing capacity are most widely used criteria to evaluate traffic signal control settings. The objective to minimize signalized intersection delay cannot necessarily minimize emission. This is because emission is closely related to stops besides the impact from delay.

Cycle length is an important signal control parameter influencing on signalized intersection performance. This paper aims to investigate cycle length optimization at isolated signalized intersections from the viewpoint of emission. To better understand characteristics of the emission based setting, it will be compared with the delay based setting.

Literature Review

There are few literature about traffic signal control setting from the viewpoint of emission. Most of the existing works concern on developing estimation models of vehicle emission or optimizing control parameters by using some specific models estimating emission.

Oneyama, et al. (2001) developed a methodology for NO_x estimation by investigating a relationship between engine condition and emission. The model parameters were calibrated using data measured by real running pattern on a chassis dynamometer. By utilizing the developed NO_x estimation model, the cycle length minimizing NO_x was calculated at a given green time split under a balanced directional flow condition. The optimized value was found different with that for minimizing delay. The difference varied due to saturation degree and directional demand ratio.

Li, et al. (2004) proposed an emission based signal control method optimizing signal cycle length and green time split. By establishing an integrated objective function for traffic flow quality, fuel consumption and emission, the optimized cycle length and green time split were calculated. The cycle length minimizing emission was found comparably longer than that minimizing delay especially at higher saturation level.
As shown above, existing research results have shown some difference between the emission based and the delay based optimum traffic signal settings. However, root source of such difference was not discussed in detail. Also, the conditions where the emission based setting is advantageous and reasonable should be further investigated.

Methodology

Components of emission at a signalized intersection. At a signalized intersection, E_C, E_I, E_D, E_A are defined as components of emission respectively which are attributable to contribution of cruise, idle, deceleration, acceleration. Because the contribution from per cruise distance can be regarded as constant, this study concerns on E_I, E_D, E_A.

Emission Performance Index. For the emission based traffic signal setting, emission performance index E is defined as total amount of emission in a certain time interval. This can be expressed by sum of emission from individual vehicles as Eq. (1)

$$E = E_I + E_D + E_A = \sum_{i=1}^{n} \frac{de_{ii}}{dt} \cdot d_{si} + \sum_{i=1}^{n} (e_{di} + e_{ai}) \quad (1)$$

where,
E: total emission (g),
n: stops,
de_{ii}/dt: emission rate (g/s) of idle mode for vehicle i, i=1, 2, 3…n
d_{si}: the stopped delay time for vehicle i,
e_{di}, e_{ai}: emission (g) for one deceleration and acceleration of vehicle i respectively.
Emission rate can be computed using the existing empirical methods of estimating emission. Here, $C_{emission}$ is defined as cycle length minimizing emission in a given time interval.

Cycle Length Optimization

This section will introduce procedures of cycle length optimization from the viewpoint of emission by using emission performance index. First, assumptions are given for modeling emission. Then emission from one approach is modeled. By summing up emission from all approaches, emission of whole intersection is obtained. Finally $C_{emission}$ is achieved.

Assumptions. In order to facilitate formulation of intersection emission modeling, the following assumptions are made about intersection configuration, traffic demand, and signal phasing. Also emission is supposed to be produced within vicinity of intersection. Green time split is assigned to keep the same v/C. Assumed intersection configuration and signal phasing are illustrated in Figures 1 and 2, respectively.

Other assumptions are as follows:
(1) Uniform arrival and under-saturated,
(2) Through traffic only,
(3) The same saturation flow rate (S) for all approaches,
(4) Balanced demands for opposite approaches of major and minor streets.

Intersection Emission Modeling. In time interval T ( ), emission of the total intersection can be expressed as Eq. (2). It is a sum of emission from all approaches of major and minor streets.

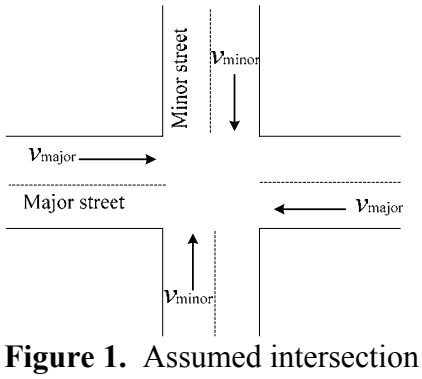


Figure 1. Assumed intersection configuration



Figure 2. Assumed signal phasing

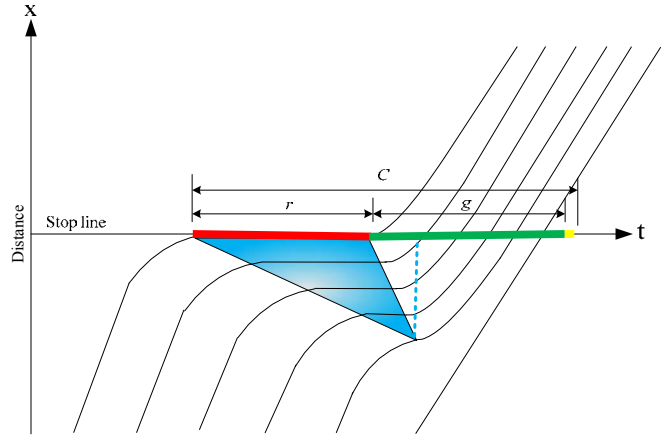


Figure 3. Time-distance diagram.

$$E_{total} = 2 \cdot \frac{T}{C}(E_{major}(C) + E_{\min or}(C)) \tag{2}$$

where E_{total}: total emission (g) in time interval T, $E_{major}(C)$, $E_{minor}(C)$: emission (g) from approaches of major and minor streets in one cycle respectively, and T/C: the number of cycles during time interval T.

Hence emission from an approach needs to be modeled.

Estimation of approach emission. A traffic flow from an approach of a signalized intersection in a cycle is shown in Figure 3. Based on Eq. (1), the total emission from it can be computed by summing up emission from individual vehicles.

Assuming homogeneous vehicle type and uniform cruise speeds, deceleration and acceleration rates, Eq. (1) can be simplified as Eq. (3).

$$E = \frac{de_i}{dt} \times \sum_{i=1}^{n} d_s + n \times (e_d + e_a) \tag{3}$$

where E: total emission (g) for a traffic flow from an approach, de_i/dt: emission rate for idle mode (g/s), n: stops, $\sum d_s$: total stopped delay time (s), and e_d, e_a: emission (g) for one deceleration and acceleration respectively.

Under the assumption of uniform arrival, Eq. (4) is utilized to calculate total stopped delay time and stops.

$$\sum_{i=1}^{n} d_s = \frac{C(1-\frac{(C-L)\theta}{C})^2}{2(1-\frac{v}{S})}, \quad n = v \times \frac{1-\frac{(C-L)\theta}{C}}{1-\frac{v}{S}} \tag{4}$$

where L: total lost time per cycle (s), C: cycle length(s), θ: green time split, v: traffic flow rate (veh/h), and S: saturation flow rate (veh/h of green time).

Function of intersection emission estimation. For simplifying formulation of intersection emission modeling, two variables are defined in Eq. (5).

$$X = \frac{v_{major}}{S - v_{major}}, Y = \frac{v_{minor}}{S - v_{minor}} \tag{5}$$

where v_{major}, v_{minor}: traffic demands of approaches along major and minor streets respectively, and S: saturation flow rate (veh/h of green time).

After transformation, Eq. (2) can be expressed as a function of cycle length as Eq. (6).

$$E_{total} = \frac{de_i}{dt} T\left[X \cdot S \cdot (1-\theta)^2 + Y \cdot S \cdot \theta^2\right] C + T \frac{\frac{de_i}{dt} X \cdot S \cdot L^2 \cdot \theta^2 + \frac{de_i}{dt} Y \cdot S \cdot L^2 \cdot (1-\theta)^2 + 2(e_d + e_a) X \cdot S \cdot L \cdot \theta + 2(e_d + e_a) Y \cdot S \cdot L \cdot (1-\theta)}{C}$$
$$+2T\left[\frac{de_i}{dt} X \cdot S \cdot L \cdot \theta \cdot (1-\theta) + \frac{de_i}{dt} Y \cdot S \cdot L \cdot \theta \cdot (1-\theta) + (e_d + e_a)(1-\theta) S \cdot X + (e_d + e_a)\theta \cdot S \cdot Y\right] \tag{6}$$

where T: time interval (s), L: total lost time per cycle (s), θ: of green time split of major street, X, Y: defined in Eq. (5), de_i/dt: emission rate for idle mode (g/s), and e_d, e_a: emission (g) for one deceleration and acceleration, respectively.

$C_{emission}$ derivation. By putting differential of total emission equal to zero with regard to cycle length, $C_{emission}$ can be obtained in Eq. (8).

$$\frac{\partial E_{total}}{\partial C} = 0 \tag{7}$$

$$C_{emission} = \sqrt{L^2 \frac{X\theta^2 + Y(1-\theta)^2}{X(1-\theta)^2 + Y\theta^2} + 2L \frac{(e_d + e_a)}{\frac{de_i}{dt}} \frac{X\theta + Y(1-\theta)}{X(1-\theta)^2 + Y\theta^2}} \tag{8}$$

According to formation of $C_{emission}$, it is sensitive to total lost time, degree of saturation and emission rates.

A variable is declared as Fe which is expressed in Equation (9). Fe stands for how many seconds of the stopped delay can produce the same amount of emission as that from one deceleration and acceleration.

$$Fe = \frac{(e_d + e_a)}{\frac{de_i}{dt}} \tag{9}$$

With respect to $C_{emission}$, Fe is of particular significance. Fe depends on emission type, vehicle's engine type, vehicle's initial cruise speed, etc.

Difference with the emission based optimum cycle length. Under the assumption of uniform arrival, the delay based optimum cycle length is the required minimal cycle length which is expressed as Eq. (10)

$$C_{min} = \frac{L}{1 - Y} \tag{10}$$

where L: total lost time per cycle (s), and Y: sum of the degree of saturation for all critical phases.

Difference can be identified through comparison between Eqs. (8) and (10). The main cause is that derivation of $C_{emission}$ takes into consideration impacts of both delay and stop whose optima do not coincide.

Application of $C_{emission}$

This section will explore application of $C_{emission.}$ Conditions will be identified where the emission based cycle length settings are reasonable and advantageous.

Through discussions above, the minimal emission under $C_{emission}$ is obtained through decreased stops with sacrificing delay performance.

As aforementioned, E_I, E_D, E_A stands for components of emission respectively which are attributable to contribution of idle, deceleration, acceleration. E_I is associated with delay. E_D and E_A depend on stops. From the viewpoint of emission, on one hand longer cycle length can result in lower stop rate due to higher portion of green time in given time interval. This will cause lower E_D, E_A. On the other hand, longer cycle length will bring increased delay level, and this arouses higher E_{Idle}.

Therefore in order to better explore application of $C_{emission}$, sensitivity of E_I, E_D, and E_A with regard to cycle length will be analyzed.

"Marginal Cost" and "Marginal Benefit". The absolute value of gradient $\partial E_I/\partial C$ can be regarded as "marginal cost" which is an increased amount of E_I with increase of one unit of cycle length. Under homogeneous vehicle type and aforementioned assumptions, this "marginal cost" is constant as shown in Figure 4.

The absolute value of gradient $\partial(E_D+E_A)/\partial C$ can be regarded as "marginal benefit" which is a decreased amount of E_D and E_A with increase of one unit of cycle length. Under homogeneous vehicle type, uniform cruise speeds and the aforementioned assumption, "marginal benefit" can be shown in Figure 5. Due to decreasing and declining stop rate, it decreases and its sensitivity drops off.

Values of k are highly dependent on vehicles' cruise speeds, vehicle's engine type. To investigate their influences, methods to estimate emission rates are needed.

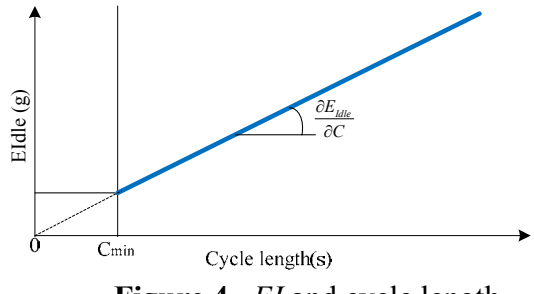


Figure 4. *EI* and cycle length.

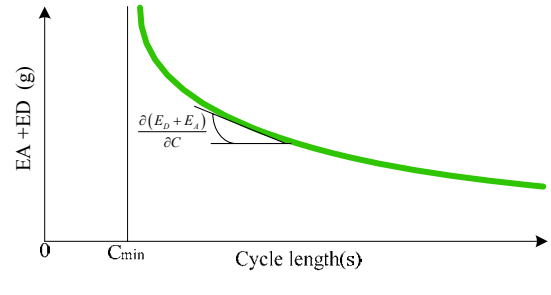


Figure 5. E_D, E_A and cycle length.

Ratio of "marginal benefit" to "marginal cost" is defined as *k* in Eq. (11).

$$k = \left| \frac{\partial \left(E_D + E_A \right)}{\partial C} \right| / \left| \frac{\partial E_I}{\partial C} \right| \tag{11}$$

Methods to estimate emission rate. Different from other performance measures, it is very difficult to directly measure emission amount from all vehicles' exhaust pipes. Hence, emission of a traffic flow is estimated by sum of emission amount from individual vehicles as shown in Eq. (1). For a single vehicle which experiences a stop before stop line, the produced amount of emission can be expressed in Eq. (12).

$$f = f_1 + f_2 + f_3 + f_4 \tag{12}$$

where *f*: total emission (g), f_1, f_2, f_3, f_4: emission (g) from cruise, acceleration, deceleration and idle modes respectively.

Oneyama, et al. (2001) developed an empirical method for NO_x estimation by investigating relationship between heavy vehicle's engine condition and emission. As expressed in Equations (12-1), (12-2), (12-3) and (12-4), this method models emission rates based on vehicle's driving modes considering change of spot speed. Parameters in equations have been calibrated using empirical data measured by the real running pattern on chassis dynamometers.

$$Cruise: \frac{df_1}{dt} = c_1 v + c_2 v^3 + c_{4r+} \tag{12-1}$$

$$Acceleration: \frac{df_2}{dt} = c_1 v + c_2 v^3 + c_{3a} a v + c_{3b} a + c_{4r+} \tag{12-2}$$

$$Deceleration: \frac{df_3}{dt} = c_{4r-} \tag{12-3}$$

$$Idle: \frac{df_4}{dt} = c_{4ri} \tag{12-4}$$

where *v*: spot speed (m/s), *a*: acceleration rate (m/s^2), and c_1, c_2, c_{4r+}, c_{3a}, c_{3b}, c_{4r-}, c_{4ri}: general parameters.

Similarly, Kato, et al. (2002) proposed an empirical method for NO_x estimation with regard to small passenger vehicles. As expressed in Eqs. (13-1), (13-2), (13-3) and (13-4), this method models emission rates based on vehicle's driving modes considering change of spot speed. Parameters in equations have also been calibrated using empirical data measured by the real running pattern on chassis dynamometers.

$$Cruise: \frac{df_1}{dt} = r_c \tag{12-5}$$

$$Acceleration: \frac{df_2}{dt} = p\alpha + qv + r_a \tag{12-6}$$

$$Deceleration: \frac{df_3}{dt} = r_d \tag{12-7}$$

$$Idle: \frac{df_4}{dt} = r_d \tag{12-8}$$

where v: spot speed (m/s), a: acceleration rate (m/s^2), and r_c, p, q, r_a, r_d: general parameters.

Using these methods, the values of k can be estimated to indentify conditions where significant emission reduction can be obtained under $C_{emission}$,

Impact of Vehicles' Cruise Speeds. Emission from one deceleration and acceleration is strongly associated with vehicle's initial cruise speed when approaching intersections.

Table 1. Cruise speed and *Fe*.

Vehicle type	Passenger vehicles			Diesel heavy vehicles		
Cruise speed(km/h)	One deceleration and acceleration(g)	1 second idle(g)	Fe	One deceleration and acceleration(g)	1 second idle(g)	Fe
20	0.1095	0.0018	61	0.1567	0.0104	15
30	0.181	0.0018	101	0.2962	0.0104	28
40	0.2637	0.0018	147	0.4765	0.0104	45
50	0.3575	0.0018	199	0.6975	0.0104	67
60	0.4626	0.0018	258	0.9593	0.0104	92
70	0.5789	0.0018	323	1.2618	0.0104	121

Table 2. Fe_{mix}.

Heavy vehicle percentage	0%	5%	10%	15%	20%	25%	30%	35%	40%
Fe_{mix}	147	123	107	96	87	80	75	70	67

Impact of Vehicles' Cruise Speeds. Emission from one deceleration and acceleration is strongly associated with vehicle's initial cruise speed when approaching intersections.

Fe for a single vehicle is calculated using Eqs. (12-2), (12-3), (12-4), (12-6), (12-7), (12-8) and listed in Table 1. Emission type is NOx, and parameter values are in appendix.

For both passenger vehicle and heavy vehicle, significant changes of *Fe* can be shown with regard to cruise speed. For high cruise speed, high values of k mean significant emission reduction with imposing certain increase delay on drivers. Under this condition, application of $C_{emission}$ is recommend.

Impact of Heavy Vehicle Percentage. Under the same driving conditions, heavy vehicle will produce much more amount of emission, e.g. NO_x than small passenger vehicle as shown in above Table 1. For intersections nearby industry zones or ports, traffic fleet is characterized of a high heavy vehicle percentage.

One interesting feature shown in Table 1 is that on one hand, emission rates of heavy vehicle are higher than that of small passenger vehicle. On the other hand, *Fe* for heavy vehicle is lower than that of small passenger vehicle. For a traffic fleet with higher heavy vehicle percentage, this feature makes its ratio of "marginal benefit" to "marginal cost" comparably lower.

To reflect this impact, one concept "mixed emission rate" is declared. *Fe* of "mixed emission rate" is calculated using the following Eq. (13).

$$\mathrm{Fe}_{mix} = \frac{\rho(e_d + e_a)_{HV} + (1-\rho)(e_d + e_a)_{PC}}{\rho(\frac{de_i}{dt})_{HV} + (1-\rho)(\frac{de_i}{dt})_{PC}} \tag{13}$$

where
ρ is percentage of heavy vehicle(%),
HV and *PC* stand for heavy vehicle and small passenger vehicle respectively.
Then for given cruise speed, e.g. 40km/h, mixed *Fe* is calculated and shown in Table 2.
In Table 2, one important finding is that Fe_{mix} decreases with increase of percentage of heavy vehicle. Also under this condition, obvious emission reduction can be obtained by applying $C_{emission}$.

Conclusions and recommendations

This paper explores fundamental issues of cycle length optimization from the viewpoint of emission. The exploring includes the following procedures. First performance index of the emission based policy is established. Second emission of the whole intersection is modeled as a function of cycle length. Consequently $C_{emission}$ is obtained. Then, regarding application of $C_{emission}$, "marginal cost" and "marginal benefit" are adopted. Finally it is recommended to be applied on conditions with high cruise speed or high heavy vehicle percentage. This study can serve as a basis of designing cycle length settings from the viewpoint of emission.

For future work, traffic control parameters will be optimized regarding various demand distributions, arrival patterns and new engine development.

References

Akcelik, R. (1982). "New approximate expressions for delay, stop rate and queue length at isolated signals." *In: Proc. Int. Conf. On Road Traffic Signalling, Institute of Electrical Engineers, London*, 9–13.

Hirokazu Kato, et al. (2002). "Emission estimation model based on vehicle driving mode for the purpose of environmental evaluation of improved traffic measures." *in JSCE 30th,* 135-141.

Kawashima, H, Tsubota, Y. and Watanabe, K. (2001). "A study on the relationship between micro adaptive signal control and emission reduction." *21st Annual Meeting of Traffic Engineering Research,* 117–120.

Li, X. et al. (2004). "Signal timing of intersections using integrated optimization of traffic quality, emissions and fuel consumption.", *In Transportation Research Part D 9*, 401-407.

Oneyama, H, Oguchi, T. and Kuwahara, M. (2001). "Estimation Model of Vehicle Emission Considering Variation of Running Speed." *Proceeding for the Eastern Asian Society for Transportation Studies*.

Tsai-Yun Liao et al(1996). "Optimal traffic signal strategy for fuel consumption and emissions control at signalized intersections." *European Transport Conference*

Appendix:

Parameters for emission estimation methods

Oneyama method			Kato method		
Parameter	Value	t value	Parameter	Value	t value
c_1	7.22E-04	25.59	p	1.16E-03	178
c_2	2.31E-06	53.79	q	2.99E-04	5.9
c_{3a}	6.19E-03	97.03	r_c	6.93E-03	
c_{3b}	3.98E-03	35.77	r_a	1.79E-03	39.2
c_{4i}	1.04E-02	39.09	r_d	3.58E-03	
c_{4r+}	3.39E-03	28.54			
c_{4r-}	7.64E-03	7.02			
R^2	0.995		R^2	0.598	
Sample size	2041		Sample size	3657	

Influence of the Number of Through Lane Groups on Saturation Flow Rates at Signalized Intersections

Wanchao Zhang[1], Yang Wang and Xiaokuan Yang

[1] Transportation Research Center, Beijing University of Technology, Beijing, China; PH: (8610) 6739-1678; Email: zhangwanchao2008@emails.bjut.edu.cn

Abstract

Multi-through lanes have been used to reduce vehicle delays and queue lengths at signalized intersections with relatively high through movement demands. Limited research has been conducted to study the through lane group capacity issues and how they relate to the number of lanes in the through lane groups. This study analyzed the influence of the number of lanes in the through lane group using 11 sites in Beijing. The one-way analysis of variance (ANOVA) test was used to determine statistical significance of the number of lanes' influence on saturation flow rates. In addition, regression analysis was conducted to determine the relationship between the number of lanes in the through lane group and the saturation flow rates. The results showed that the influence of the number of lanes in the through lane group on saturation flow rate was remarkable in Beijing. The relationship between saturation flow rate and the number of lanes in the through lane group was a quadratic function. At the end of paper, the default value of saturation flow rate and the adjustment factor for the number of lanes in the through lane group was given based on the traffic conditions in Beijing.

Key words: saturation flow rate; number of lane; adjustment factor; signalized intersection

Introduction

Signalized intersections are key elements of urban traffic networks. They are often

bottlenecks where delay and traffic congestion mostly occur. Therefore, signal timings and intersection channelization based on traffic conditions are most important processes during the intersection design. As a result, how to determine the saturation flow rates for each lane group has been the key to intersection capacity analysis.

The HCM2000 has defined the capacity clearly: (1) The capacity of a facility is the maximum hourly rate at which persons or vehicles reasonably can be expected to traverse a point or a uniform section of a lane or roadway during a given time period under prevailing roadway, traffic, and control conditions. Reasonable expectancy is the basis for defining capacity. That is, the stated capacity for a given facility is a flow rate that can be achieved repeatedly for peak periods of sufficient demand. Capacity is not the absolute maximum flow rate observed on such a facility. Driver characteristics vary from region to region, and the absolute maximum flow rate can vary from day to day and from location to location. (2) The saturation flow rate represents the number of vehicles per hour per lane that can pass through a signalized intersection if the green signal was available for the full hour, the flow of vehicles was never halted, and there were no large headways. According to the definition of capacity and saturation flow rate, the capacity of a facility was based on the saturation flow rate, and the definition was also the guideline of data collection when we studied the saturation flow rate.

The HCM2000 segmented the intersection into lane groups that consider both the geometry of the intersection and the distribution of traffic movements. A saturation flow rate for each lane group is computed according to Eq. 1 (HCM, 2000).

$$s = s_o N f_w f_{HV} f_g f_p f_{bb} f_a f_{LU} f_{LT} f_{RT} f_{Lpb} f_{Rpb} \quad (1)$$

s — saturation flow rate for subject lane group, expressed as a total for all lanes in lane group (veh/h);

s_o — base saturation flow rate per lane (pc/h/ln);

N — number of lanes in lane group;

f_* — adjustment factor for the variables′ influence on saturation flow rate;

HCM2000 supposed that the saturation flow rate for each lane group was not changed along with the number of lanes in the lane group (HCM 2000). Therefore, this assumption needs to be studied to approve that weather the assumption was suited for the traffic condition in Beijing or not. There were two questions: 1) Was the saturation flow rate weather effected by the number of lanes in lane group? 2) Was the impact negative or positive? The objective of this paper was to study the influence of the number of lanes in the lane group on the saturation flow rate and to conduct the quantitative analysis for this influence.

Literature Review

HCM 2000 assumed that the relationship between the saturation flow rate for each lane group and the number of lanes in lane group was a simple multiple function. This model ignored the probability of the influence of the number of lanes in the lane group on the saturation flow rate. There were many research fruits on the model of saturation flow rate in recent years. Ackeret (1996) studied the operational characteristics of single and multiple left-turn lanes and compared them to adjacent through-lanes on the same intersection approach. The conclusion of this paper supports the HCM 1994 suggestion that the ideal through-lane saturation flow rate is 1900 (pc/h/ln). However, the field observations suggest that the left-turn factor for single left-turn lanes with protected left-turn signal phasing should be 1.0, and dual left-turn lanes should be a minimum of 0.98 per lane. The results represent a significant increase over the suggested left-turn factor of 0.95 as recommended within the HCM 1994. The triple left-turn lane observations were found to continue to support an average left-turn factor of 0.95. Sando and Moses (2009) analyzed the influence of a number of geometric factors found at 15 triple left-turn lane sites in Florida on saturation flow, lane usage, and lane utilization. The results showed that triple left-turn lanes on downgrades and angle of turn less than 90° were the two factors that mostly contributed to the high saturation flow while triple left-turn lanes located on one-way streets and on approaches that are curved had the lowest saturation flow. It was also found that street type significantly influenced saturation flow results when combined with other variables. Courage et al. (2002) studied the triple left-turn lanes at signalized intersections. The results of these analyses suggested that triple left turns can produce substantial operational benefits, however, an increase in crash rates was noted at some locations. It was also observed that existing traffic models do not treat triple left turns explicitly and can therefore only offer an approximation of their operation.

At present, the research on the saturation flow rate of the lane group were mainly focused on the multiple turn lanes. There was limited study on the influence of the number of lanes in through lane group on the saturation flow rate. Therefore, the aim of this paper is to investigate the relationship between the number of lanes in the through lane group and the saturation flow rate in Beijing.

Methodology and Data Analysis

The research objective of this paper was only the through lane group. Therefore, the researchers only studied the influence of the through lane number on the saturation flow rate at signalized intersections. The procedure of this research includes four steps: (1) analyze the influence of lane width on the saturation flow rate in Beijing;

(2) confirm weather considered the lane width or not based on the result of step (1) when conducting the study on the impact of the number of lanes in through lane group on the saturation flow rate in Beijing; (3) data collection conducted at 11 typical signalized intersections in Beijing; (4) The one-way analysis of variance (ANOVA) test was used to determine statistical significance of the number of lanes' influence on saturation flow rate. In addition, regression analysis was conducted to determine the relationship between the number of lanes in through lane group and the saturation flow rate.

Data Collection. According to the definition of the basic condition given by the HCM 2000, 11 typical signalize intersections in Beijing's fifth ring road were selected. (Figure 1) There were eight typical signalized intersections which through lane groups include only one through lane and three typical signalized intersections which through lane groups include multiple through lanes, such as 2, 3, and 4. (Tables 1 and 2).

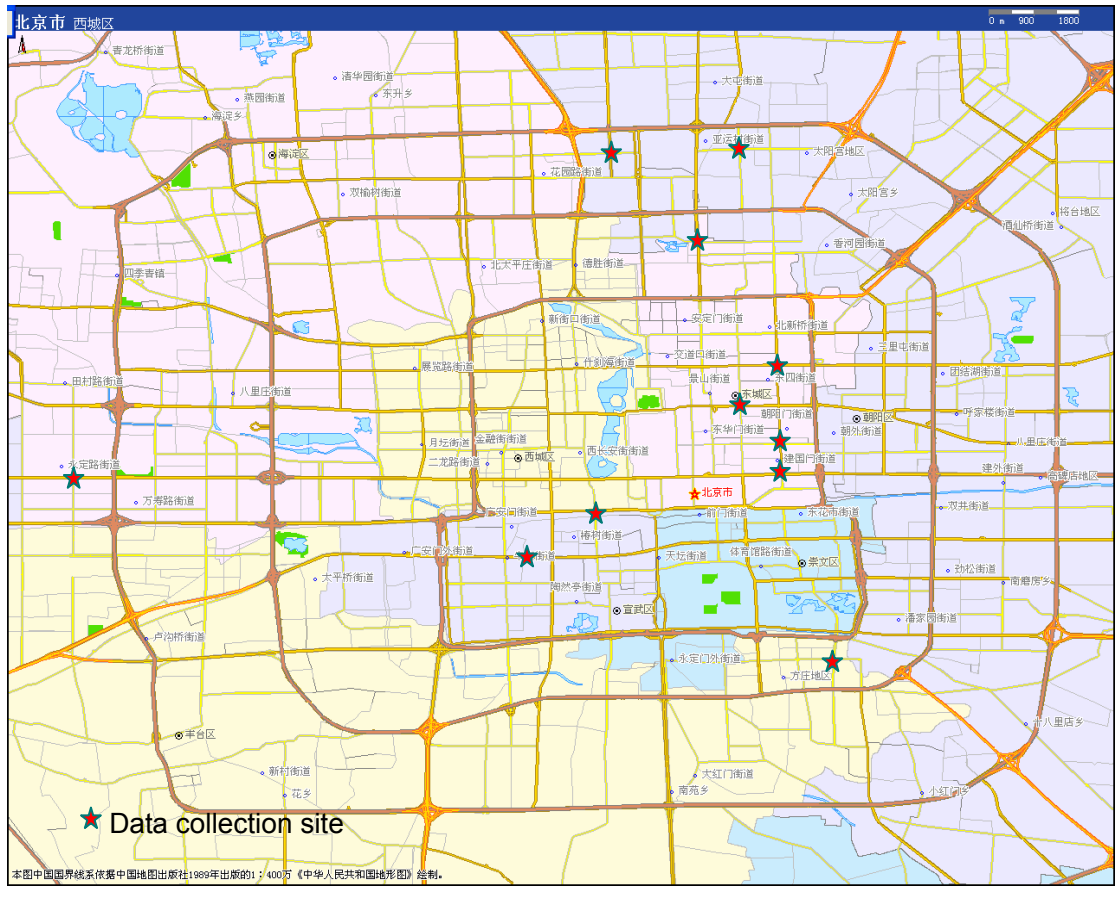


Figure 1. Figure of data collection Site Distribution.

Table 1. Number of typical signalize intersection with one through lane.

Through lane width (m)	Number of study site
3.5	2
3.25	2
3.0	2
2.85	1
2.65	1

Table 2. Number of typical signalize intersection with multiple through lane.

Number of lanes in through lane group	Number of study site
2	1
3	1
4	1

This paper used video cameras to collect the saturation flow headway. According to HCM 2000, vehicles are recorded when their front axles cross the stop line. The measurement starts at the beginning of the green time or when the front axle of the first vehicle in the queue passes the stop line. Saturation flow, however, is calculated only from the headways after the fourth vehicle in queue passes the stop line to eliminate the impact of start delay on the saturation flow rate. Data collected at 11 typical signalized intersection in Beijing has shown that the value of headways were stable after the third vehicle in queue passes the stop line (See Figure 2.)

Data Processing. The quality of data was the key element to ensure the correctness of the investigation conclusion. Errors could not be avoided during the course of the data collection and process. According to the characteristic and error causes, the error could be classified into three types, such as random error, systematic error, and mistake error. There were three measurements to judge the quality of data. Precision reflects the size of random error; trueness reflects the size of systematic error and accuracy reflects the combined effects of random error and systematic error. In order to enhance the quality of data, the exceptional data that exist in the samples should be eliminated before the data analysis. There were three methods to identify the exceptional data, such as Paŭta test, Grubbs test, Dixon test, Chauvenet test, Nair test, and T-test. This paper adopts the Paŭta test to identify the exceptional data because of the large number of samples.

Impact Factor Analysis. The analysis of variance (ANOVA) test was conducted to determine statistical significance of the variable (lane width and number of lanes) influence on saturation flow rate. In addition, regression analysis was conducted to

determine the relationship between the variable (lane width and number of lanes) in the through lane group and the saturation flow rate.

Influence of Lane Width on the Saturation Flow Rate

The effect of lane width on average saturation flow rate was estimated by means of a one-way analysis of variance (ANOVA) for the five lane width categories described in Table 1. The ANOVA was performed for the average saturation flow rates. All analyses were performed at the 95-percent confidence level. Lane width (category) had a statistically significant effect on average saturation flow (p-value < 0.01). (Tables 3 and 4).

According to the result of one-way analysis of variance for the five lane width categories, lane width (category) had a statistically significant effect on average saturation flow. Therefore, lane width must be considered when conducting the study on the impact of the number of lanes in the through lane group on the saturation flow rate in Beijing. In other words, the data used in the research on the influence of number of lanes in the through lane group must be collected from the approach where the through lane width must be 3.5 m.

Based on the above result, regression analysis was conducted to determine the relationship between the lane width and the saturation flow rate. The analysis result shows that there was a good linear relationship between saturation flow rate and the lane width (Figure 2). Therefore, the lane width adjustment factor for the saturation flow rate could be obtained (Table 5).
The linear relationship between saturation flow rate and lane width was described as:

$$y = -0.4391x + 3.6009 \quad R^2 = 0.9973$$

y – headway，s

x – lane width，m

Table 3. Saturation flow rate statistics by lane width category.

Lane width category	N	Sum	Mean	Standard error
3.5	206	424.781	2.062044	0.223607
3.25	143	311.44	2.177902	0.461817
3	138	313.985	2.275254	0.418462
2.85	157	370.588	2.360433	0.379697
2.65	198	481.659	2.432621	0.327116

Table 4. Result of one-way analysis of variance.

Variance source	SS	df	MS	F	P-value	F crit
Between groups	7.845142	4	1.961285	5.590853	0.000203	2.387541
Within group	200.3082	837	0.350803			
Total	208.1534	841				

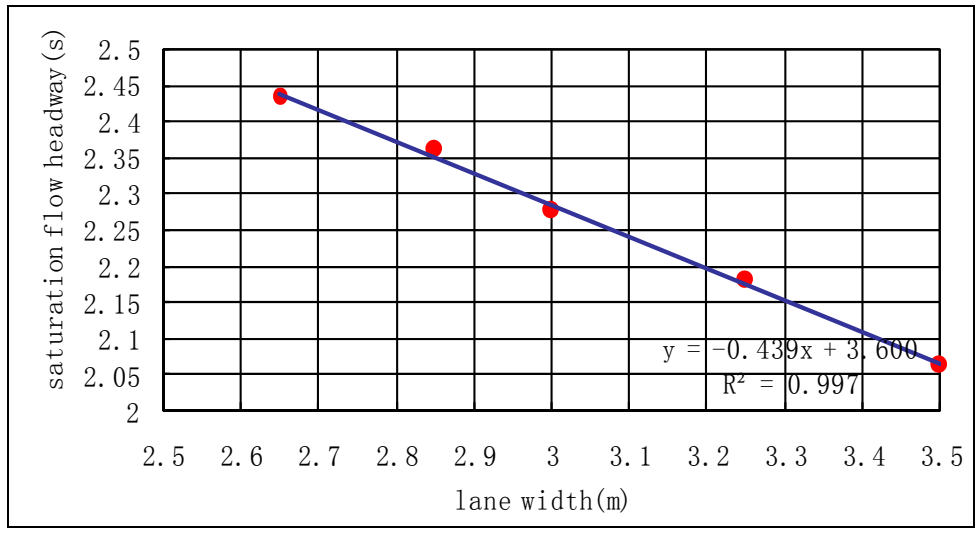


Figure 2. Linear relationship between saturation flow rate and the lane width.

Table 5. Lane width adjustment factor for the saturation flow rate.

Lane width (m)	Saturation flow rate (pc/h/ln)	Adjustment factor f_w
3.5	1746	1
3.25	1653	0.95
3	1582	0.9
2.85	1525	0.87
2.65	1480	0.85

Influence of the Number of Lanes in the Through Lane Group on the Saturation Flow Rate

The study on the influence of the number of lanes in the through lane group on the saturation flow rate should choose the same lane width (3.5 m) to collect the data, because of lane width has a significant effect on saturation flow rate. Three typical signalized intersections which width of lane in the through lane group was 3.5 m were selected (Table 2).

Table 6. Saturation flow rate statistics by lane count category.

Lane count category	N	Sum	Mean	Standard error
3.5-1	206	424.781	2.062044	0.223607
3.5-2	123	250.389	2.035683	0.3221
3.5-3	111	221.719	1.997468	0.443306
3.5-4	132	246.576	1.868	0.143315

Table 7. Result of one-way analysis of variance.

Variance source	SS	df	MS	F	P-value	F crit
Between groups	3.249772	3	1.083257	4.030103	0.007473	2.620593
Within group	152.6736	568	0.268791			
Total	155.9233	571				

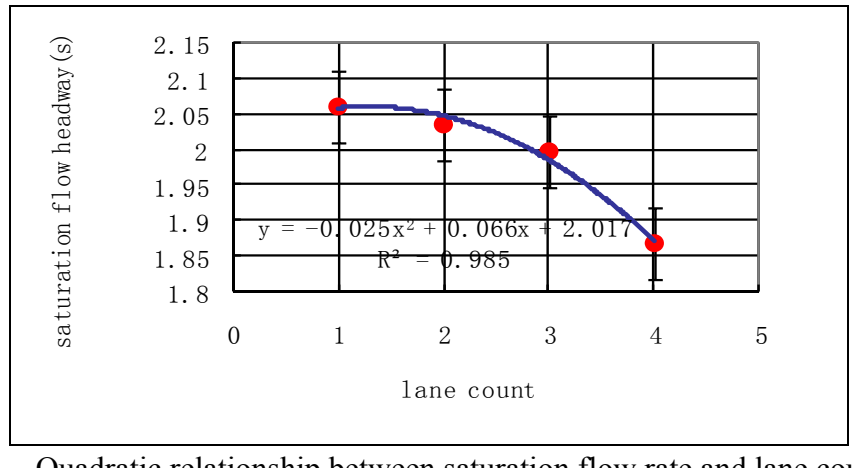


Figure 3. Quadratic relationship between saturation flow rate and lane count.

The effect of number of lanes in the through lane group on average saturation flow rate was estimated by means of a one-way analysis of variance (ANOVA) for the four lane count categories described above (Table 2). The ANOVA was performed for the average saturation flow rate. All analyses were performed at the 95-percent confidence level. The number of lane (category) had a statistically significant effect on average saturation flow (p-value < 0.01) (Tables 6, Table 7).

According to the result of one-way analysis of variance for the four lane count categories, the number of lanes in the through lane group had a statistically significant effect on average saturation flow rate (p-value < 0.01). Based on the

above result, regression analysis was conducted to determine the relationship between the number of lanes and the saturation flow rate. The analysis result shows that there was a good quadratic relationship between saturation flow rate and the number of lanes (Figure 3). Therefore, the number of lanes in the through lane group adjustment factor for the saturation flow rate could be obtained (Table 8).

The quadratic relationship between saturation flow rate and lane count in the through lane group was described as:

$$y = -0.0258x^2 + 0.0669x + 2.017 \qquad R^2 = 0.9858$$

y – headway，s

x – number of lanes in through lane group

Table 8. Lane number in through lane group adjustment factor.

Lane count	Saturation flow rate（pc/h/ln）	Adjustment factor f_n
4	1927	1.07
3	1802	1.02
2	1768	1.01
1	1746	1

Table 9. Comparison of saturation flow rate values of this research to HCM.

Lane width(m)	Adjusted saturation flow rate（pc/h/ln）	
	HCM 2000	Current research
3.5	1879	1746
3.25	1826	1653
3	1773	1582
2.85	1742	1525
2.65	1699	1480

Table 10. Comparison of saturation flow rate values from this research to HCM.

The number of lanes in through lane group	Adjusted saturation flow rate（pc/h/ln）	
	HCM 2000	Current research
4	7516	7708
3	5637	5409
2	3758	3536
1	1879	1746

Conclusions

Table 9 compares the saturation flow rate estimates based on HCM 2000 procedures to those measured in the current research. The table shows that the measured saturation flow rate values are generally lower than those obtained from HCM 2000 procedures. Table 10 shows that although the measured signal lane saturation flow rate values are lower than those obtained from HCM 2000 procedures, the lane group saturation flow rate are higher than those obtained from HCM 2000 procedures when the number of the through lane group equals 4. Because of HCM 2000 ignored the influence of the number of lanes in the bthrough lane group. Also, these differences between the HCM 2000 adjusted saturation flow rates and the measured saturation flow rates may be partially explained by the difference of vehicle characteristics and driver behavior between the United States and China.

With respect to the basic condition of signalized intersection, this paper defines the lane width 3.5 m as the basic condition because the maximum through lane width was 3.5 m in Beijing. Therefore, more study needed in other cities in China where lane width was higher than 3.5 m.

Acknowledgment

The work reported in this paper was performed as part of Project of Beijing Municipal Institute of City Planning & Design Study on The Capacity of Signalize Intersection in Beijing. The findings and conclusions of this paper are those of the authors and do not necessarily reflect the views of NCHRP or its sponsors.

References

Ackeret, K. W. (1996). *Analysis of Saturation Flows at Single, Dual and Triple Left-Turn Lanes*, Ph.D. dissertation, Univ. of Nevada, Las Vegas.

Courage, K., Stephens, B., Gan, A. and Willis, M. (2002). *Triple Left-Turn Lanes at Signalized Intersections*, Final Report Prepared For The Florida Department of Transportation.

Sando, T. and Moses, R. (2009). "Influence of intersection geometrics on the operation of triple left-turn lanes." *Journal of Transportation Engineering,* ASCE.

Transportation Research Board. *Highway Capacity Manual* 2000, National Research Council, Washington, D.C.

The Impact Analysis of Intersection Sight Distance on Vehicle Speed

Miaomiao Liu[1], Guangquan Lu[2] and Yang Li[3]

[1] Graduate Student, [2] Associate Professor, [3] Undergraduate Student, School of Transportation Science and Engineering, Beihang University, XueYuan Road No.37, HaiDian District,Beijing, China 100191; Email: liumiao-0605@163.com

Abstract

Sight distance and vehicle speed are closely related to traffic safety, and good sight distance not only enable driver to judge the road traffic environment correctly, determine the correct driving behavior, but also determines effective operating time of the driving behavior. This paper made an in-depth analysis to intersection sight distance and turning characteristics, and carried on a statistical analysis to relationships among stopping distance, landscape static sight distance and the difference between them with vehicle speed, which took Mudanyuan Intersection of Haidian District in Beijing as an example and used mathematical models of the landscape static sight distance and the stopping distance. From the experiments we find that intersection sight distance have a strong correlation with vehicle velocity. Intersection static sight distance and sight distance difference have positive correlations with turning right vehicle velocity, while have negative correlations with direct vehicle velocity. Besides, it can be seen by combining the existing data that the greater static sight distance and sight distance difference are, the higher the maximum speed which is allowed and enable ensure the safety of traveling and turning right vehicle velocity are, but direct vehicle speed is lower, which has certain relations with driver psychology. Direct driving in the situation of smaller static sight distance and sight distance difference, drivers will be more anxious and then they are more inclined to speed up to avoid transportation conflicts and accidents; but if static sight distance and sight distance difference are greater, drivers will control accelerator and brake calmly, and the moving velocity will be lower.

Key words: sight distance; turning characteristics; mathematical model; statistical analysis

Introduction

The sight distance of road providing for driver is one of the most important safety factors. Good sight distance not only enable driver to judge road traffic environment correctly, and determine correct driving behavior, but also determines effective operating time of the driving behavior.

There are lots of researches on intersection sight distance, and mostly theoretical studies: intersection sight distance based on the laws of motions and the equations are flexible to changes, especially assumptions; thus traffic engineers obtain results from clear scientific equations, which can be easily defended (Awadallah, 2009). Sight obstruction due to the opposite turning vehicles may contribute to significant increments of the critical gap for both left-turn and U-turn drivers. With the sight distance problem, the drivers' left-turn follow-up time is also significantly increased compared to those without the problem (Yan and Radwan, 2007). The sight distance along the major road for a passenger car at a stop-controlled intersection be based on a distance equal to 7.5 s of travel time at the design speed of the major road. Longer sight distances are recommended for minor-road approaches that have sufficient truck volumes to warrant consideration of a truck as the design vehicle (Harwood et al.) A skeleton of the road, road geometry alignment design of horizontal, vertical section and cross-section and geometry alignment combination are permanent. The irrational road geometry alignment design and inconsequent geometry alignment combinations may evoke traffic accidents (Liu, 2003). A piece-wise linear clear zone boundary can be determined with ease to a considerable level of accuracy (Lovell et al., 2001).

This paper made an in-depth analysis to intersection sight distance and turning characteristics, and carried on a statistical analysis to relationships among stopping distance, landscape static sight distance and the difference between them with vehicle speed, which took Mudanyuan Intersection of Haidian District in Beijing as an example and used mathematical models of the landscape static sight distance and the stopping distance.

The Analysis of Intersection Sight Distance and Turning Characteristics

Intersection Sight Distance. The line of sight envelope curve can be used to guarantee sight distance (Figure 1).

The stopping sight distance is the shortest distance needed between driver seeing the obstacle and being able to stop in front of it. It is stipulated that stopping sight distance is driver's moving sight distance. It needs to determine landscape static sight

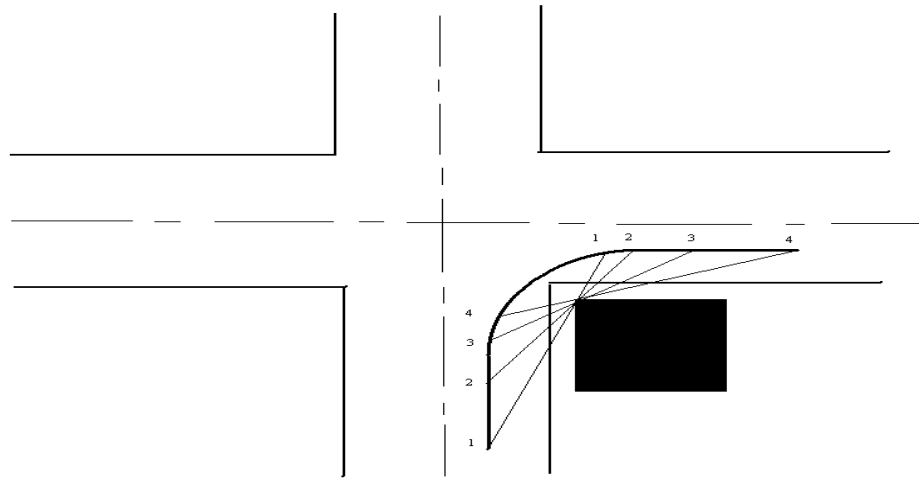


Figure 1. Line of sight envelope curve.

distance that landscape provides to driver in order to determine the visual impact of road traffic safety. When the height of landscape is higher than driver's line of sight, moving sight distance completely under effects of static sight distance; while the height of landscape is below a certain driver's line of sight, the driver can get his expected length of sight distance through the top of landscape, no matter the moving sight distance is greater or less than the landscape static sight distance.

Turning Characteristics. Before the driver is about to enter the intersection, the main concern is safety-related factors, such as turning radius, situation of opposite incoming vehicle, and thus adjust vehicle speed to pass the intersection safely. Drivers must look at inside landscape of intersections to obtain such information. The judge on the situation of intersection road is a continual movement process along with the constant moving of vehicle.

In the process of obtaining road information, driver's line of sight constantly tangents to inside landscape and forms continuous sight distance. When objects inside of bend obscure driver's sight, the driver will be uncertain about the preceding road state and reduce vehicle speed, moreover, unceasingly take brake measure. When the bend sight is good, the driver will reduce scope of vehicle deceleration control and speed change is little. A good sight can improve intersection traffic efficiency and ensure vehicle to pass through safely by the maximum speed.

The Sight Distance Model

Method Analysis. When driver makes a turn on certain radius curve and inside landscape blocks driver's sight, the driver will take certain measures to pass through safely. In the entire curve length scope, the most dangerous section is the first half of

the bend. It is discovered in the investigation and study process that the number of accidents occurred in small radius bend is clearly higher than that on large radius bend, therefore we take the small radius curve as model foundation and divide bend accidents into a vehicle accident and two vehicles collision accident (Lu and Tang, 2001). When the vehicle goes on the curve which inside landscape blocks sight, the landscape will have an impact on road driving.

Model Establishment. We selected the rectangular coordinate system as platform of model establishment and put four corners, landscape and vehicles into the coordinate system separately based on certain relations. Landscape impacted on speed of the intersection mainly referred to the inside landscape affected driver's sight. Figure 2 is the stopping sight distance schematic drawing.

Stopping sight distance formula:

$$s_t = \frac{Vt}{3.6} + \frac{KV^2}{254\Phi} + S_0 \tag{1}$$

In the formula: t—reaction time; V—vehicles' speed, km/h; K—brake service factor of the first vehicle, generally between 1.2-1.4; φ—longitudinal frictional coefficient, 0.6; S—safe distance, 5-10 m.

The formation of landscape static distance is determined together by the landscape position and the vehicle position (Figure 3).

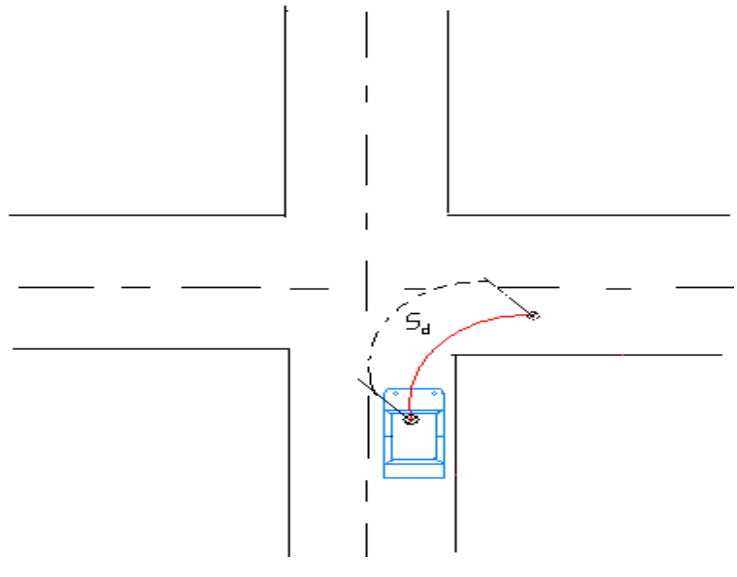


Figure 2. Stopping sight distance schematic drawing.

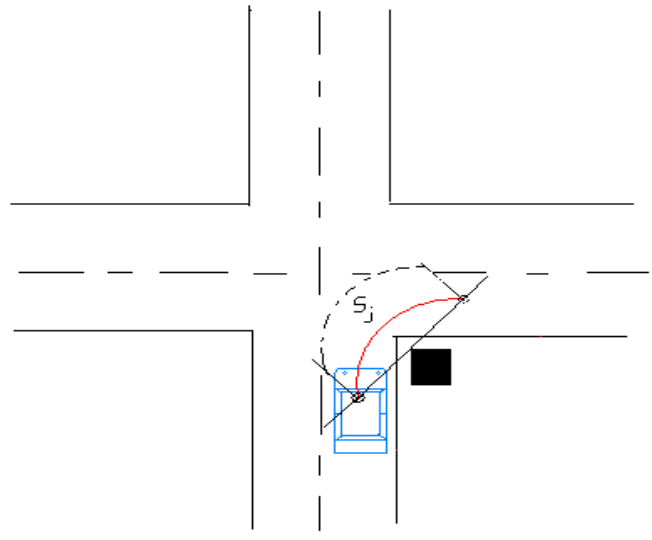


Figure 3. Landscape sight distance schematic drawing.

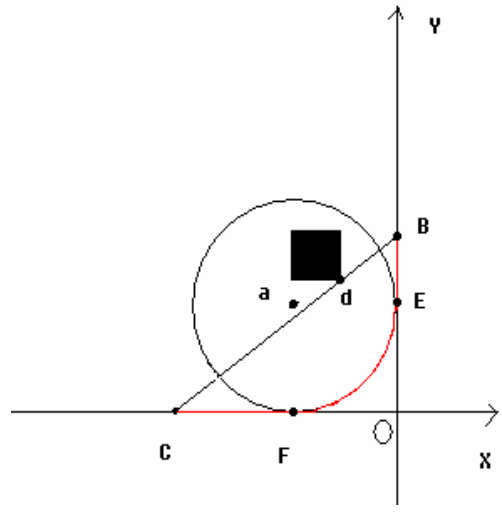


Figure 4. Calculation schematic drawing of landscape static sight distance.

The computation method of landscape static apparent distance is shown in Figure 4.

As is shown in Figure 4, B is driver's position, $d(m, n)$ is point covered driver's vision, a is the center of rounding path and $C(x, 0)$ is the end point of static apparent distance formed by the pilot B and the building point d. The length of red streak is landscape static apparent distance, and vehicle' extension rounding path radius R is 20 m.

The formula of landscape static apparent distance S_j is as follows:

$$S_j = BE + CF + EF = (y_0 - R) + |x_0| - R + \frac{\pi R}{2} \tag{2}$$

Because the straight line equation of AB is:

$$y - y_0 = \frac{(n - y_0)}{m} x \tag{3}$$

and

$$x_0 = \frac{my_0}{y_0 - n} (x_0 < 0) \tag{4}$$

$$S_j = y_0 + \frac{my_0}{n - y_0} + (\frac{\pi}{2} - 2)\ R \tag{5}$$

Derivate of S_j and make it equal to 0, it can be known when $y_0 = n + \sqrt{-mn}$, S_j obtains the minimum. Namely the landscape static apparent distance of this position ($y_0 = n + \sqrt{-mn}$) is smallest and safe speed allowed is lowest.

Example of Mudanyuan Intersection in Beijing

The Introduction of Mudanyuan Intersection. Through many intersections' investigation, Mudanyuan Intersection of Haidian District in Beijing had been taken as the example of study, which each direction's static apparent distance was bigger than the stopping sight distance. The landscape which needed to be studied was as follows: North—newsstand; East—blue structural fence; South—bus stop; West—the height of landscape was low, thus did not do the research. Figure 5 is the street intersection schematic drawing.

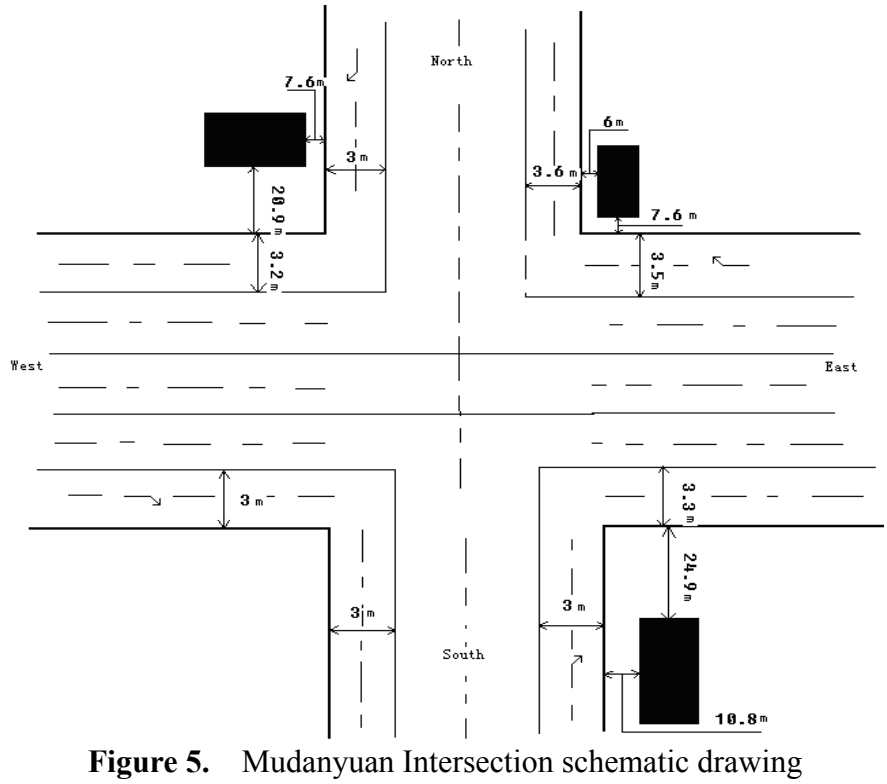


Figure 5. Mudanyuan Intersection schematic drawing

Speed and Sight Distance of the Intersection. Selected Bushnell Velocity instrument to gather vehicle speed data, smallest landscape static apparent distance spot as survey place and carried on 5 days of data acquisition. Chose three time sections (morning, noon and evening) everyday and selected 12 right-turn vehicles and 12 straight vehicles to carry on the survey (all survey vehicles were selected under the circulation of traffic and non-traffic light influences) in four directions of each time section. Besides, it could eliminate passage width influence to adopt east and west comparison, north and south comparison.

Statistical Analysis of the Example. Established samples of four directions' right-turn speed and straight speed separately, and then carried on non-parametric test.

Non-parametric test. It could be known by using SPSS software to do non-parametric test that accompanying probabilities of four directions' right-turn speed and straight speed were both 0.000, which smaller than significance level 0.05, therefore rejected zero supposition and considered that four directions' right-turn speed and straight speed both had remarkable difference, namely each direction's speed characteristic had obvious differences. This possibly had relations with static apparent distance and apparent distance difference. Following made the correlation analysis using statistics software.

Turning right speed and sight distance-correlation analysis. Following is the scatter diagram of turning right speed and sight distance:

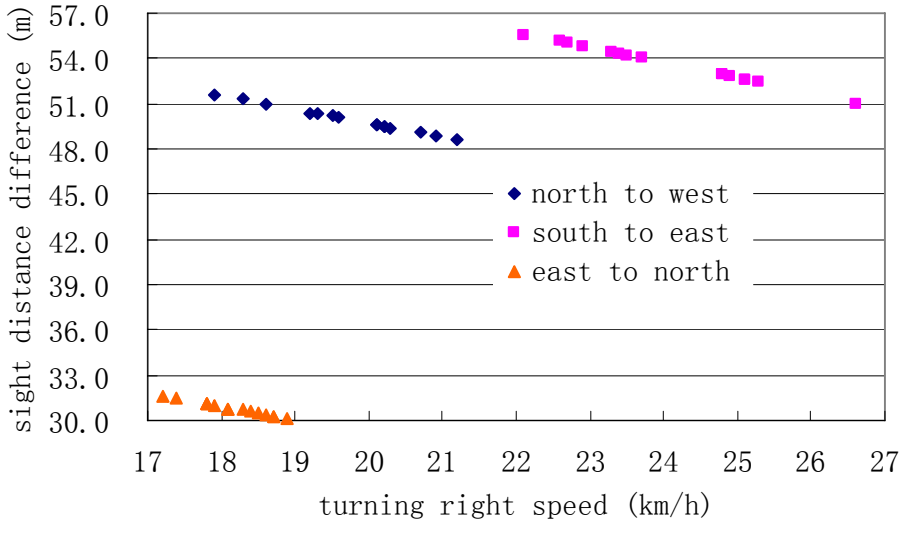


Figure 6. Scatter diagram of turning right speed and sight distance.

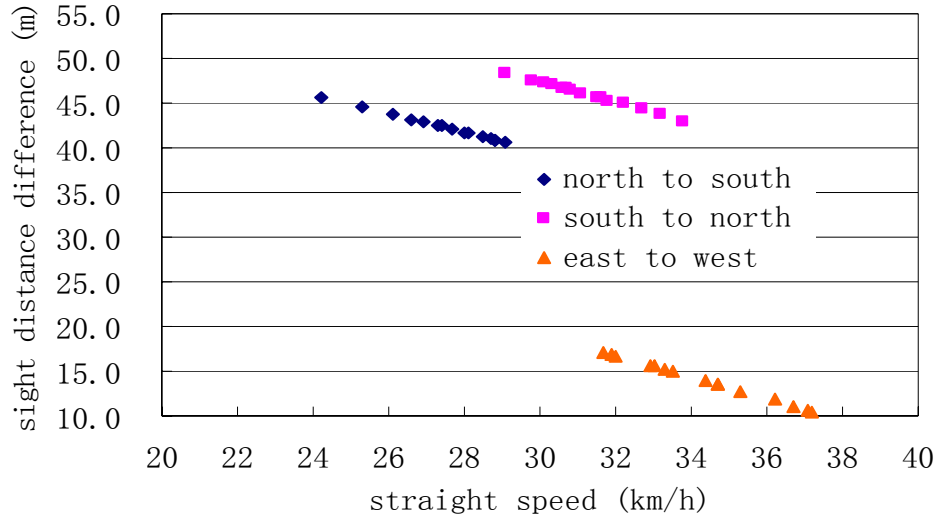


Figure 7. Scatter diagram of straight speed and sight distance.

It was known by using SPSS software that relevant coefficient of intersection right-turn speed and right landscape static apparent distance, apparent distance difference was 0.816, 0.722 respectively. Statistical test's associated probability was smaller than 0.01, namely intersection right-turn speed and right landscape static apparent distance were obviously related and positive correlation, so were the right-turn speed and sight distance difference. The bigger static sight distance and sight distance difference were, the bigger driver's sight scope was when right turning, and the driver might guarantee that changed safely by the high speed, therefore the speed would be bigger.

Straight speed and sight distance-correlation analysis. The scatter diagram of straight speed and sight distance is as follows:

May know using SPSS software analysis: relevant coefficient of intersection straight speed and right landscape static sight distance, sight distance difference was –0.574, –0.728 respectively. Statistical test's associated probability was smaller than 0.01, namely intersection straight speed and right landscape static sight distance were obviously related and negative correlation, so were the straight speed and sight distance difference. Driver could not observe lateral incoming vehicle and pedestrian condition promptly when direct driving in the case of small static sight distance and sight distance difference, which had certain influence to driver's psychology.

It is thought that the smaller right landscape static sight distance and sight distance difference are, the smaller sight scope is when direct driving in green light situation, and driver will have certain tense feeling and favors to travel fast, therefore the speed

will be bigger; Otherwise, driver's vision scope is bigger, whose tense feeling will not be too intense, therefore he or she will control accelerator and brake gently, and the straight speed will be instead smaller. The conclusion waits for further investigation and study.

Conclusions

When running on the road, vehicles' speed change must be gentle and avoid presenting great speed change scope. If driver's sight is covered by the landscape of roadside, his psychology will be affected by each kind of sight distance change, which can lead to moving velocity change.

Street intersection apparent distance and vehicle moving velocity have a strong correlation. The intersection turning right vehicle velocity and static sight distance, distance difference both have positive correlations, but direct vehicle velocity and static sight distance, distance difference both have negative correlations. It can be seen from the existing data that: the bigger right landscape static apparent distance is, the higher right-turn vehicle speed and the bigger right-turn apparent distance difference, at the same time, the maximum speed allowed under the guarantee security's condition is also higher; but the bigger right landscape static apparent distance and distance difference are, the lower straight speed will be, which have certain relations with driver's psychology. In the situation of small static apparent distance and small apparent distance difference, driver will be more anxious in straight and will be favor to pass fast to avoid conflict and accident. When static apparent distance and apparent distance difference is big, driver will control accelerator and brake gently, therefore the moving velocity will be lower.

Acknowledge

This work was partly supported by the National Natural Science Foundation of China (50808008) and Training Scheme for Excellent Researcher in Beijing "Intersection vehicle motion—micro-behavior model and simulation". The authors also wish to thank students from Beihang University who helped doing much data acquisition work.

References

Awadallah, F. (2009). "Intersection sight distance analysis and guidelines." *Transport Policy*, 143-150.

Yan, X. D., and Radwan, E. (2007). "Effect of restricted sight distances on driver behaviours during unprotected left-turn phase at signalized intersections." *Transportation Research Part F*, 10, 330-344.

Harwood, D. W., Mason, J. M., Brydia, R. E. (1999). "Design policies for sight distance at stop-controlled intersections based on gap acceptance." *Transportation Research Part A*, 10, 199-216.

Long, F. Y. (2009). "The sight distance analysis of the overpass road intersection." *Logistics engineering and management*.

Liu, X. D. (2009). *Study on the Cloud Model of operating Speed Based on Road Alignment and Sight Distant*, Jilin University Press.

Lovell, D. J., Jong, J. C., and Chang, P. C. (2001). "Clear zone requirements based on horizontal distance considerations." *Transportation Research Part A*, 35, 391-411.

Shi, N., (2005). *Road Conditions and Traffic Safety*, Tongji University Press.

Zhu, S. Y. and Wang, H. (2006). *Road Safety Evaluation Factor Analysis*, Chongqing Jiaotong University Press.

Liu, D X. (2003). "Linear modular design of the highway operator safety." *Guangxi Communication Science and Technology*.

Lu, F., and Tang, X. F. (2001). "Three-dimensional animated simulation demonstration of dangerous highway sections." *Public Security University Journal*.

Yu, J. Y. and He X. H. (2003). *Statistical Analysis and SPSS Application*, Posts & Telecom Press.

Liu, C. S. (2006). *Automobile Tire and Road Surface Friction Coefficient Studies*.

Wang, J. L. (2006). *Traffic Safety Factors Analysis in Road Lighting Design—Based on the Visual Characteristics of the Pilot Study*, Lighting design.

Intersection Safety Evaluation Model

Shengxue Zhu[1,2], Jian Lu[1]and Gu Wang[1]

[1] School of Transportation, Southeast University, Nanjing 210096, China; E-mail: zhushengxue601@126.com
[2] Faculty of Transportation Engineering, Huaiyin Institute of Technology, Huai'an 223003, China

Abstract

The objective and subjective factors affecting intersection safety is analyzed and the amount of weighted conflict points in a certain potential conflict area is selected as the intersection safety index. The intersection safety evaluation method is established with other affecting factors correction, such as geometric design, safety facilities. The rationality of the method is validated to show that it can analyze the intersection safety conveniently which needn't the traffic accident statistical data and traffic conflict survey.
Key Words: traffic safety; intersection; traffic conflict; safety evaluation

Introduction

Intersection is the key point of the road network, which functions as the joint and hub of road traffic flow. The safety condition of intersections has a great influence on the traffic safety of the road network in the region. In order to improve the traffic safety of the road net, and decrease the number of traffic accidents, it is necessary to reasonably evaluate the intersection safety conditions. At present there are mainly two types of intersection safety evaluation methods: one is direct evaluation which is based on the statistics of traffic accidents, the other is indirect evaluation which is based on traffic conflict technique. Direct evaluation method mainly includes accidents absolute number method, accident rate method, accident ratio quality control method, experience model method, regression model and etc (Zhu et al 2009). However, when these methods applied, the evaluation results' accuracy is affected

because of the long period and inaccuracy of statistical data. The indirect evaluation method based on traffic conflict technique is a relatively new traffic safety evaluation method (Xiang et al 2008), which is mostly applied to evaluate intersection traffic safety by the ratio of conflict number and traffic flow. However, this method also has shortcomings, such as the inaccuracy of the observed traffic conflicts numbers, and the lack of proper division of traffic conflicts according to different traffic conflicts' various effects on traffic safety.

Many studies show that the direct evaluation methods based on traffic accidents statistics and indirect evaluation methods based on traffic conflict technique can evaluate the safety of the intersection. However, with these methods, the researchers cannot give fast and correct evaluation of the intersection traffic safety condition, or they cannot evaluate at all for the lack of accidents statistical data. Therefore, it is necessary to develop an effective intersection safety evaluation method in order to research further on intersection traffic safety improvement measures and to improve road traffic safety. The study establishes intersection safety evaluation model which is formulated by the number of potential conflict points in per unit conflict area and other intersection safety affecting factors such as geometric design, safety facilities.

Intersection Safety Affecting Factors

Intersection safety affecting factors includes macro factors and micro factors (Pan et al 2008) . The macro factors are intersection potential conflicts, traffic flow conditions, factors related to people, environmental conditions and etc. The micro factors include the number of conflict points, traffic volume, vehicle type ratio, traffic law abiding conditions of pedestrians and drivers, intersection longitudinal grade, number of crossed roads, cross angle, visibility distance, number of lanes, width of lanes, turning radius, various traffic signs, markings, signal lights, pavement friction coefficient, pavement types, lightening conditions and etc. The major intersection safety affecting factors are obtained through employing AHP, experts' consultation, and field investigation, which is divided into three types.

Conflict points. The number of any intersection conflict points is constant .There are three types of conflicts (Xiang et al 2008).
C1-Merging conflict: Mainly shows as conflicted vehicles merge from similar directions, may lead to collision between two vehicle heads;
C2-Diverging conflict: Mainly shows as conflicted vehicles diverge from opposite directions, may lead to rear collisions;
C3-Crossing conflict: Mainly shows as conflicted vehicles approach to each other in crossed manner, may lead to collision between vehicle head and vehicle side.

Geometric Design. longitudinal grade, cross angle, visibility distance, lane setting, intersection channelization.

Safety Facilities. Geometric alignment, signs and markings, pavement, lightening conditions.

Safety Evaluation Model of Intersection

Model establishment. According to analysis the intersection safety affecting factors, the intersection safety evaluation model is formulated by combining three components : the number of weighted potential conflict points in per unit area, the influence coefficient of geometric design, the influence coefficient of safety facilities.The intersection comprehensive evaluation model is as follows:

$$R = \frac{100\sum \partial_i * R_i}{S} * f_j * f_s \quad (1)$$

where,

R= comprehensive safety index of the intersection;
∂_i=Weights of the different types conflict points;
R_i=number of the different types conflict points;
S=functional area of the intersection, calculation show in (Lu et al 2009) ;
f_j=influence coefficient of geometric design;
f_s=influence coefficient of safety facilities.

Weights of the different types conflict points. The intersection inner conflict points fundamentally decide whether the intersection is safe or not. Since different types of conflict points vary from the severity, it is necessary to consider not only mummer but also type of conflict points in the process of safety index calculation. Therefore, weights of the different types conflict points can be calculated according to conflict points' severity of different types.

Many studies show that collision accidents caused by diverging conflict points are generally rear collisions; while accidents caused by merging conflict points are generally side collisions and accidents caused by crossing conflict points are generally right angle collision. Therefore, diverging conflict accidents are the least serious one, and merging conflict accidents the medium, and crossing conflict accidents are the most serious. According to conflict points' severity, the Weights of the different types conflict points are obtained, which is shown in Table 1.

Table 1. Weight of different type conflict points.

Types of conflict points	Crossing conflict. points	merging conflict points	Diverging conflict, points
weight	3.5	1.5	1

Table 2. Weighted value of geometric design adjusting coefficient.

Longitudinal grade	Cross angle	Visibility distance	Lane setting	Intersection channelization
0.12	0.20	0.30	0.20	0.18

Influence Coefficient of Geometric Design. Geometric design influencing factors include longitudinal grade, cross angle, visibility distance, lane setting, intersection channelization (Li et al., 2007). The influence coefficient can be gained through analyzing and calculation of the actual measurement and experts' scoring, which is shown as follows:

$$f_j = \sum_j \alpha_j * A_j \tag{2}$$

$$A_j = 1 + \frac{100 - F_s}{100} \tag{3}$$

where,

f_j =influence coefficient of geometric design;
A_j= minor factors of geometric design;
α_j= weight of minor factors;
F_S= score of the *s*th minor factors.

According to regression analysis of expert' estimation, weight of influence coefficient of geometric design is obtained as shown in Table 2.

Influence Coefficient of Safety Facilities. Many studies shows that safety facilities also have influence on intersection safety (Liu et al., 2001; Tindale et al., 2004; Chu et al., 2006; Xiao et al., 2006), the adjusting coefficient of safety facilities is as follows:

$$f_s = \sum_s \varphi_s * A_s \tag{4}$$

$$A_s = 1 + \frac{100 - \phi_{sl} F_{sl}}{100} \tag{5}$$

where,

f_s=influence coefficient of safety facilities;
A_s= minor factors of safety facilities;
φ_s= weight of the s^{th} minor factors;
ϕ_{sl} = weight of the l^{th} sub- minor factor of the s^{th} minor factors;
F_{sl}= score weight of the l^{th} sub- minor factor of the s^{th} minor factors.

According to regression analysis of expert' estimation, the weight of safety facilities is obtained as shown in Table 3.

Table 3. Weight of safety facilities factors.

Minor factors	Weighted of minor factors	Sub-minor factor	Weighted of sub-minor factor
Signal light	0.25	Location of signal lights	0.28
		Visibility of signal light	0.50
		Rationality of Signal Phase	0.22
Sign	0.22	Visibility of sign	0.45
		Location of sign	0.33
		Information of sign	0.22
Marking	0.24	Visibility of marking	0.58
		Location of marking	0.42
Pavement	0.15	Pavement unevenness	0.43
		Pavement friction	0.57
Street light	0.14	Location of lights	0.60
		Integrity of lights	0.40

A certain intersection, it is not necessary to locate all of the traffic facilities. For example, it is not necessary to locate the signal light when the traffic volume is small in the intersection. As for these conditions, the present evaluation measures will regard the missing item as full score, which will lead to a relatively high score. Therefore, this study employed a method to distribute the relative weighted value of the absent evaluation t objects proportionally to other evaluation objects. This method can reflect more scientifically and rationally the variation of each evaluation objects weight under the circumstance of absent intersection facilities. The Equation is shown as followers:

$$\varphi_s^{'} = \frac{\varphi_s}{1 - \sum_k \varphi_k} \tag{6}$$

where,

φ_S ' = after adjusting weight of the s^{th} minor factors;

φ_S = before adjusting weight of the s^{th} minor factors;

$\sum_k \varphi_k$ =sum of weights of absent facilities of the minor factors.

Model Verification

In order to verify the rationality of the evaluation model, 15 intersections are selected to score by 5 experts from 0 to 100. The relativity analysis of the value calculated by the model and the score estimated by the experts is shown in Figure 1.

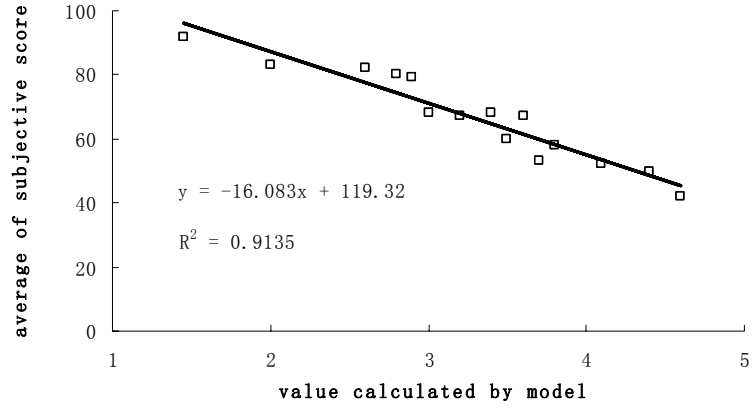


Figure 1. Relativity analyses of value calculated by model and subjective score.

According to Figure 1, the general score calculated by the evaluation model has a good relativity with the experts' subjective score. The model is proved to be reasonable.

Conclusions

Intersection acts as important node of road network. The study divides factors affecting intersection safety into three groups, and establishes intersection safety evaluation model and calibrates parameters of the model. With the value calculated by the model and the score estimated by the experts, the reasonability of the model is demonstrated. Because of limited data, the model structure and parameter value are still not adequate and the model still needs improvement, which will be further improved and fulfilled in the research.

Acknowledgements

This study was supported by Colleges and Universities in Jiangsu Province plans to graduate research and innovation (No. CX09B_059Z) and China Department of Transportation (No. 200831800008).

References

Chu X.M., Yan X.P., Mao Z. (2006). “Automatic classify method of traffic sign.” Journal of Traffic and Transportation Engineering, 91-96.

Li T.H, Wu H.J(2007). “Causes and countermeasures of highway traffic accidents in long straight line combined with sharp curve.” *China Journal of Highway and Transport*, 35-40.

Liu H. X. (2001). "Ergonomics of Intersection Traffic Signal Installation."*Journal of Traffic and Transportation Engineering*, 100-103.

Lu J., Xiang, Q.J.,Zhang G.Q.,and Yuan L.(2009) "Theory and Method for Highway Intersection' Traffic Safety Design." *Science Press*.

Pan, Fuquan, Lu, Jian John, Xiang, Qiaojun (2008) "Influencing factors for safety level of service and its principal component analysis." *American Society of Civil Engineers,* 1710-1715

Tindale, S. A, Noyce D. A., Seyfried R K. (2004). "Toolbox on Intersection Safety and Design."*: Institute of Transportation Engineers.*

Xiang, Q.J., Lu C., Wu Q., Lu J.(2008) "Traffic Safety Evaluation on Highway Intersection Based on Severity Division of Traffic Conflict ."*Journal of Highway and Transportation Research and Development*, 128-131.

Xiao R. M.. et al (2006) "Design of traffic safety signs on mountain roads "*Journal of Chang'an University (Natural Science Edition*, 63-67.

Zhu S.X. , Lu J., Xiang, Q.J.,Y L.L. (2009) "Intersection Safety Evaluation Method Based on Bayesian network" *IEEE Computer Society*, 234-237.

Measuring the Saturation Flow, Lost Time and Passenger Equivalent Values of Signalized Intersection

Henk J. van Zuylen[1] and Jie Li[2]

[1] Hunan University, Changsha PR China, h.j.vanzuylen@tudelft.nl
[2] Delft University of Technology, Delft, The Netherlands, Email: ljlj369@msn.com

Abstract

In the past important differences have been found in the saturation flows in China and The Netherlands. Chinese saturation flow rates appeared to be 20 to 30% lower than the ones in The Netherlands. In order to confirm these conclusions, the methods for the estimation of saturation flow rates and other characteristics of flows at signalized intersections are compared. The regression method appears to give the most informative results, while headways give very detailed information but don't give an unbiased estimation of the saturation flow rate. The third method, the product limit method, uses also the information of unsaturated flow conditions and can give a direct estimation of the probability distribution of saturation flow rates and other characteristics.

Introduction

Intersections are the main capacity bottlenecks in urban road networks. Different road users have to share the intersections space, which causes conflicts and consequentially unsafety and long delays. Signalization with traffic lights can improve the conditions. The operation of signalized intersections is determined by the traffic flows, the composition of the traffic flows, the geometrical structure of the intersection – i.e. number of lanes, allocation of lanes to traffic streams, pedestrian crossings etc – the structure of the control scheme, and the saturation flow.

The saturation flow rate is defined as the maximum number of vehicles that can pass a stop line per hour during the green phase. This maximum flow rate is present a few seconds after the start of the green phase and is assumed to continue until the queue has dissolved. In the beginning of the green phase the drivers need time to react and to accelerate. This is often taken into account as a start up loss, a part of the green

phase that is not effectively used to serve the traffic. At the end of the green phase the traffic flow can still continue during the first seconds of the yellow phase, giving an extension of the effective time that a flow has access to the intersection.

Often a distinction is made between the shown green time t_g and the effective green time t_{eff} as

$$t_{eff} = t_g - \lambda_1 + \lambda_2 \tag{1}$$

where λ_1 is the start lag and λ_2 is the extension at the end, the end lag.

The time between the end of an effective green phase for one direction and the start of the effective green phase of the following direction in the control scheme is called lost time t_l.

The flows v_i, saturation flow rates s_i and lost times $t_{l,i}$ are the quantities that determine the traffic control. The minimum cycle time C_{min} is given by

$$C_{\min} = \frac{\sum_i t_{l,i}}{1 - \sum_i v_i / s_i} \tag{2}$$

Where the sum of t_l is over all transitions of the signals and the sum in the denominator is over the flows in the so-called conflict group: the group of flows that are all pair wise conflicting so that they need to have separate green phases. There are several conflict groups; the conflict group with the highest sum $\sum_i v_i / s_i$ is called the dominant conflict group. That conflict group determines the possibilities to control the intersection. The sum should always be less than 1 in order to have a feasible, not oversaturated intersection control.

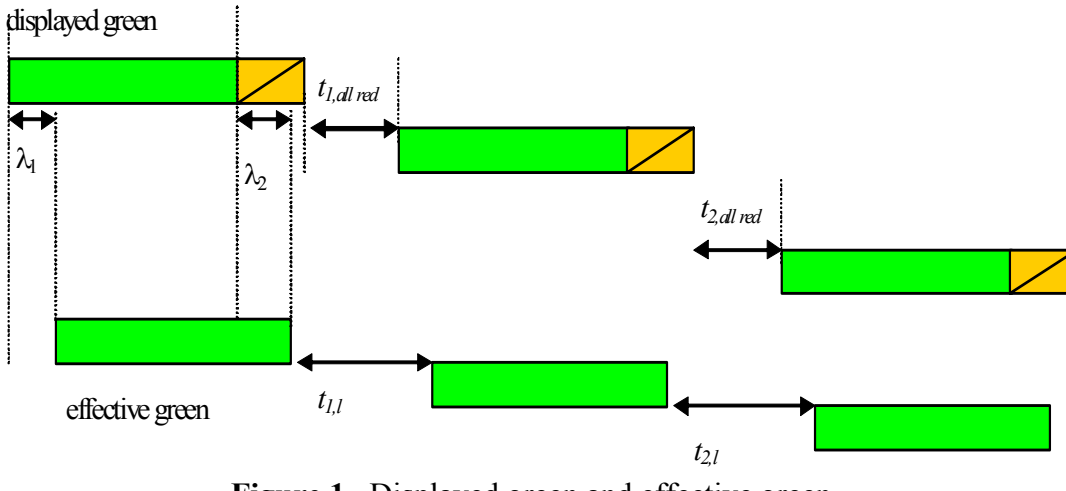


Figure 1. Displayed green and effective green.

The formula (2) shows the importance of the saturation flow rate and lost time for the traffic control. The saturation flow rate can also be seen as the inverse of the time headway between consecutive vehicles. That is the way HCM 2000 defines saturation flow: the inverse of the headways between vehicles after the fourth in the queue. It is clear that different classes of vehicles have different (average) headways. A bus or (articulated) lorry will need more time to pass an intersection than an ordinary passenger car. To deal with this difference, the concept of *passenger car unit* (pcu) is introduced, the ratio of the average time needed by a vehicle of a specific class and the average time needed by a passenger car (Kimber et al. 1985).

There are several ways to determine the saturation flow, the start and end lags and the pcu values. Branston and van Zuylen (1978) developed a sophisticated, but still simple method to determine simultaneously saturation flow rate, pcu values, start and end lags. Before 1978 the method most used was to count vehicles in periods of 5 seconds and take the average, ignoring the first time period in the green phase and all time periods that were not fully saturated (TRRL 1963). Nowadays, the method described in the Highway Capacity Manual (HCM 2000) is most commonly used, although very few traffic engineers really measure saturation flows. Instead of verification with real traffic flows, they use approximate models to calculate the saturation flow from lane width, gradient, presence of bus stops etc.

This paper presents methods to verify the model calculation and to determine the saturation flow from observations. In the next section the regression method which is an extension of the method described by Branston and van Zuylen (1978) will be explained. The disturbances of the traffic flows, especially remaining conflicting vehicles at the start of the green phase and lane changing on a short distance before the intersection, can have an important influence of the start lag and saturation flow rate. Such factor is taken into account in the new method. Important in the measurement of the saturation flow rate is that the observations should be done in time steps that are fully saturated. It means that the queue still exists. It might be difficult to determine whether the conditions for a fully saturated observation period are still fulfilled after the front of the queue has moved backwards. It is possible that after some time the flow over the stop line is more determined by a bottleneck upstream. With help of first order flow theory this question is analyzed in section 3. Finally another saturation flow rate estimation method, using also the observations of unsaturated time periods, will be introduced. The paper ends with a discussion of the influence of several factors on the determination of the saturation flow rates.

The paper combines a methodological approach with experiences in two countries: China and The Netherlands. The difference between driving conditions and driver behavior between two countries gives rise to important extensions of the previous methodologies for the estimation of saturation flow rates.

The Practical Situations for the Evaluation of the Saturation Flow Rates

A main incentive to analyze saturation flow rates comes from the experience of the first author of this paper. When he analyzed the congestion on the first ring road of Shanghai and the second ring road in Beijing, he found that the inappropriate traffic control at the off-ramps of the ring roads is an important reason for the congestion. The traffic control did not perform well due to two reasons: The structure and timing of the control scheme did not suit the traffic conditions and the saturation flow rate was 20 to 30% lower than the saturation flow rate of comparable situations in The Netherlands and USA (van Zuylen H.J. et al. 2003). Later research compared the saturation flow rates in 3 Chinese and 2 Dutch cities with similar results: the lanes at Chinese intersections have a significant lower saturation flow rate than the ones in The Netherland (van Zuylen et al. 2009, Li et al., 2010).

The reasons for this difference are the structure of the traffic control, the more complicated traffic situations at Chinese intersections and the driving culture. The most important issues for this paper are the methodology for the assessment of the saturation flow rate and related characteristics of the traffic and traffic control. The examples are taken from the research done in China and The Netherlands.

The Operational Definition of Saturation Flow

From the discussion in the previous section the number of vehicles that crosses the stop line during a time period during green or yellow is

$$n_i + n_{i1}p_1 + n_{i2}p_2 + \ldots = s(T_i - \lambda_1\delta_{i1} + \lambda_2\delta_{i2}) + \varepsilon_i \quad (1)$$

where

n_i = number of passenger cars that cross the stop line in time period T_i
n_{i1} = number of vehicles of class 1 that cross the stop line in time period T_i
n_{i2} = number of vehicles of class 2 that cross the stop line in time period T_i
p_1 = the pcu value for vehicles in class 1
p_1 = the pcu value for vehicles in class 2
s = the saturation flow
T_i = duration of the observation period; the length of this period should be limited to the part with in the green phase when the vehicles are counted that pass at the end and after the end of the green phase
λ_1 is the start lag
λ_2 is the end lag
δ_1 = 1 if the observation period starts at the beginning of the green phase, otherwise $\delta_1 = 0$
δ_2 = 1 if the observation period ends at the end of the green phase, otherwise $\delta_2 = 0$
ε_1 = the error dues to non systematic deviations of the traffic process from the supposed model.

If the parameters p_1, p_1, s, λ_1, λ_2 are normally and independently distributed, these parameters can be estimate by multiple linear regression, i.e. a least square optimization. A similar approach using a minimization of Theil's measure of difference between observations and model is adopted by Radhakrishnan (2009) for the determination of saturation flow rate and pcu values.

There is some approximation involved in formula (1). The effect of the start up of the queue is found in the first few vehicles whose drivers need some time to react and accelerate. That is the normal case as observed in the Netherlands and other countries. There is some discussion about the length of the time in which the effect of the start lag is still observable. HCM 2000 assumes that after the fourth vehicle the start effects have stabilized, while Bonneson (1992) found that only after 8 vehicles the stabilization exists.

The number of the first vehicles that experience the start up delay is different in the Chinese intersections and the Netherlands ones. While in the Netherlands only the first two or three drivers were affected by the start up delay, on the Chinese intersections four to six drivers were slowed down in the start phase. The reasons for that were also investigated by Li et al. (2010). They found that the start lag is influenced by the number of conflicting vehicles m_1 still on the intersection at the start of the green phase.

$$n_i + n_{i1}p_1 + n_{i2}p_2 + \ldots = s(T_i - (\lambda_1 + q_1 m_1)\delta_{i1} + \lambda_2\delta_{i2}) + \varepsilon_i \quad (2)$$

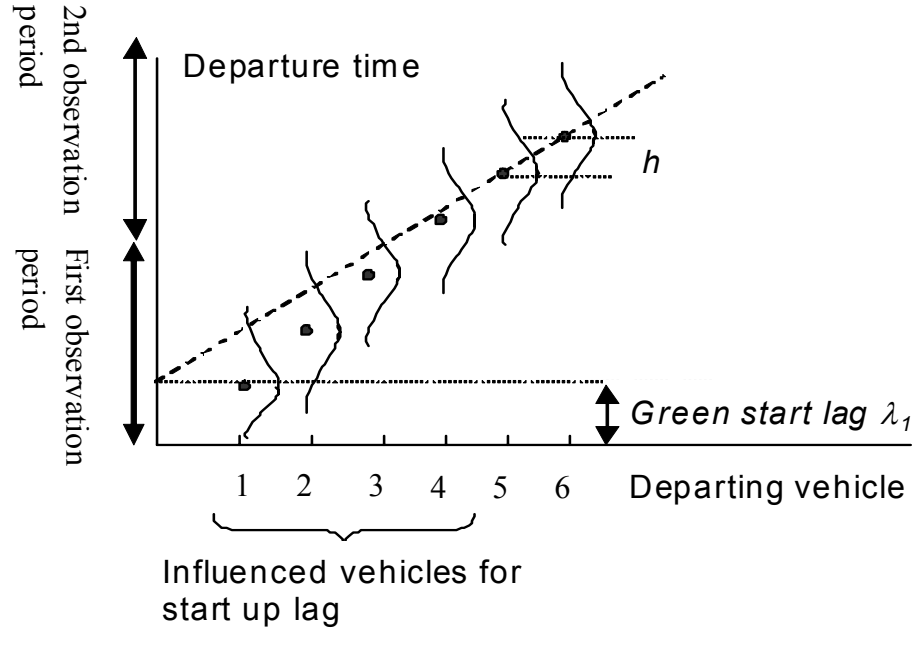


Figure 2. Probability distribution of passing moments and expectation value.

Notice: the probability distribution of the first vehicle(s) is skewed since they can not pass before the beginning of green.

The value of the regression parameter can also be determined from observations. Li et al. (2010) found that $s.m_1 \cong 0.33$, i.e. each remaining vehicle at the start of the green phase reduces the number of passing vehicles with 0.33. The same effect of the influence of remaining conflicting vehicles has been reported by Long (2005) who described the effect as an increase of start lag from 0.5 to 1.1 s.

Another mechanism that should be included in the expression for the saturation flow is the fact that in China it is not uncommon that drivers change lanes just before the intersection. Lane changing means that one vehicle for some time occupies two lanes. It is likely that such behavior will reduce the number of cars that can pass the stop line. If these mechanisms are not considered separately, the estimated saturation flow rate would be biased and lower than the reality, because the impact of illegal overtaking is ignored. After including the number of illegal lane changing m_2, the expression for the number of vehicles that can pass the intersection becomes:

$$n_i + n_{i1}p_1 + n_{i2}p_2 + \ldots = s(T_i - (\lambda_1 + q_1 m_1)\delta_{i1} + \lambda_2 \delta_{i2}) - q_2 m_2 + \varepsilon_i \quad (3)$$

In Eq. (3), q_2 is the coefficient that gives the reduction of number of vehicles passing the stop line per illegal change of lane. For an intersection in Changsha, Li et al. (2010) found the value $q_2 = 1.4$: i.e. every lane changing before the stop line prevents 1.4 vehicles to pass. Table 1 presents the result of the analysis of saturated traffic flow. It shows that the saturation flow rate is underestimated if the influence of conflicting vehicles and illegal lane changing are ignored.

Table 1. Analysis of observations of saturation flow on an intersection in Changsha.

Results	Saturation (s)	Start lag (λ1)	p1	n1	q1	m1	q2	M2
With adjustment	1570	1.86	1.9	8	1.4	54	0.33	13
Statistic test		R2:0.972	F:297.92	P: 0	error variance 2.84			
Without adjustment	1510	3.11	1.9	8				
Statistic test		R2:0.965	F:430.770	P: 0	error variance 3.90			

As mentioned before, the saturation flow can only be observed when the queue is still present. However, after the beginning of the green phase the front of the queue moves backwards, the traffic that passes the stop line is in free flow condition and the flow is determined by the demand or capacity of the upstream bottleneck. It is possible that another upstream bottleneck with lower capacity determines the flow, e.g. a bus stop or road geometry with fewer lanes. This situation will be analyzed in the next section using first order traffic flow theory.

First order traffic flow theory

First order traffic flow theory explains the macroscopic behavior of traffic in terms of the fundamental relation between average speed u, flow q and density d:

$$q = d*u \tag{4}$$

This so-called fundamental relationship between flow and density is often assumed that a shape like the one shown in Figure 3.

One important quantity is the maximum flow v_m, the saturation flow, with the corresponding density d_v. The other important quantity is the maximum density d_m, the density of queued vehicles.

The reality is a little bit more complicated than figure 3 shows. Traffic appears to have different behavior when driving in free flow conditions and driving from a queued situation. In free flow the maximum flow is higher than in the situation that a queue is discharging. The fundamental diagram is more realistic given in Figure 4.

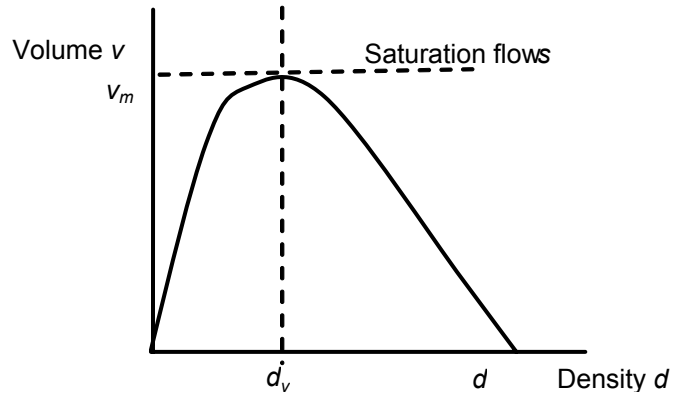


Figure 3. The relationship between density d and volume v in a simple fundamental diagram giving.

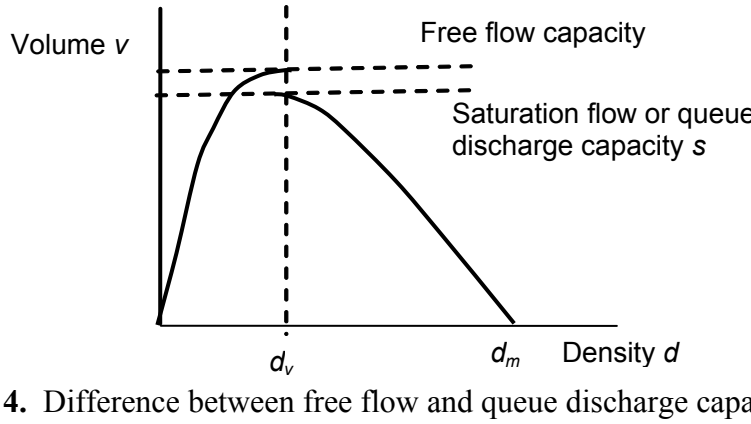


Figure 4. Difference between free flow and queue discharge capacity in fundamental diagram.

In the fundamental diagram three states can be distinguished: the arrivals (assumed for simplicity that traffic arrives randomly) with volume q_1. The speed of the traffic that arrives at the back of the queue is slightly dependent on the density and in the fundamental diagram the speed is given by $u_1 = \partial v / \partial d \big]_{d=d_1}$

In
Figure 5 three states can be distinguished:

(1)the free flow of the arriving traffic at the back of the queue
(2)the queuing traffic
(3)the queue discharging traffic

The state 3 is the state where the saturation flow can be observed. The transition between the back of the queue moves backwards with a speed given by the tangent of angle α_1 in
Figure 5 a, given by $v_1 /(d_m - d_f)$ (see e.g. Hoogendoorn 2004). The transition between the queued state 2 and the moving traffic with flow rate s has a density wave speed $\tan(\alpha_2) = s/(d_m - d_v)$. At the point where the two density waves meet (x_m) the queue is over and the flow at the stop line is just the flow v_1. The moment that the queue has dissolved is $t_q = Q_{max}\ t_b / (1 - v_1 t_b)$ seconds after the start of the effective green, where the time t_b is the time difference between consecutive cars in the queue to go in motion. That is the speed of the density wave between state 2 and 3 expressed in vehicles. Q_{max} is the length of the queue at the start of the green phase.

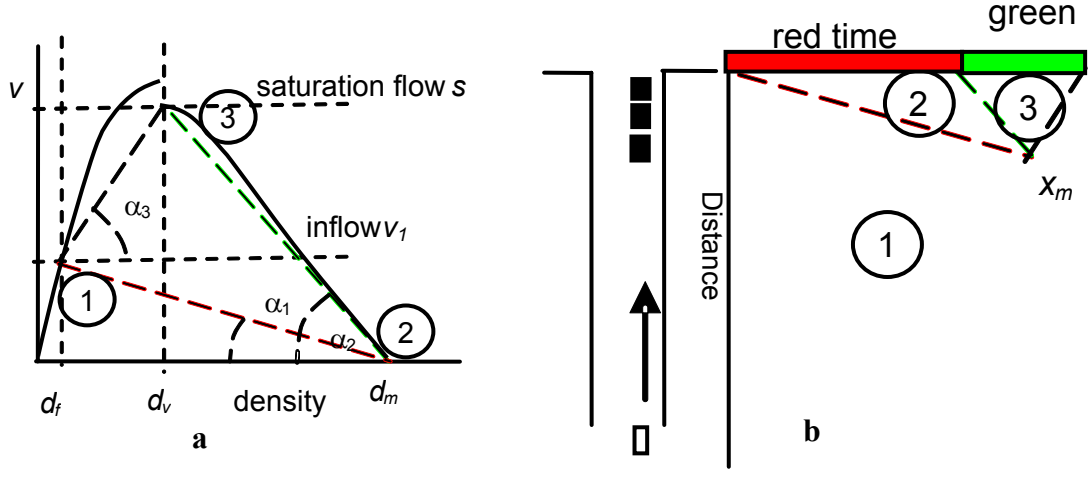


Figure 5. Different states of the traffic in front of the stop line.

It is important for the saturation flow analysis to know the duration time of the saturation flow. In the case of a uniform road profile, the saturation flow can be measured until the last previously stopped vehicle crosses the stop line. Often roads are close to an intersection upstream or the upstream road profile is narrower than that at the intersection. That means that the capacity of the upstream bottle neck can be measured after some time of the green phase as shown in Figure 6.

Giving a formula for the time T during which the real saturation flow can be measured is not complicated if the precise relation between d and v is known. In reality the behavior of the traffic is not so exactly described by a simple deterministic relationship. The transition between region II (queue) and III (traffic in saturation flow) is given by $s/(d_m - d_v)$, which contains quantities that can easily be determined or at least estimated. The speed of the transition between state III and IV is $(s - c_3)/(d_v - d_{IV})$, where c_3 is the bottleneck capacity. The estimation of d_{IV} is more difficult and needs some direct observations of the number of vehicles on the road close to the intersection after the queue is fully in motion.

Most important in this analysis is that the traffic engineer has to take care in drawing conclusions from a decreasing saturation flow when he observes a decrease of the value of the saturation flow during the progress of the green time.

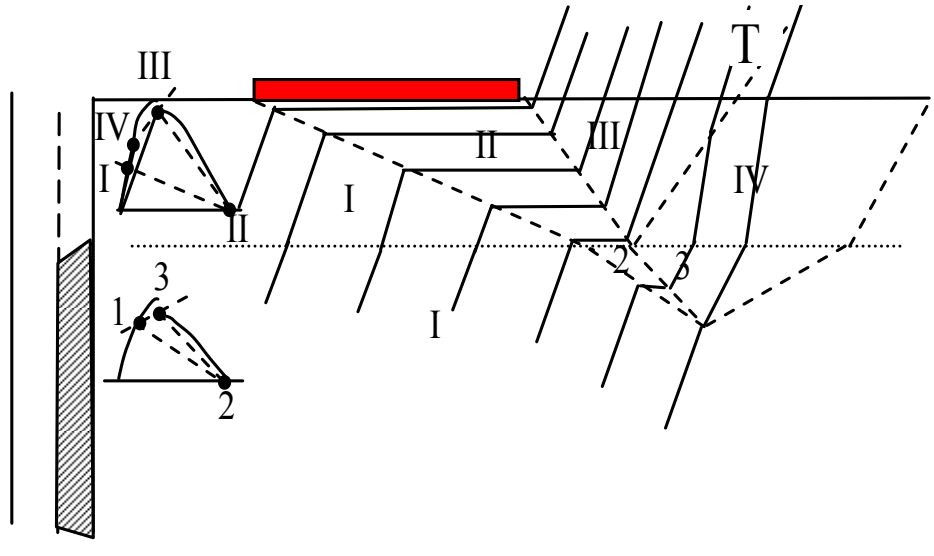


Figure 6. Flow at the stop line is the capacity of the upstream bottleneck after time T.

Headways

An alternative to the measurement of number of vehicles that pass the stop line during a certain time is to measure individual headways. HCM (2000) recommends to measure headways after the fourth vehicle. Potts et al. (2007) and Benneson (1992) found that the start up effects even take longer than the first four vehicles. They use the headways after the eighth vehicle. As Niitymäki and Pursula (1997) and Li et al. (2010) show, the effects of the start delay on headways depends on local conditions and often does not last as long as the 8th vehicle. Often the headways appear to grow after a certain number of vehicles passing the stop line in a queue.

For the analysis of the differences in driving behavior, the headways give a perfect methodology. For 4 Chinese and 4 Dutch intersections the headway distributions

have been measured and analyzed (Li et al. 2010). Significant differences in the distribution of the headways have been found, as shown in Figure 7.

However, the conversion from headways into saturation flow rate is not so straight forwards as suggested in HCM 2000. As Kimber et al. (1985) discussed, saturation flows derived from headways are different from the ones derived directly from regression. This applies also to the estimation of PCU values. The assumption that the saturation flow rate is just the inverse of the headway is too simplistic. In the following it will be shown that the inverse of the headway underestimates the saturation flow rate because of the statistical variation of headways.

The number cars that can pass in a certain time t is $\sum_i^n h_i$. The quantity h is stochastic, not all headways are the same. The probability $P_n(T)$ that in a certain time T n vehicles have past the stop line can be calculated as

$$P_n(T) = p(\sum_i^n h_i \le T).p(\sum_i^{n+1} h_i > T) \tag{5}$$

i.e. the probability $p(\sum_i^n h_i \le T)$ means that the summation of n headways is less than T and that of n+1 headways is longer than T. If the headways are mutually independent, the variance of the sum of headways is proportional to the number of headways in the sum. So the variance of n headways is $n*\sigma$, where σ is the variance of a single headway. For computational simplification reasons, the probability distribution of the sum of n headways is assumed to vary between $n.\bar{h} - \sigma\sqrt{3n}$ and $\overline{n.h} + \sigma\sqrt{3n}$

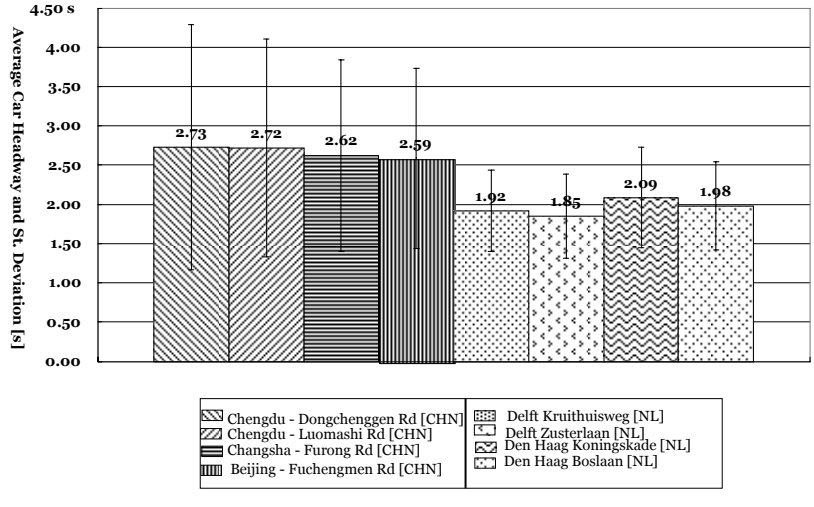


Figure 7. Average car headways and standard deviations on four Chinese and four Dutch intersections.

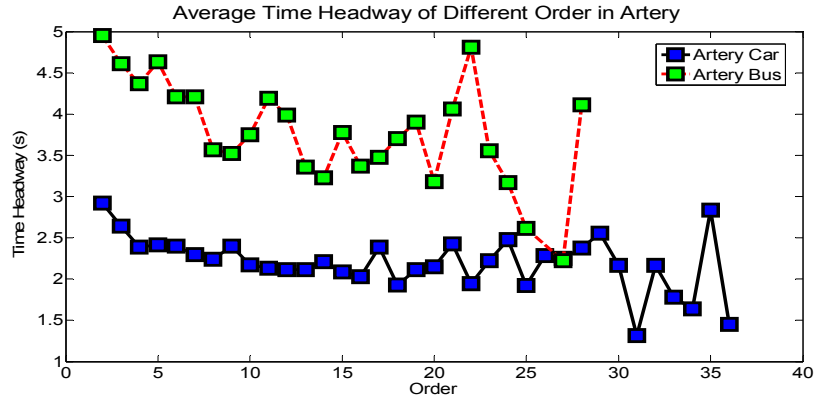


Figure 8 Time headways of vehicles and busses depending on the position in the queue during the red phase in Changsha (CN)

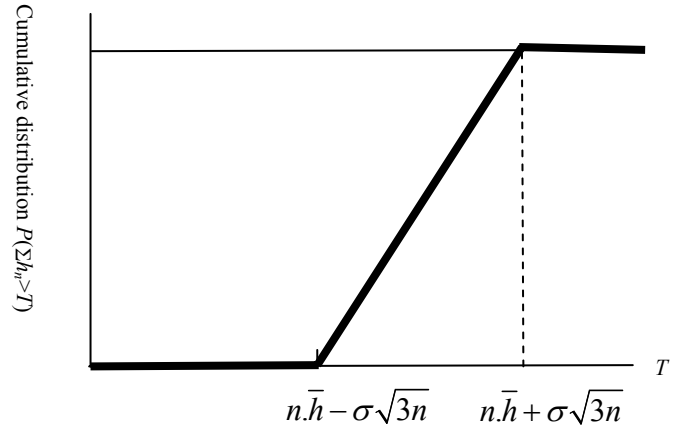


Figure 9 Cumulative probability distribution of headways, assuming a uniform probability distribution with standard deviation σ.

The probability of $p(\sum_i^n h_i \le T)$ is 1 - $p(\sum_i^n h_i > T)$, i.e.

$$p(\sum_i^n h_i \le T) = 0 \quad \text{if } T < n.\bar{h} - \sigma\sqrt{3n}$$

$$= 1 \quad \text{if } T > n.\bar{h} + \sigma\sqrt{3n}$$

$$= (T - n.\bar{h} + \sigma\sqrt{3n})/2\sigma\sqrt{3n} \quad \text{if } n.\bar{h} - \sigma\sqrt{3n} \le T \le n.\bar{h} + \sigma\sqrt{3n} \qquad (6)$$

Similarly the probability $p(\sum_i^{n+1} h_i > T)$ is

$$p(\sum_i^{n+1} h_i > T) = 1 \quad \text{if } T < (n+1).\bar{h} - \sigma\sqrt{3(n+1)}$$

$$= 0 \quad \text{if } T > (n+1).\bar{h} + \sigma\sqrt{3(n+1)}$$

$$= 1 - (T-(n+1)\bar{h}+\sigma\sqrt{3(n+1)})/2\sigma\sqrt{3(n+1)}$$
$$\text{if } (n+1).\bar{h}-\sigma\sqrt{3(n+1)} \le T \le (n+1).\bar{h}+\sigma\sqrt{3(n+1)} \quad (7)$$

Using these formulas, the average number of vehicles $\bar{n}$ that can pass the stop line in a time period T can be determined under the assumption that the period start just after the passing of a vehicle (so that the next vehicle needs a full headway).
There two possibilities for the position of the two probability curves with respect to each other, depending on the value of n and σ.
$n.\bar{h}-\sigma\sqrt{3n} < (n+1).\bar{h}-\sigma\sqrt{3(n+1)}$ or $\bar{h} > \sigma(\sqrt{3(n+1)}-\sqrt{3n})$ as assumed in Figure 10 or just the opposite $n.\bar{h}-\sigma\sqrt{3n} > (n+1).\bar{h}-\sigma\sqrt{3(n+1)}$. The relation between the rightmost two points is always as shown in Figure 10.

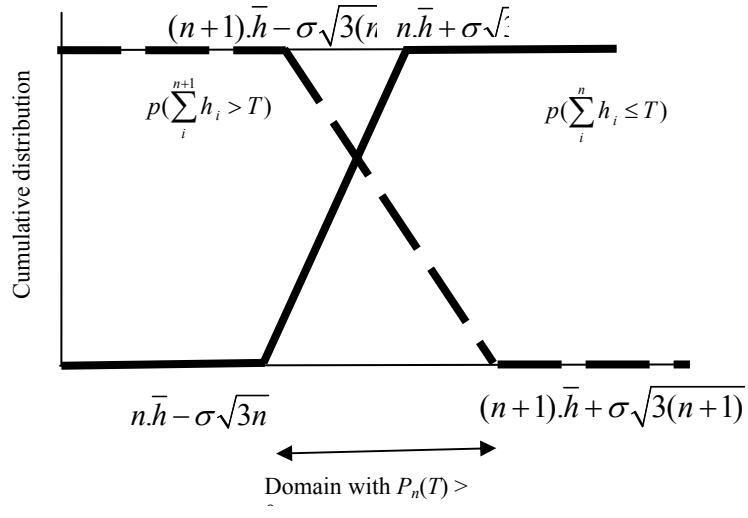


Figure 10. Probability of n headways fitting in a time interval T.

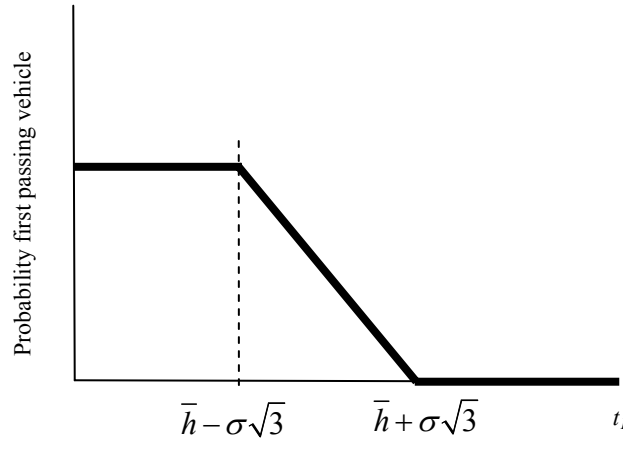


Figure 11. Probability distribution of the first passing vehicle.

If a vehicle is counted when the rear axis passes the stop line, the first vehicle can pass in the beginning of the time period T, say at moment t_1. In remaining time T-t_1 the probability to count n vehicles is calculated with formulas 5, 6 and 7. The probability distribution of the passing of the first vehicle looks as visualized in Figure 11.

The result of a few numerical checks shows that the average number $\bar{n}$ of vehicles that pass during a time T is larger than $T/\bar{h}$, or

$$\bar{h} > T/\bar{n} \text{ and } \bar{s} > 1/\bar{h}$$

The conclusion is that measuring headways is not a valid way to determine saturation flow rates if one assumes that $\bar{s} = 1/\bar{h}$, since saturation flows should always be based on the maximum number of vehicles that can pass the stop line in a certain interval of the green time. A few numerical examples show that the expected throughput of the stop line is at least less that $T/\bar{h}$ with 0.5 vehicle if the standard deviation of the headways is very low. With increasing standard deviation, this difference becomes larger. The conclusion is that the measurement of headways is important and relevant for the quantification of the traffic behavior, but that the link between the headway and the saturation flow deserves special attention.

As an example, the various headways are compared with the saturation flow rates at three Chinese cities using the same observations. Apart from two cases, the saturation flow that is directly estimated with the regression method is larger than the saturation flow calculated as $3600/\bar{h}$.

Table 2. Comparison of average saturation flows and the inverse of the average headway

City	Intersection	Lane	Saturation Flow	St.D.	Average headway $\bar{h}$	St.D.	$3600/\bar{h}$
Beijing	DongZ.M	TH_L	1415	187	2.5782	0.7564	1396
		TH_R	1736	229	2.3665	0.4464	1521
Chengdu	BaB.J._South	TH_L	1457	94	2.5008	0.7592	1439
		TH_R	1557	135	2.4694	0.7902	1458
	DongC.G	TH_L	1573	74	2.4177	0.8556	1489
		TH_R	1436	92	2.4729	0.9472	1456
Changsha	FuR.L_N.2008	TH_L	1335	59	2.6904	0.9792	1338
		TH_M	1526	42	2.4118	0.7604	1493
	FuR.L_N.2009	TH_L	1436	37	2.6175	0.8873	1375
		TH_M	1487	41	2.4121	0.9188	1492
	FuR.L_S.2009	TH_L	1552	65	2.5371	1.1861	1419
		TH_M_L	1556	76	2.5060	1.0655	1437
		TH_M_R	1615	80	2.4328	1.1622	1480
		TH_R	1831	159	2.2947	1.1148	1569

The next section discusses how to estimate the distribution of saturation flows both in under saturated and fully saturated time periods.

Estimating the Saturation Flow With the Product Limit Method

The product limit method (PLM) is a method to estimate parameters from a cumulative probability. The method has been developed by Kaplan and Meier (Cox and Oakes, 1984). The product limit method estimates the probability distribution of the saturation flow rate from observations from all flows passing the stop line during time intervals, not only the saturated flow rates.

Observations of flow rates are *right censored* which means that flow rates are lower than the saturation flow (in the case that the flow is not saturated, i.e. when the queue is no longer present upstream of the intersection). Also the observations of such flows give information about the saturation flow: the saturation flow is larger than the observed flow rate in an under saturated flow condition.

Right censoring of the observations means that each observed flow $f_i(T_i)$ in period T_i is limited by $f_i(T_i) \leq s.T_i$. In the PLM an indicator δ is introduced. The value of δ is 1 if the observed flow is under saturated, otherwise the value of δ is 0.

Hyde and Wright (1986) have independently developed a method to determine saturation flow rates from asymptotic values of observed saturation flow rates during a green phase. They were not able to determine whether the observed flow was saturated or not because there was an upstream bottleneck. The method developed for this situation can be seen as a simplified version of the product limit method.

A probability distribution function of the saturation flow rate, $f(s)$ can be estimated from the observation. In a fully saturated time period, a certain saturation flow rate s_i, the likelihood of this observation is $L(s_i) = f(s_i)$. In an under saturated flow period, the flow rate q_i is lower than the saturation flow rate. The probability of observation of a unsaturated flow rate v which is less than the saturation flow rate, is given by Bayes' theorem

$$P(v \mid \delta = 1) = \frac{P(\delta = 1 \mid v)P(v)}{P(\delta = 1)} \tag{7}$$

The probability distribution of the free flow rate $P(v)$ can be determined from observations upstream of the intersection or it can be assumed to be a Poisson distribution determined by the average flow. The probability $P(\delta = 1)$ is the probability that in an observation period the queue has disappeared, it depends on the cycle time, green time and load ratio of the stop line and can be determined from $\frac{1}{n}\sum_{i=1}^{n}\delta_i$. The probability that a certain observed flow rate is less than the saturation flow rate is given by

$$P(\delta = 1 \mid v) = \int_{v}^{\infty} f(x)dx \tag{8}$$

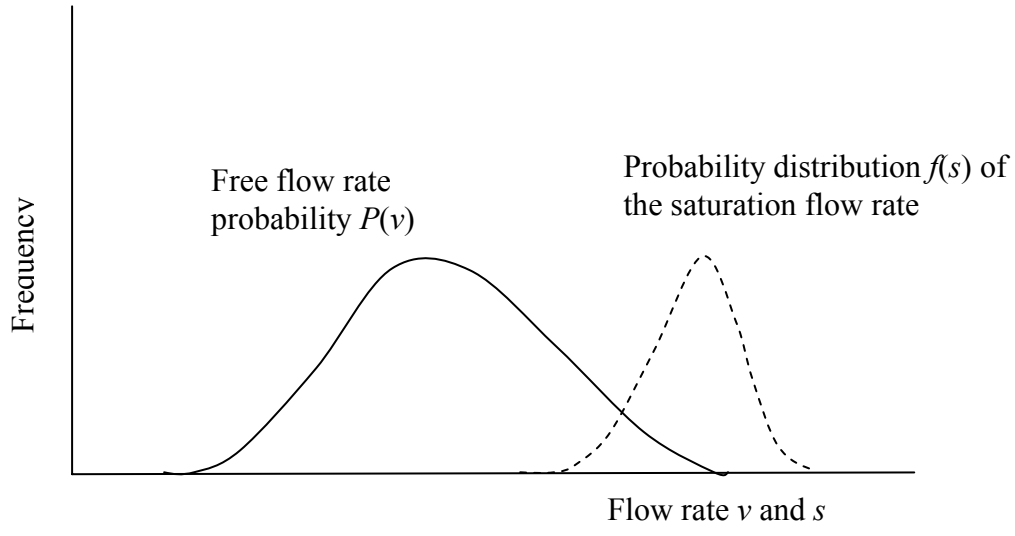


Figure 12. The probability distribution of the demand flow rate and the saturation flow rate.

Both $P(v)$ and $P(\delta = 1)$ are not influenced by the distribution of the saturation flow rate and can be ignored in the likelihood definition. Combining all together, the likelihood of the observation of a certain flow rate is given by

$$L\{v_i\} = \prod_{i=1}^{n} \left[\int_{v_i}^{\infty} f(x)dx \right]^{\delta_i} f(v_i)^{1-\delta_i} \tag{9}$$

If for $f(s)$ a certain probability distribution is assumed characterized by a few parameters, e.g. a normal distribution with an average value and a standard deviation, the likelihood (9) of a certain set of observed flows $\{v_i\}$ can be optimized to find these parameters. Chen et al. (2009) used the product limit method to determine the distribution function of the saturation flow rate from simulations for lanes with mixed traffic. The limitation of the product limit method is that it does not give an easy way to include explanatory factors to explain the variation in saturation flow from cycle to cycle. The method of Branston and van Zuylen (1978) has this possibility, but this method does not have the possibility to include information from under saturated observation periods. Especially the observations of the end of the green phase, which can give information about the end lag of the green phase, are difficult to observe on intersections that are well controlled, i.e. without overflow queues at the end of the green phase.

Example

Both regression method and HCM method are used to estimate saturation flow rates for a situation in Changsha, an intersection at a bridge onramp (at the Wuyi Road). The observation time is from 11:45am to 13:15pm on 29, Dec. 2009. The calculation results are shown in Table 3. Saturation flow rates both on the onramp of the bridge and the Wuyi road are larger than those calculated by HCM method.

Table 3. Comparison of saturation flow rates calculated by two methods.

Method		s(veh/ln/h)	R^2	F	p-value	error variance	Standard deviation
Ramp	Regression	1630	0.95	513.0	0	0.473	
	HCM method	1536					189
Wuyi Rd (Arterial)	Regression	1732	0.98	1908.6	0	1.417	
	HCM method	1676					147

The saturation flow rate distribution in Wuyi Rd in Changsha was estimated by Product Limit Method based on two methods. The traffic flows were measure besides the stop line section. The measurement interval is every 5 vehicles. The applied mean saturated flow rate was set at 1800 pcu/h, which is a representative criterion value for one urban street lane in China. A normal distribution with a standard deviation of 10% was chosen to generate stochastic saturated flow rate. In this example, about 25% of the observed flow rates were saturated flows. Three capacity distributions were calculated:

(1) Empirical Distribution, which is the normal distribution of saturation flow rates with a standard deviation of 10%;
(2) The estimated distribution of capacity using the PLM and assuming normally distributed capacity values;
(3) Estimated distribution of capacity using PLM and a parameter free distribution of capacity values.

The fact that the PLM method gives a 3 to 4% higher value for the saturation flow rate might be explained by the fact that the lane capacity is lower for cars driving in a congested state than for cars in free flow. PLM takes also the free flow conditions into account.

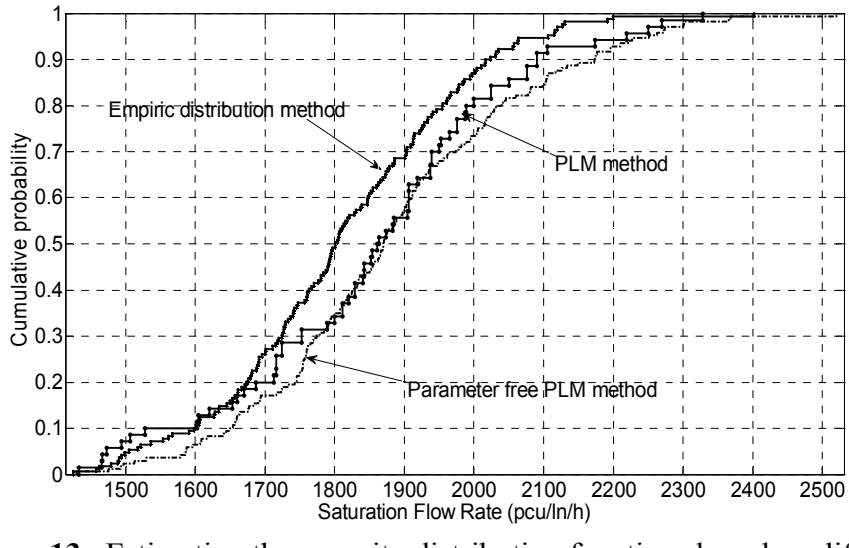


Figure 13. Estimating the capacity distribution functions based on different estimation methods.

Conclusion

Three different methods have been compared for the determination of the saturation flow rate of lanes and approaches of signalized intersections. The significant differences that have been found between intersections in China and The Netherlands require a careful analysis of the methods to estimate the saturation flow rate from observations. The three methods are
(1)the regression method,
(2)the time headway method,
(3)the product limit method.

The first method is only applicable for observation periods that are fully saturated. It has the advantage that several explanatory factors can be included, such as driver discipline and quality of the traffic control. It also offers the possibility to analyze start up delays and the utilization of the first seconds of the yellow phase. The possibility to describe effects of disturbances of flows in an analytical way is important, since especially in China disturbances of saturation flows are quite common. This reduces the average saturation flow, but should not be seen as a factor to determine a lower saturation flow rate. The regression method shows which factors, with corresponding measures, can be expected to improve the saturation flow rates.

The time headway method gives a bias if used to estimate the saturation flow rate. It gives in general an underestimation of the saturation flow rate. However, it gives an excellent possibility to analyze the impact of start up delays on the different vehicles in the queue and to analyze the differences between busses and passenger cars. The product limit method can deliver a probability distribution of the saturation flow rate. Both the observation of saturated flow and unsaturated flow can be used to determine the saturation flow rate distribution. Influences of disturbances of the traffic flow on the distribution function are included, but can not be directly separated.

Research is now going on to fuse the advantages of the three methods to get a method for an unbiased estimation of the distributions of the saturation flow rate, start and end lags, pcu values, and the influences of disturbances.

References

Bonneson, J. A., “Aodeling queued driver behavior at signalized junctions.”Transportation Research Record 1365, Transportation Research Board, 1992.

Branston, D., H.J. van Zuylen (1978). “The estimation of saturation flow, effective Green time and passenger car equivalents at traffic signals by multiple linear regression.” *Transpn Res.Vol*1. (12), 47-53

Chen Xiaoming, Shao Chungfu, Li Da, and Dong Chunjiao (2009). "Capacity reliability of signalized intersections with mixed traffic conditions." *Tsinghua Science And Technology* 14 (3) 333-340.

HCM 2000 (2000)

Hyde, T., and C. C. Wright (1986). "Extreme value methods for estimating road traffic capacity." *Transpn. Res. B Vol*. 208. No. 2, 125-138.

Hoogendoorn, Serge P.(2004). "Traffic flow theory and simulation." Delft University of Technology Course Note 4821.

Kimber, R.M., M.McDonald, N. Hounsell (1985). "Passenger car units in saturation flows: Concept, definition, derivation." *Transpn. Res:B Vol*. 19B, No. 1, 39-61, 1985

LI Jie, Henk J. van Zuylen, YuSen Chen, Ruihua LU (2010) "Comparison of driver behavior and saturation flow in china and the Netherlands." Paper submitted to ITE journal ITS.

Long, G. (2005). "Start-up delays of queued vehicles." *Transportation Research Record* No. 1934, Transportation Research Board of the National Academies, Washington,D.C., 125–131.

Niittymäki, Jarkko, and Matti Pursula (1997)."Saturation flows at signal-group-controlled traffic signals." *Transportation Research Record* 1572 24-32.

Radhakrishnan, Padmakumar, Tom V. Mathew (2009). "Passenger car units and saturation flow models for highly heterogeneous traffic at urban signalised intersections." *Transportmetrica* 1944-0987.

Potts, Ingrid B., John F. Ringert, John D. Zegeer, Karin M. Bauer, Douglas W. Harwood, and David K. Gilmore (2007) TRB meeting Washington.

TRRL, (1963) "A method of measuring saturation flow at traffic signals." Transport and Road Research Laboratory Road Note No. 34, London: HMSO.

van Zuylen, Henk J., YuSen CHEN, LI Jie, and Ruihua LU (2009) "A comparison of driver behavior and saturation flow in China and the Netherlands." ITS World Congress, Stockholm

van Zuylen, Henk J., CHEN Yu Sen, and LI Min Wei (2003). "Optimizing traffic control for intersections with diminishing saturation flow." *Sadayuki, Tsugawa and Masayoshi* (ed) CTS2003, 10th IFAC symposium on control in transportation systems 29 – 34 Tokyo.

One Approximate and Simple Method for Signal Timing of Overlap Phasing*

Jingyuan Wang[1] and Wencai Liu[2]

[1] Associate Professor, Civil Engineering College, Shenzhen University, Nanhai Ave 3688, Shenzhen, Guangdong, P.R. China, 518060;
Email: wangjingyuan_01@sina.com
[2] Civil Engineering College, Shenzhen University, Nanhai Ave 3688, Shenzhen, Guangdong, P.R. China, 518060; MVA Transport Consultants (Shenzhen) Co. Ltd, Shenzhen, Guangdong, P.R. China, 518033

Abstract

The control program design of overlap phasing was studied for the non-equilibrium traffic flow at a signalized intersection. One approximate and simple method for signal timing of overlap phasing was presented. The application principles and conditions of overlap phasing were discussed starting from its concept. The design and timing methods of overlap phasing were constructed on the basis of the timing method of conventional signal phases. The application of the method was illustrated. The result showed that overlap phasing reduced total flow ratio, cycle, delay, and increased capacity. The research provided a theoretical basis for the design of overlap phasing. The effectiveness of time-space resources utilization and service level can be improved at intersections using the research approach.
Key Words: signalized intersection; overlap phasing; signal timing

The non-equilibrium of traffic flow shows two differences of degree of saturation at an intersection. One is the difference between different approaches of the same intersection, and the other is the difference between different directions of the same approach. In fact, these differences are general and universal as a result of time-space characteristics and direction characteristics of traffic flow. So time-space resources are wasted at an intersection. Overlap phasing is an effective way to solve the waste of time-space resources caused by non-equilibrium of traffic flow in a signalized intersection.

*Foundation: Project supported by the National Natural Science Foundation of China (Grant No: 50908150); the Natural Science Foundation of Guangdong, China (No. 9451806001002319).

Overlap phasing was used in the signal design by British technical staff earlier achieving sound effects (Jin, G., and Zhao, B. 2000). After that, various forms of overlap phasing for different traffic flow conditions were created abroad (2006). These studies paid more attention to the application conditions and the advantages and disadvantages of overlap phasing (2002). The timing method of overlap phasing was almost neglected (NIATT 2009). Domestic research showed similar features with little specialized literature for overlap phasing. Only two literatures (Jin, G., and Zhao, B. 2000, 2002) discussed the application mechanism and feasibility of overlap phasing in accordance with the domestic document retrieval. Generally speaking, the appropriate method for the design of overlap phasing has not been formed home and abroad.

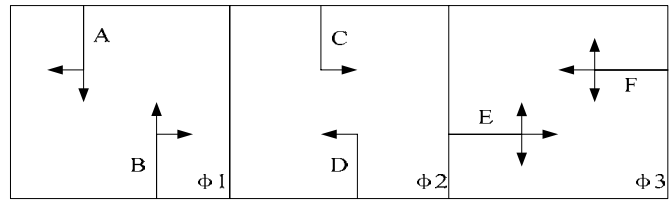
One approximate and simple method for overlap phasing was studied on the basis of a timing method of conventional signal phases. The application principles and conditions of overlap phasing were discussed through analyzing the traffic flow characteristics at an intersection. The method for signal timing of overlap phasing was deduced with the principle of same saturation according to the signal timing method of non-overlap phasing, taking cross intersection as an example. On this basis, the general method and procedure for signal timing of overlap phasing were summed up. Finally, the application of the method for signal timing of overlap phasing was illuminated.

Setting Principle of Overlap Phasin

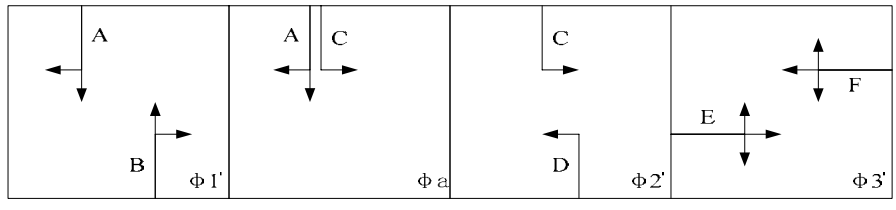
Basic Concept. Overlap phasing is "later start and earlier off (or earlier start and earlier off)" in comprehensible expression. That is to say that one or more traffic flows of the next phase are pre-started before one phase ends, or one or more traffic flows of the previous phase are pre-cut off before one phase begins (2002). In other words, overlap phasing is been set between the end of the previous control state and the start of the next control state making two connected traffic control states overlap. In another way, an additional control state is introduced in the signal cycle (Jin, G., and Zhao, B. 2000).

Figure 1(a) is the conventional three-phase in cross intersection and (b) is the four-phase with overlap phase inserted Φa between Φ1 and Φ2 in (a). In Φa traffic combination flow B was pre-ended, and C was pre-discharged before Φ2 starting, so Φ1 and Φ2 were overlapped.

Setting principle.The following instructions were illustrated for convenient analysis and presentation.



(a) Conventional three phases in cross intersection



(b) Overlap phasing

Figure 1. Conventional phase and overlap phasing.

In Figure 1, A represents the southbound through and right-turn traffic flow combination, B represents the northbound through and right-turn traffic flow combination, C represents the southbound left-turn traffic flow, D represents the northbound left-turn traffic flow, E represents the westbound through and right-turn and left-turn traffic flow combination, F represents the eastbound through and right-turn and left-turn traffic flow combination. It is assumed that the relationships between traffic flow ratio as follow: $Y_A \succ Y_B, Y_C \succ Y_D, Y_E \succ Y_F, Y_A - Y_B \succ Y_C - Y_D$.

In fact, it is general that traffic flows with different flow ratio are included in one phase as shown in Figure 1 (a). The difference of flow ratio will cause the waste of time-space resources at an intersection. The following analyses were carried out with Figure 1 (a) as an example. The time required to release A is longer than B for $Y_A \succ Y_B$. So A is the critical traffic flow in Φ1, and the green time of Φ1 is the time required for A s release. It also means that B s release finishes earlier than A. In a period time of the latter half of Φ1, the space resource occupied by B previously is idle. At this moment, C of Φ2 which doesn't conflict with A can be released in theory. However, C can only be released after the release of A (Φ1 converting to Φ2), because of the restriction of signal phase. There are similar problems for traffic flow combination pairs C and D, and E and F. A, C, and E are the critical traffic flows according to the assumptions. So the total flow ratio of the intersection is:

$$Y = Y_{\mathrm{A}} + Y_{\mathrm{C}} + Y_{\mathrm{E}} \tag{1}$$

Y represents the total traffic ratio of intersection.Y_A, Y_C, Y_E represents the traffic flow ratio of A, C, E in sequence.

The key to solving above problem is to utilize the idle time for the non-critical traffic flow fully (Wencai 2009). The conventional phase (a) was reformed according to the characteristics of traffic flow for achieving the goal. The control mode of overlap phasing (Figure (b)) was formulated by adding a new control state Φa between the Φ1 and Φ2. The detailed method is as following. When Φ1' converted to Φa after B's

release finished, A continued to pass through and C began to release. The signal converted from Φa to Φ2' after A s release finished. After D s release finished, finish releasing, the signal converted to Φ3'....., and so on. The effective green time of Φ1' and Φ2' should be long enough for the release of B and D, respectively. The remainder of A by the end of Φ1' and C by the end of Φ2' would be released in Φa. The length of Φa should be determined by the residual flow ratio of A with the assumption of $Y_A - Y_B \succ Y_C - Y_D$.

According to the above information, C can be started to release (Φa) immediately after B's release (Φ1') finished, without having to wait until the end of Φ1. Then the waste of space-time resources in the intersection caused by the conventional phase was avoided. Under certain conditions, the total flow ratio at intersection with overlap phasing shown in Fig. (b)

$$Y^{'} = Y_{\mathrm{A}} + Y_{\mathrm{D}} + Y_{\mathrm{E}} \tag{2}$$

The total flow ratio of the intersection is lower after using overlap phasing. The reduction of the total flow ratio will bring to shorter cycle, smaller delay, larger capacity and maximize the efficiency of time-space resources at intersection with constant total loss time between the critical traffic flows.

Application Conditions. Three application conditions for overlap phasing can be summarized from the above analysis. Firstly, the number of initial phase is generally 3 or more. Secondly, there are both non-equilibrium of traffic flow obviously in two adjacent phases. Thirdly, the traffic flows with larger flow ratio in two adjacent phases do not conflict with each other.

The first and third items are definite. However, it is necessary to provide one feasible means for judging the degree of non-equilibrium mentioned in the second item. What degree is obvious or not? For this purpose, further analysis was made according to Figure 1.

In Figure 1, the difference between A and B (or C and D) is greater, and the duration of Φa is more longer. The difference is smaller, the duration is shorter. The setting of Φa is possible and necessary only when the difference reaches a certain level. There is an interval between Φ1' and Φa after B s release finishes and before C s release starts if the phase style of signal intersection is shown as Figure 1 (b). A continued to pass through during this interval. It can clearly be seen that the remainder of A by the end of Φ1' can be released in I and Φa. The remainder of C by the end of Φ2' can be released in Φa and the interval between Φa and Φ2' similarly. Therefore, the difference of release time between A and B (or C and D) caused by their difference of flow ratio mustn t be less than a green interval to make the setting of Φa possible. And Φa should last a period of time in practice (the minimum duration of one signal phase is not discussed here).

Calculation for Signal Timing of Overlap Phasing

The method for signal timing of overlap phasing was deduced with the principle of

same saturation based on the signal-timing method of conventional phase according to the setting principle taking the cross intersection as an example. Then, the general method for signal timing of overlap phasing was summed up.

Timing of Conventional Three-phase.

Design of overlap phasing form. Appropriate forms of overlap phasing should be designed firstly when its application conditions are met. One new control state is inserted between the two adjacent phases containing imbalance traffic flows to release the non-conflict flows with larger flow ratio. Φ1 and Φ2 are adjacent as shown in Figure 1(a). A and B are not balanced in Φ1, and the flow ratio of A is greater than B. C and D are not balanced inΦ2, and the flow ratio of C is greater than D. Furthermore, A and C don´t conflict. So, Φa was inserted between Φ1 and Φ2 to release A and C. Overlap phasing was formed (Figure 1 (b)).

Timing of overlap phasing. Following is the timing method of overlap phasing. Signal timing was carried out with conventional phase. Then, signal timing of overlap phasing was obtained through adjustment of signal time with conventional phase according to the setting principle of overlap phasing. Explanation was illustrated by Fig.1 in detail.

Firstly, the signal timing of the intersection controlled by a conventional phase was calculated using the conventional method of F.V. Webster-B.M. Cober (F-B). This time the program meets the principles of same saturation and minimum delay (2006). According to the F-B method,the optimal cycle length C shown as Figure 1 (a) is

$$C = \frac{1.5L + 5}{1 - Y} \tag{3}$$

In the formula, L represents the total lost time in a cycle(s), and Y represents the total flow ratio of the intersection.

The effective green time g_e is equal to the green time of the display g assuming that the lost of time of phase change being equal to the change interval. Then the green time of phase i is

$$g_i = g_{ei} = \frac{y_i}{Y}(C - L) \tag{4}$$

In the formula, g_i represents a green phase (or traffic flow) (s), g_{ei} represents effective green time(s), and y_i represents traffic flow ratio.

In Figure 1(a), $g_1=g_A$,$g_2=g_C$,and $g_3=g_E$.

Then, signal timing was adjusted according to the setting principle with the principle of same saturation. The signal time was adjusted in three steps for Figure 1(b). The

green time of Φ1'and Φ2' were long enough for the release of B and D, respectively. The values of y_B and y_D were taken into Eq. 4, then $g_{1'}=g_B$, $g_{2'}=g_D$ The remainder of A by the end of Φ1' and C by the end of Φ2' would be released in Φa and change interval. So, $g_a = \max[(g_A - g_B) - I, (g_C - g_D) - I] = (g_A - g_B) - I$. ③ The traffic flows in Φ3' were same as in Φ3. So the green time didn't change. That is $g_{3'}=g_3=g_E$. The cycle length of overlap phasing C' is

$$C' = g_A + g_D + g_E + 3I \tag{5}$$

The cycle length was shortened $g_C - g_D$ when using overlap phase.

General method and Procedure for Signal Timing of Overlap Phasing. Signal timing of overlap phasing was discussed above taking Figure 1 as an example. On this basis, the general method and procedure for signal timing of overlap phasing were summed up. The conventional phase is designed for the signal intersection. The signal time is calculated using the conventional method (F-B method). It is judged whether the application conditions of overlap phasing (shown as 1.3) are met in the intersection controlled by conventional phase. The form design and signal timing of overlap phasing is carried out if the application conditions is met. The form of overlap phasing is designed. One new control state is inserted between the two adjacent phases containing imbalance traffic flows to release the non-conflict flows with a larger flow ratio. The signal timing of overlap phase is determined. The signal time of the two adjacent phases before and after the overlap phasing is calculated according to their second critical traffic flow respectively. The time of the inserted phase is calculated according to the larger difference between the flow ration of the critical and the second critical traffic flow in the two adjacent phases. The time of other phase is equal to the time of the conventional phase. The procedure is shown as Figure 2.

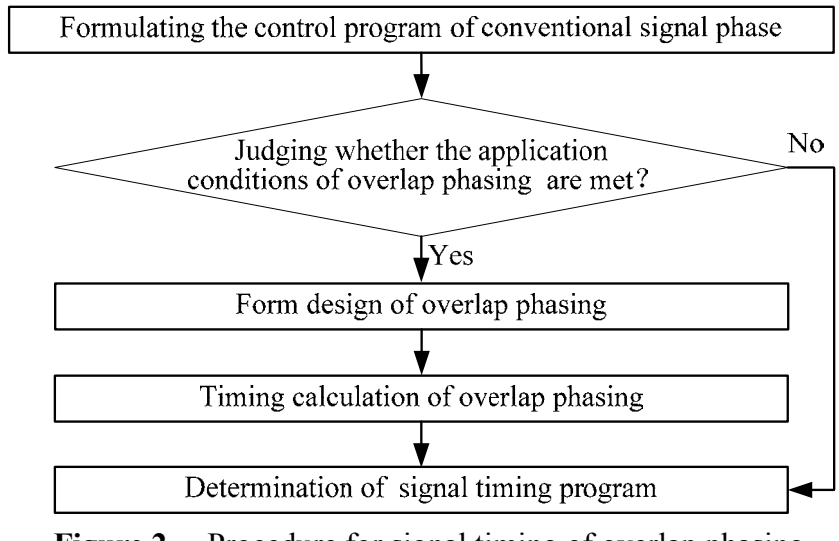


Figure 2. Procedure for signal timing of overlap phasing.

Table 1. Intersection geometry and traffic characteristics.

Flow Direction	Southbound			Northbound			Eastbound			Westbound		
Lane Function	TR	R	L	TR	R	L	TR	R	L	TR	R	L
Number of lanes	1	1	1	1	1	1	1	1	1	1	1	1
Traffic Volume (veh)	798	798	399	608	608	228	304	304	277	228	228	181
Flow Ratio	0.42	0.42	0.21	0.32	0.32	0.12	0.16	0.16	0.15	0.12	0.12	0.10

Numerical Illustration

The characteristics of traffic flow are shown as Table 1 for the example of the cross intersection. The task is to design the signal control program of this intersection without considering the impact of pedestrians.

It is assumed that the design saturation flow rate for all lanes is 1900 vphg (TRB 2000). The traffic flow ratios of each lane are shown in Table 1.

Signal Timing of Conventional Phase. The initial form of the phase for this intersection was determined (Figure 1(a)) according to the characteristics of traffic flow. It can be seen from Table 1 that y_A=0.42,y_B=0.32,y_C=0.32,y_D=0.12,y_E=0.16,and y_F=0.12. The key traffic flow combinations are y_A, y_C, and y_E. The total flow ratio of the intersection is Y=0.79. The cycle lost time L=12s if change interval I= 4s assuming the lost time of phase change being equal to the change interval. By Eq. 3, the optimal cycle length is

$$C = \frac{1.5 \times 12 + 5}{1 - 0.79} = 110\,\text{s} \tag{6}$$

Through Eq. 4, $g_1 = g_A = 52$ s, $g_2 = g_C = 26$ s, $g_3 = g_E = 20$ s, and $g_B = 40$ s, $g_D = 15$ s, and $g_F = 15$ s.

Design and Timing of Overlap Phasing.

Judgment of application condition. The three application conditions defined above were judged one by one for Figure 1(a). (1) The number phase is 3. (2) InΦ1, $y_A \succ y_B$, $g_A - g_B = 12 \succ I = 4$. In Φ2, $y_C \succ y_D$, $g_C - g_D = 11 \succ I = 4$. A and B, C and D can both be considered unbalanced traffic flow combinations obviously. 3) A and C don't conflict. So overlap phasing is suitable for this intersection.

Form design of phase and signal timing. According to the analysis result of application conditions and the setting principles of overlap phasing, Φa was inserted between Φ1 and Φ2 to release A and C, then overlap phasing was formed (Figure 1(b)). The change interval is 4 s.

The signal timing for the control program shown in Figure 1 (b) was carried out. The green time of Φ1' was long enough for the release of B. So, $g_{1'}=g_B=40$s. The green time of Φ2' was long enough for the release of D. So, $g_{2'}=g_D=15$s ; $g_a = \max[(g_A - g_B) - I, (g_C - g_D) - I] = 8$ s. $g_{3'} = g_3 = 20$ s. The cycle length of overlap phasing is $C' = g_A + g_D + g_E + 3I = 95$ s which is 15s shorter than that of the conventional phase.

The total flow ratio of the intersection controlled by overlap phasing is $Y' = Y_A + Y_D + Y_E = 0.74 < Y = 0.79$.

Conclusion

The inappropriate phase is one of the main reasons causing the inefficiency of intersections in China. Overlap phasing is an effective way to solve the waste of time-space resources caused by non-equilibrium traffic flow at intersections. The control program design of overlap phasing at signalized intersections was studied on the basis of the timing method of the conventional signal phase. The application principles and conditions of overlap phasing were discussed starting from its concept. The design and simple timing methods of overlap phasing were constructed, and their applications were illustrated. The results showed that overlap phasing reduced total flow ratio, cycle, delay, and increased capacity. The research results provided a theoretical basis for the design of overlap phasing, and the utilization effective of time-space resources can be improved at intersections.

References

Liu, W. (2009). “Research on ripple change timing method.” *Shenzhen: Civil Engineering College*, Shenzhen University.

Jin, G., and Zhao, B. (2000). “The application of ripple change in the design traffic signal.” *Traffic Management Research*, (2):13-15.

Peng, G., and Zhang, Y. (2002). “Study of the way of setting ripple changes.” *China Journal of Highway and Transpor*t, 12(Sup.): 59-62.

NIATT (2009). Signal Timing Tutorial.
http://www.webs1.uidaho.edu/ niattproject/backgroundbasic3.htm.

TRB (2000). Highway capacity manual (HCM) 2000. National Research Council, Washington, D.C.

Wang, J. (2006). “Study on practical methods of optimization for integrating timing and spacing resources in signalized intersections.” *Nanjing: Transportation College*, Southeast University.

Arterial Signal Timing and Coordination: Sensitivity Analyses and Partition Techniques

Wei (David) Fan[1] and Zong Z. Tian[2].

[1]Assistant Professor, Ph.D., P.E, Department of Civil Engineering,The University of Texas at Tyler,3900 University Blvd.Tyler, TX 75799, PH: 1-903-565-5711, Fax: 1-903-566-7337, Email: wfan@uttyler.edu.
[2]Associate Professor, Ph.D., P.E, Department of Civil & Environmental Engineering, University of Nevada, Reno, Reno, NV 89557, PH: 1-775-784-1232, Fax: 1-775-784-1390, Email: zongt@unr.edu.

Abstract

This paper addresses two issues regarding arterial signal coordination and traffic progression: 1) It presents a sensitivity analysis of determining the optimal network cycle length under various combined options chosen among increment in cycle length, whether and how to allow uncoordinated, and ways to conduct offset optimization; and 2) It compares the effectiveness of three different network partition techniques which can be potentially incorporated into software packages. A case study is undertaken using a real world arterial in the City of Tyler, Texas. Numerical results indicate that using the Quick Offset Optimization command during the process of optimizing cycle length may not be sufficient enough to find the best signal timing plans and this is particularly true if intersections are allowed uncoordinated to a certain extent (ranging from Rarely, to Sometime, to Often). It is also found that the three different partition techniques (i.e., Coupling Index, Strength of Attraction, and Coordinatability Factor) differ only slightly and can produce the same solutions in all scenarios.

Key Words: traffic signals; signal coordination; network partition

Introduction

Several signal timing software packages are available for developing optimal signal timing solutions. Generally, each of these software packages has its unique optimization features which can be classified into two main categories based on the

optimization objectives: bandwidth based and delay based. As self-explained, bandwidth-based software packages intend to maximize the bandwidth when deriving an optimized signal timing solution. Examples of such software include PASSER II (PASSER V User Guide, 2002, Chang, E.C.P.et.al., 1988, Messer, C.J.et.al., 1973), MAXBAND (Little, J.D.C.et.al., 1981), and MULTI-BAND (Gartner, N.A, 1991, Sripathi, H.K.et.al., 1995). Examples of delay based software packages include TRANSYT 7F (TRANSYT 7-F User's Manual, 2006) and SYNCHRO (Husch, D. and Albeck, 2006). Delay-based software packages attempt to minimize system delays as the primary objective and bandwidth is usually a by-product of the optimization process. When evaluating and comparing specific signal timing plans developed by these software packages, microscopic traffic simulation models such as CORSIM/NETSIM (TSIS User's Manual, 2003) are generally used (Yang, X.K, 2001; Tian, Z. and Urbanik, 2007*)* .

Many efforts have been made on combining the advantages of both bandwidth-based and delay-based solutions. For instance, Tian and Urbanik (Tian, Z. and Urbanik, 2007) developed a heuristic-based system partition technique to improve signal coordination and traffic progression. By dividing a large signalized arterial into several subsystems, their proposed approach achieved the maximum bandwidth efficiency while still improved performance measures such as delays, stops and travel speeds. Lan et al. (Lan, C.J.et.al., 1994) developed a methodology to optimize signal coordination on arterials by simultaneously considering the two potentially conflicting criteria (i.e., delay minimization and progression bandwidth maximization criteria). Their strategy could either maximize the progression bandwidth while maintaining a certain user-defined level of service, or minimize the delay at some degree of loss in bandwidth. Wallace and Courage (Wallace, C.E. and Courage, K.G, 1982) proposed an approach to using a measure of the quality of progression perceived by the drivers. A policy was presented to maximize the number of progression opportunities (PROS) available on an arterial signal system. Compared with the maximal bandwidth solutions, the resulted arterial progression design showed decreased stops and delays without suffering the loss of perceived progression associated with direct minimization of stops and delays.

Several studies have been conducted to compare different software packages such as TRANSYT 7F, PASSER II, PASSER V and SYNCHRO (Yang, X.K, 2001, Tian, Z. and Urbanik, 2007, Alam, S.M.M, 2004) and guidelines for selecting signal timing software are also presented (Chaudhary, N.A., Kovvali, V.G. and Alam, S.M.M.,2008). Although bandwidth may still be viewed with high regard in the traffic engineering community in judging the quality of a timing plan and deemed as one of the most visible indicators to individual drivers (Tian, Z. and Urbanik, 2007), delay-based signal timing plans that also accounts for bandwidth maximization is becoming increasingly popular. In particular, SYNCHRO, a delay-based macroscopic analysis and optimization software, has been becoming increasingly used in industry for developing signal timing plans. SYNCHRO can be used to analyze the capacity for signalized intersections, coordinate networks and arterials, and generate time-space diagrams interactively. It allows for quick generation of

optimum timing plans and optimizes cycle lengths, split times, offsets and phase sequence to minimize stops and delay (Husch, D. and Albeck, J., 2006).

It is noted that SYNCHRO generally optimizes a signal timing plan by assuming a specific vehicle travel speed on links along an arterial. However, due to variations in drivers' behavior and vehicle characteristics, the actual vehicles' travel speeds are becoming highly variable and thus may become undesirable for coordination (Tian, Z. and Urbanik, 2007). To develop better coordinated signal timing plans, traffic engineers and researchers have recognized the necessity of dividing a large system into smaller subsystems, a technique called system partition (Husch, D. and Albeck, J., 2006).

System partition has proven to be an effective technique for developing coordinated signal timing plans, and it has received more attention recently (Husch, D. and Albeck, J., 2006, Tian, Z. and Urbanik, 2007). However, this technique has not been well addressed in most software packages. SYNCHRO is perhaps the only software that has an internal algorithm for system partition. In particular, most arterial signal timing studies are conducted focusing on bandwidth-maximization with very little effort on delay-based arterial signal timing studies, although SYNCHRO has been increasingly used as a commercial software package for the purpose of developing signal timing solutions. Let alone, SYNCHRO has various optimization commands and options. For example, when optimizing the cycle length, SYNCHRO sets quick offset optimization as default for obtaining preliminary solutions, while the ultimate optimal solution may have to be obtained through the extensive offset optimization command. Furthermore, SYNCHRO provides various combined options, such as increment in cycle length (2 or 5 or 10 seconds), whether and how to allow uncoordinated, and ways to conduct offset optimization. However, very few studies have examined the sensitivity of such options.

As such, the purpose of this study is to investigate two research issues that have not well addressed thus far: 1) the sensitivity of various combined options chosen among increment in cycle length, whether and how to allow uncoordinated, and ways to conduct offset optimization; and 2) the effectiveness of three different network partition techniques which can be potentially incorporated into software packages. These issues are addressed through a case study at a real world arterial in the City of Tyler, Texas.

The remainder of this paper is organized as follows: Section 2 provides some discussions of three different system partition techniques as well as outlines some basic features of the SYNCHRO software. Section 3 undertakes a case study using a real world arterial in the City of Tyler. Finally, Section 4 includes a summary and discussion of future research directions.

System Partition Techniques

Several attempts have been made over the years to develop a factor to determine where system partition (i.e., signal progression breaks) should occur. For example, Hook and Albers (Hook, D. and Albers, A, 2008) proposed and briefly compared three different approaches to determining breakpoints in signal progression using five randomly chosen links in the City of Fort Collins as a case study. These three different methodologies that were originally presented by Hook and Albers (Hook, D. and Albers, A, 2008) are adopted and discussed here in detail. A comparison of the three approaches is conducted in Section 3 through a case study.

Coupling Index. The simplest among the three system partition methods is probably the Coupling Index (CI) method derived from the Newton's law of gravitation which states that the attraction between two bodies is proportional to their sizes and inversely proportional to the distance squared. The formula for calculating the Coupling Index is shown below.

$$CI = \frac{V}{D^2} \tag{1}$$

where: CI = Coupling Index (unitless); V = Traffic Volume for the Analysis Time Period (1000 vehicles/hour); D = Link Distance (miles)

It should be pointed out that CI is a data at the link level along the arterial or in the whole network. A higher number generally means a higher need to coordinate adjacent signals that composes the link. In principle, there is no absolute CI at which coordination should or should not occur, but Table 1 shows the rule of thumb: If a CI is less than 1, then there is no need to coordinate. If a CI falls in between 1 and 50, then coordination is desirable. A CI greater than 50 indicates that signal coordination is critical.

Strength of Attraction As self-explained by its name, the Strength of Attraction (SoA) refers to the desirability of coordinating two adjacent intersections. It is a function of the platoon interference (such as on-street parking maneuvers), link traffic volume, link travel speeds and link distance. The following formula is presented for calculating the strength of attraction of any link between two adjacent intersections.

$$AF = I * V * \left(\frac{S}{D}\right)^2 \tag{2}$$

where: AF = Strength of Attraction (unitless); I = Platoon Interference (unitless); V = Traffic Volume for the Analysis Time Period (vehicles/hour); S = Link Travel Speed (miles/hour); D = Link Distance (feet)

In particular, as platoon progresses down the street, it will encounter various degree of interference under different roadway situations. Hook and Albers (Hook, D. and

Table 1. Coordinatability methodology (*Adopted from Hook and Albers, 2000*).

	Coordination		
Methodology	Not Desirable	Desirable	Critical
Coupling Index	<1	1 to 50	> 50
Strength of Attraction	< 0.5	0.5 to 2	> 2
SYNCHRO	< 20	20 to 80	> 80

Albers, A, 2008) gave a good rule of thumb for this factor. For example, the platoon interference factor of 2.0 is recommended for roadways without parking, 1.5 for roadways with parallel parking, and 1.0 for roadways with angled parking. Again, there is no absolute strength of attraction at which coordination should or should not occur. The borderlines drawn for the coordinability based on the Strength of Attraction is given in Table 1.

Coordinatability Factor. An increasingly popular signal timing and intersection analysis software, SYNCHRO (Husch, D. and Albeck, J. 2006) has an internal methodology to calculate the Coordinatability Factor (CF) and uses it to measure the desirability of coordinating the intersections. The calculation of CF is highly empirical and several variables are involved, including link travel time, link traffic volume, link distance, vehicle platooning, vehicle queuing, and cycle lengths. In particular, what makes the calculations of CF in SYNCHRO different from the strength of attraction method is that it also explicitly considers the cycle length and vehicle queuing adjustment. For example, the potential for vehicle queues exceeding the available storage is also considered in determining whether coordination is warranted.

SYNCHRO specifies a good rule of thumb regarding the CF to provide indications whether an intersection should be coordinated with others in the system. The criteria based on all the variables as mentioned above are used to determine a CF on a scale from 0 to 100 or more. For example, a CF of 100 suggests that it is critical to coordinate the signal with others; while a CF below 20 means that the signal should be operated independently. Table 1 again summaries such rule of thumb.

Synchro Features. This section provides some background information regarding the basic features of SYNCHRO, particularly how SYNCHRO performs the calculation, evaluation and optimization process to quickly generate optimum signal timing plans by optimizing cycle lengths, split times, offsets and phase sequence.Figure 1 shows the optimization flows embedded in SYNCHRO.

As can be seen, Step 2 involves the system partition process by dividing a network into multiple subsystems. SYNCHRO uses the CF for such partition processes. In Step 3, SYNCHRO finds the best cycle length by calculating the lowest Performance Index (PI) in which the total delay and vehicle stops are directly accounted for. It should be noted that such minimum PI is obtained for each single run. A single run

can be conducted corresponding to a specific scenario resulted by selecting a specific combined option among increment in cycle length, whether and how to allow uncoordinated, and ways to conduct offset optimization (Husch, D. and Albeck, J. 2006). Since different options may result in different minimum PIs, the sensitivity analyses should be conducted in order to find the overall minimum PI (which yields the best cycle length) among all possible options. Also in Step 3, the SYNCHRO User Guide (Husch, D. and Albeck, J. 2006) states that a quick offset optimization is usually sufficient to determine how a cycle length in study performs so that the "best" cycle length can be identified. However, such practice only evaluates and compares a certain amount of signal timing solutions existed under the quick offset optimization option. A large amount of scenarios are actuallyexcluded and their signal timing solutions are discarded without extensive evaluations. As a result, this Step may yield a local optimal rather than the best cycle length because it potentially wipes out the best overall signal timing plan of a different cycle length. In this regard, although Step 4 uses a thorough offset optimization to optimize the offsets and phase sequences on basis of the "best" cycle length determined in Step 3, it is very likely that the best signal timing plan overall cannot be achieved. In the following sections, this paper will conduct the sensitivity analyses and test out all three (rather than the quick offset only) offset optimization processes to see whether a quick offset optimization does suffice or not.

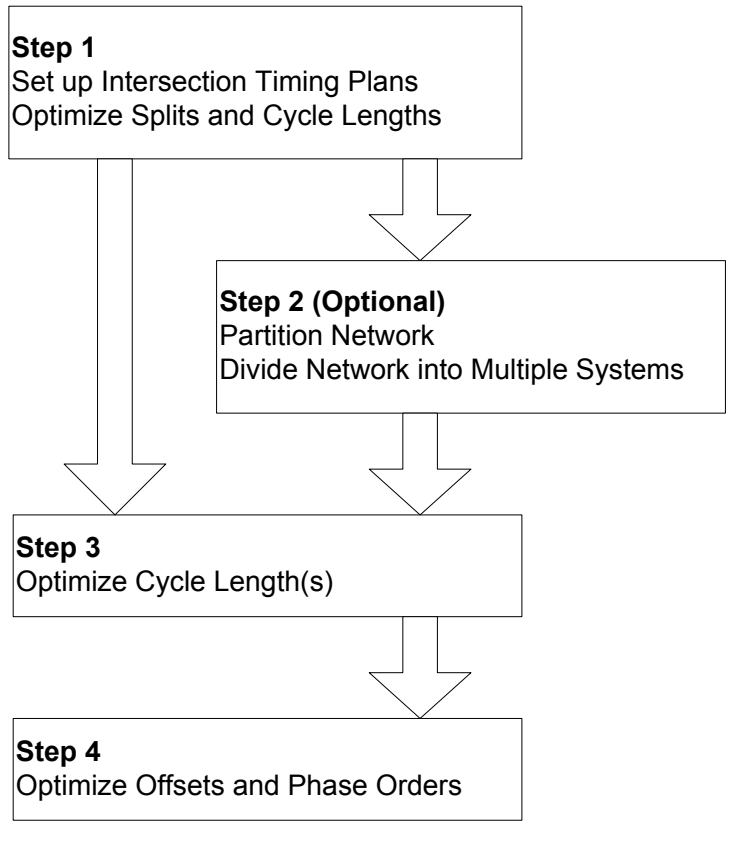


Figure 1. SYNCHRO Optimization steps (*10*).

Case Study

Site Description. A case study is presented in this section to compare all three system partition techniques as well as conduct the sensitivity analyses of the arterial signal coordination as mentioned before.

The test site is the South Broadway Avenue in Tyler, a small-sized city located in East Texas. The South Broadway Avenue is a major north/south arterial serving traffic generated by a number of major retail shopping centers of the City of Tyler. The posted speed limit on South Broadway Avenue is uniform and set as 35 mph (56 km/h). The signal system has 10 signals and the traffic volumes used for developing signal timing plans are collected during several field trips during weekday a.m. peak hours. The system has an approximate 70/30 directional flow split with the northbound being the peak direction during the a.m. peak period. Figure 2 shows a map of the system represented in SYNCHRO and Table 2 summarizes the major traffic control characteristics of each intersection.

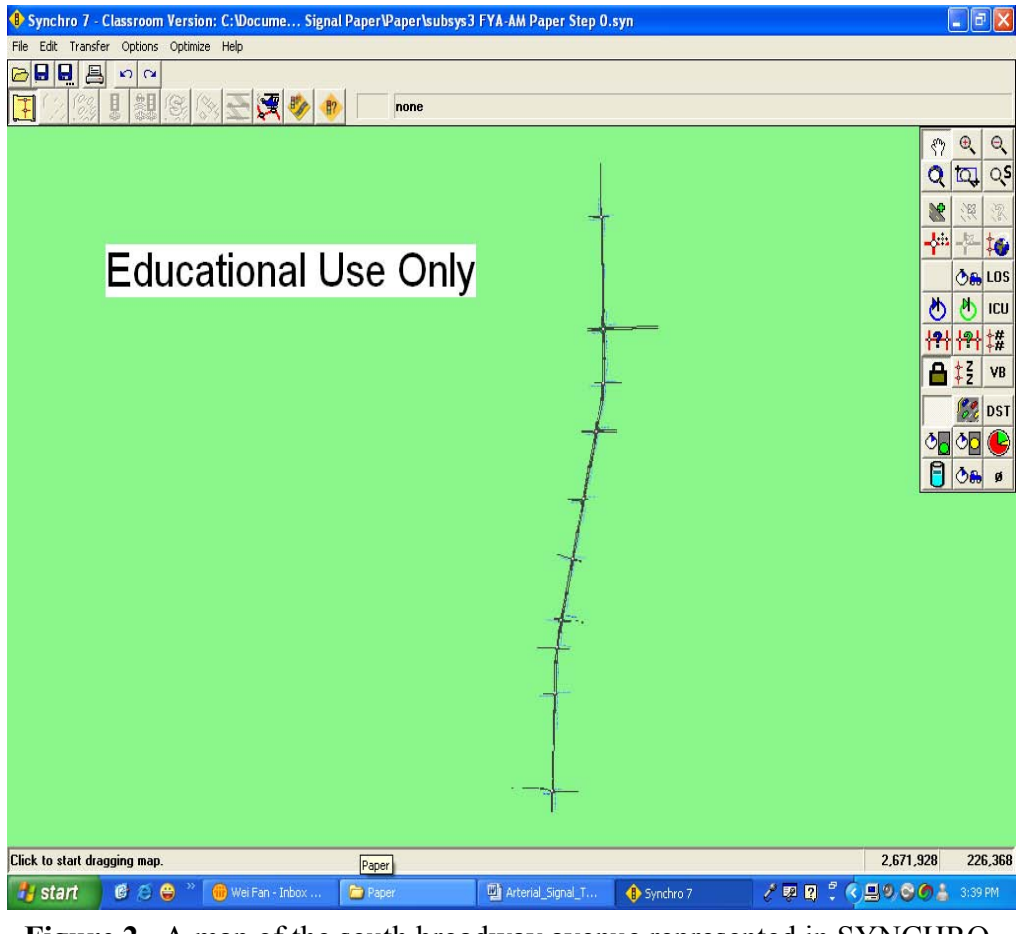


Figure 2. A map of the south broadway avenue represented in SYNCHRO.

Table 2. Summary of traffic and control characteristics.

Intersection number	Cross street name	Spacing (ft)	Main street left-turn phasing	Cross street left-turn phasing
1	Amherst St.	—	Prot./perm.[a]	Prot./perm.
2	Loop 323	3238	Prot.[b]	Prot.
3	Independence Rd.	1566	Prot.	Prot.
4	Shiloh Rd.	1410	Prot.	Prot.
5	Rieck Rd.	2015	Prot.	Prot.
6	Donnybrook Rd.	1746	Prot./D. perm.[c]	Prot./D. perm.
7	Grande Blvd.	1809	Prot.	Prot./perm.
8	R.E. Lee Drive	819	Prot.	Perm.[d]
9	Southtown Dr.	1335	Prot.	Split
10	Heritage Rd.	2810	Prot./perm.	Split

[a]Prot./perm. = Protected + Permitted.
[b]Prot. = Protected.
[c]Prot./D. perm. = Protected + Dallas Permitted.
[d]Perm. = Permitted.

Case Study Results. Using the South Broadway Ave as a case study, the sensitivity analyses of the arterial signal coordination are conducted using SYNCHRO and the effects of three different system partition techniques are also compared.

Sensitivity Analyses Results. Table 3 presents the sensitivity analyses results of optimizing network cycle length. As mentioned before, the delay-based SYNCHRO software finds the best cycle length by calculating and comparing the Performance Index (PI) and the solution with the lowest PI yields the best cycle length. As part of the sequential steps normally involved in developing optimal signal timing plans using SYNCHRO, the offsets need to be optimized after determining the optimal system cycle length. Table 4 provides solution results of the final signal timing plans for all scenarios in Table 3 during the sensitivity analyses process.

Effects on Computation Time

During the process of conducting sensitivity analyses of optimizing network cycle length as shown in Table 3, all results of the effects of sensitivity analyses on computation time are as obvious as one can expect. It is noticed that the computational time decreases significantly in the order of changing from Never (0) to Often (70), everything else being the same. Likewise, the computational time decreases noticeably as cycle length increment increases from 2 to 5 to 10 seconds, if everything else is kept the same. Also, the more extensive the chosen offset optimization, the more the computational time.

Table 3. Sensitivity analyses results of optimizing network cycle length.

				Cycle Length Increment (50~150s)		
Allow uncoordinated?				2	5	10
Never (0)	Offset opt.	Quick	PI	404 (134)[a]	412 (120)	412 (120)
		Medium	PI	385 (130)	385 (130)	385 (130)
		Extensive	PI	384 (124)	387 (135)	388 (130)
Rarely (30)	Offset opt.	Quick	PI	401 (124)	410 (130)	410 (130)
		Medium	PI	393 (124)	400 (125)	400 (120)
		Extensive	PI	396 (122)	402 (120)	402 (120)
Sometime (50)	Offset opt.	Quick	PI	401 (124)	411 (115)	413 (120)
		Medium	PI	393 (124)	399 (115)	400 (120)
		Extensive	PI	394 (114)	402 (120)	402 (120)
Often (70)	Offset opt.	Quick	PI	470 (150)	470 (150)	470 (150)
		Medium	PI	471 (150)	471 (150)	471 (150)
		Extensive	PI	472 (150)	472 (150)	472 (150)

[a]Performance Index (optimal cycle length).

Table 4. Solution results of the final signal timing plans on basis of table 3.

				Cycle Length Increment (50~150s)		
Allow Uncoordinated?				2	5	10
Never (0)	Offset opt.	Quick	PI	384	400	400
		Medium	PI	380	380	380
		Extensive	PI	384	386	381
Rarely (30)	Offset opt.	Quick	PI	393	394	394
		Medium	PI	393	397	400
		Extensive	PI	393	395	395
Sometime (50)	Offset opt.	Quick	PI	393	399	400
		Medium	PI	393	395	400
		Extensive	PI	392	395	395
Often (70)	Offset opt.	Quick	PI	469	469	469
		Medium	PI	468	468	468
		Extensive	PI	468	468	468

Effects on Solution Quality

Table 4 provides solution results of the final signal timing plans on basis of all scenarios in Table 3 during the sensitivity analyses process. That is to say, the results in Table 4 are achieved by optimizing the offsets and lead/lag phasing based on the optimal cycle length derived from Step 3 whose results have been shown in Table 3. As mentioned before, it is described that in SYNCHRO, a quick offset optimization is set as default and stated as usually sufficient to determine how the cycle length will perform. As such, the main purpose of conducting such sensitivity analyses and comparing results in Tables 3 and 4 is to see whether a quick offset optimization is sufficient or not as stated by SYNCHRO.

As can be seen, when intersections are allowed uncoordinated to a certain extent (ranging from Rarely, to Sometime, to Often), the PIs of the final optimized signal timing plans derived from the quick offset optimization are extremely close to (if not exactly the same as) those from either medium or extensive offset optimization, regardless of the chosen cycle length increment (i.e., whether it is 2, or 5, or 10 seconds). On the other hand, if intersections along the arterial must be coordinated (i.e., never allowed uncoordinated), the difference among PIs of the final optimized signal timing plans derived from the quick, medium and extensive offset optimization will become increasingly significant as the cycle length increment increases (with the difference being minimum if cycle length increment chosen as 2 seconds). This indicates that to derive the best arterial signal timing plans, extensive offset optimization rather than just the quick offset optimization command may be warranted in Step 3 (Optimize Network Cycle Length) although this may increase the computational time. However, considering that developing signal timing plans is normally conducted offline, it is worthwhile of doing so or at least trying the extensive offset optimization in Step 3.

Results Comparison among Three Different Partition Techniques.

The results comparison comprises two perspectives: Coordination Desirability and Solution Quality Comparisons. Table 5 provides coordination desirability comparisons among all three different partition techniques. Note that the coordination desirability of each links connecting two neighboring intersections are calculated based on the methodology and equations presented in Section 2. Based on the rule of thumb shown in Table 1, the coordination desirability results are included in parenthesis. As can be seen, the results from these three different partition techniques differ only slightly.

To see whether the network partition in Step 2 can possibly improve the solution quality, a series of SYNCHRO runs are also conducted. It is observed that there are no differences in the partition solutions based on the coordination desirability among all three different partition techniques for all scenarios as shown in Tables 3 and 4. This is perhaps due to that the internal cycle length optimization methodology embedded in SYNCHRO may still recommend individual intersections belonging to the same zone to be operated independently even they are partitioned into different zones (Husch, D. and Albeck, J. 2006).

Summary And Future Research

Traffic signal coordination has been an accepted and proven practice throughout the United States. It can significantly reduce overall stops, travel delays, and vehicle emissions as well as improve air quality by allowing for large groups of vehicles to efficiently flow through a series of traffic signals without stopping.

Table 5. Coordination desirability comparisons.

Link		Methodology		
Origin	Destination	Coupling Index	Strength of Attraction	Coordinatability Factor
Amherst St.	Loop 323	1.3 (N.D.[a])	0.1 (N.D.)	31 (D.)
Loop 323	Independence Rd.	8.9 (D.[b])	0.8 (D.)	60 (D.)
Independence Rd.	Shiloh Rd.	12.6 (D.)	1.1 (D.)	68 (D.)
Shiloh Rd.	Rieck Rd.	5.0 (D.)	0.4 (N.D.)	64 (D.)
Rieck Rd.	Donnybrook Rd.	7.4 (D.)	0.7 (D.)	67 (D.)
Donnybrook Rd.	Grande Blvd.	6.3 (D.)	0.5 (D.)	54 (D.)
Grande Blvd.	R.E. Lee Drive	42.0 (D.)	3.7 (C.[c])	80 (C.)
R.E. Lee Drive	Southtown Dr.	9.9 (D.)	0.9 (D.)	64 (D.)
Southtown Dr.	Heritage Rd.	2.2 (D.)	0.2 (N.D.)	53 (D.)

[a]N.D. = Not Desirable.
[b]D. = Desirable.
[c]C. = Critical.

This paper intends to investigate two research issues regarding arterial signal coordination and traffic progression that have not well addressed thus far: 1) It presents a sensitivity analysis of determining the optimal network cycle length under various combined options chosen among increment in cycle length, whether and how to allow uncoordinated, and ways to conduct offset optimization; and 2) It compares the effectiveness of three different network partition techniques which can be potentially incorporated into software packages.

A case study is undertaken using a real world arterial in the City of Tyler. Sensitivity analyses of the optimized network cycle length are conducted. It is noticed that the computational time decreases significantly in the order of changing from Never (0) to Often (70), everything else being the same. Likewise, the computational time decreases noticeably as cycle length increment increases from 2 to 5 to 10 seconds if everything else is kept the same. Also, the more extensive the chosen offset optimization, the more the computational time. It is also concluded that using the Quick Offset Optimization command during the process of optimizing cycle length may not be sufficient enough to find the best signal timing plans and this is particularly true if intersections are allowed uncoordinated to a certain extent (ranging from Rarely, to Sometime, to Often). This paper also compared three different partition techniques including Coupling Index, Strength of Attraction, and Coordinatability Factor. It is found that the coordination desirability results differ only slightly and that all three partition techniques can produce the same solutions in all scenarios.

Further research is expected to compare the three different signal partition techniques considered in this paper with the other one proposed before (Tian, Z. and Urbanik, T,

2007) for arterial signal timing. Last but not least, we need to test on more arterials and networks to make our conclusions more general because the results presented in this paper may depend on the case studies and therefore may not be general and transferable.

Acknowledgements

The authors would like to thank Justin Smith of Traffic Engineering Department at the City of Tyler for his generosity in sharing the traffic data regarding the South Broadway arterial and making this paper possible.

References

Alam, S.M.M. "Evaluation of signal timing optimization software: synchro and passer ii." Master Thesis, University of Texas at El Paso, El Paso, TX, July 2001.

Chaudhary, N.A., Kovvali, V.G., and Alam, S.M.M. "Guidelines for selecting signal timing software." http://tti.tamu.edu/documents/0-4020-P2.pdf Accessed June 25, 2008.

Chang, E.C.P., Messer, C.J., and Garza, R.U.(1988) "Arterial signal timing optimization using PASSER II-87." *ITE Journal* 58(11), 27–21.

Gartner, N.A.(1991). "MULTI-BAND approach to arterial traffic signal pptimization." *Transportation Research Part B: Methodology* 25(1), 55–74.

Hook, D. and Albers, A.(2008) "Comparison of alternative methodologies to determine breakpoints in signal progression."http://www.trafficware.com/assets/pdfs/00029.pdf, Accessed June 25, 2008.

Little, J.D.C., Kelson, M.D., and Gartner, N.(1981). "MAXBAND: A program for setting signals on arterials and triangular networks." *Transportation Research Record* 75, Transportation Research Board, Washington, D.C. 40–46.

Lan, C.J., Messer, C.J, Chaudhary, N.A., and Chang, E.C.(1992). "Compromise approach to optimize traffic signal coordination problems during unsaturated conditions." *Transportation Research Record* 1360, Transportation Research Board, Washington, D.C. 112-120.

Messer, C.J., Whitson, R.H., Dudek, C.L., and Romano, E.J.(1973) "A variable-sequence multiphase progression optimization program." *Highway Research Record* 445, Transportation Research Board, Washington, D.C. 24–33.

PASSER V User Guide. Texas Transportation Institute, College Station, TX, (2002).

Sripathi, H.K., Gartner, N., and Stamatiadis, C.(1995) "Uniform and variable bandwidth arterial progression schemes." *Transportation Research Record* 1494, Transportation Research Board, Washington, D.C.135–145.

TRANSYT 7-F User's Manual; V 10. University of Florida, Gainesville, FL, (2006).
Husch, D., and Albeck, J. Synchro Studio 7: Traffic Signal Software - User Guide. Sugar Land, TX, June 2006.

TSIS User's Manual; V 5.1. Federal Highway Administration, Washington, D.C. 2003.

Tian, Z. and Urbanik, T.(2007). "System partition technique to improve signal coordination and traffic progression." *Journal of Transportation Engineering* 133(2), February, 119-128.

Yang, X.K. (2001). "Comparison among computer packages in providing timing plans for Iowa Arterial in Lawrence, Kansas." *Journal of Transportation Engineering* 127(4), 314–318.

Wallace, C.E. and Courage, K.G.(1982). "Arterial progression: new design approach." *Transportation Research Record* 881, Transportation Research Board, Washington, D.C. 53–59.

Green Wave Control of Traffic Signal of Inducible Pre-formation at Section of Road

Shuqiao Chen[1] and Jingyi Sun [2]

[1] Associate Professor, [2] Associate Professor, Faculty of Transportation Engineering, Kunming University of Science and Technology, Yunnan Province, China, Email: chen9r@qq.com, sunjingyi2005@yahoo.com

Abstract

To realize the traffic control at the crossroads is by the red and green signal lights. In order to improve capacity of intersection, the green wave band is a useful signal control mode. This paper proposes a new kind of green waves display and control mode. There are signal lights along the road form the moving light zones, and the moving speed and length of light zones are determined and made coordinative of the signal cycle of intersection by the computer system. The driver can adjust his speed in time to join the green light zone, remain consistent with the speed of light moving and will pass the next intersection without waiting.
Key Words: green wave; pre-formation; visualization

Introduction

The red and green signal lights at the crossroads are important traffic management methods. Stop at a red light and go at the green light is a basic traffic regulation. With the increasing of vehicle quantities, the traffic congestion in modern society has become one of the most serious problems. Usually the traffic pressure concentrates on the intersection. In order to improve the capacity of intersection, green wave band is a good way. Of course in this situation, the car will pass a series of crossroads at green lights or be blocked before a series of red lights. Because there are many factors will influence the green wave band, such as pedestrians crossing the street. So we have to consider using intelligent dynamic traffic signal control to realize the green wave band.

Literature Review

The optimized studying of the green wave band mainly focus on the intersection, such as the reference paper: Study on synergy problem of dynamic lane-use assignment and the signal control at intersections (Ying Zeng, Xiaoguang Yang, Yingying Ma, 2009), established an integrated model for lane—use assignment and signal phase. There is no any studying about integrating traffic flow before passing the intersection.

The Devices Structure and Working Process of Signal Control of Pre-formation

The devices of traffic signal control of pre-formation are comprised of land coils, the display device of red and green lights along the lanes, intersection signal controller and control computer system, etc.

Along the lanes, between the adjacent intersections intervallic set underground front test coils and behind test coils, which connect with the control computer executing green wave band coordinated control program. And the computer connects with intersection signal controller which controls the crossroad signal lights and the lights along the lanes. These lights along the lanes are installed in the column of road separator barriers, and where sets the solar panels and battery charging controller, they could provide the road light display device with energy.

The underground front test coil and behind coil are set each 10 meters spaced. Each group solar LED guiding lights along the lanes is about 20 or 30 meters long, and in a group each LED lights installed on the column is paralleled. The intersection and other intersection's controller, there are processing data computer, is connected with all group solar LED lights in this section.

The working process: when the traffic flow pass through the underground coils, the data of flow, such as number, position and speed of vehicles, etc, are transmitted to the computer collection plate. Each traffic direction has a group control signals, and in a intersection has eight group signals.

The traffic data is handled by the signal control computer system and applied to get the signal timing of green wave plan. Not only it controls the signals of each entrance lane of intersection, but also controls the LED lights along the road. The LED lights will form a lights band of alternately red and green which moves to the intersection in appropriate speed. Due to the line is too long sometimes, about 0.5 to 1.5 km , and will cause the lights band signal attenuation, the lights band signal could be strengthen and amplification by solar energy which are deployed in the LED guiding lights.

In order to save energy, solar LED light will have tandem sonic switch. No traffic situations, the switch is close. Otherwise, the sonic switch open and solar LED light will get energy and work.

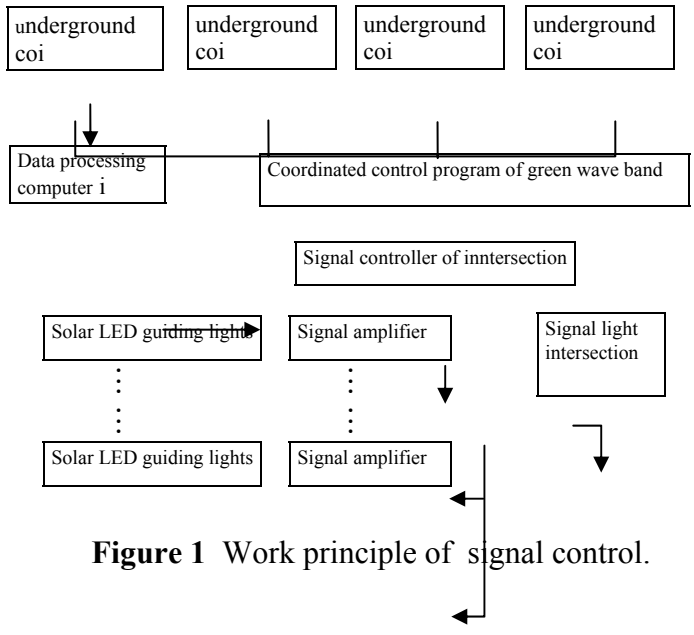


Figure 1 Work principle of signal control.

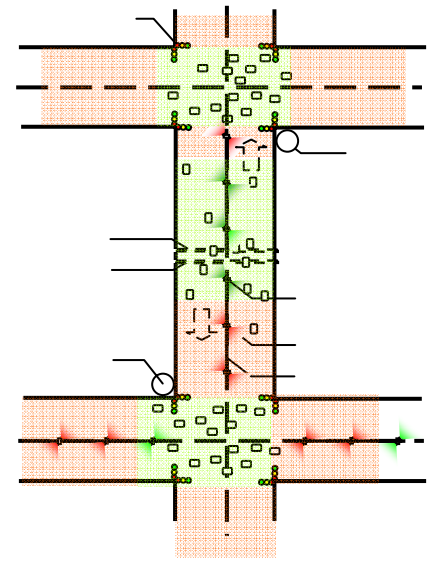


Figure 2 Signal control schematic device of pre-formation.

Driver could feel his position in green light zone or in red light zone during moving. The driver can adjust his speed in time in order to join the green light zone, and remain consistent with the speed of light moving. So he can through the next intersection without waiting. If there is no enough space to cut in green light zone, he need wait for the next period. We can restrict same driving behavior to improve the safety. This process is called "pre-formation". According to the traffic situation, the green wave band will automatically coordinate, and the signal cycle needn't keep fixed.

If the team is disturbed, the LED lights along the road could guide the traffic flow to form a new formation, and the signal control of the next intersection real-time adjust, so could achieve maximum capacity of intersection.

The Characteristic of Traffic Signal Control of Pre-formation

This control method applies to the road condition that could not solve traffic problem through renovation and expansion of road to improve traffic capacity.

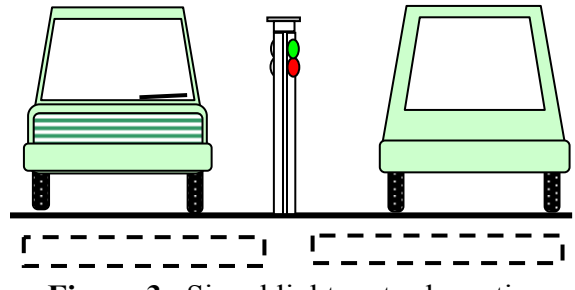


Figure 3. Signal lights set schematic.

(1) Save reconstruction cost.
(2) Dynamic adjust: The buried coils between two crossroads could real-time detect traffic information. And it is inputted into the control computer and output the signal plan to the controller in intersection. Then the dynamic adjustment of signal control could improve the efficiency of green wave band control mode, and improve traffic capacity of intersection.
(3) Reduce the pressure of driving. When the car is running in the green zone, and remains consistent with the speed of green light moving, the car will trough the intersection in the green wave band. If in the red light district of section, and with the flow velocity of red lights, then the car will meet the red light at the intersection. So the drivers can directly know whether they will meet the red light at intersection on the section, and drivers can adjusted theirs speed freely.
(4) Effect of pre-formation. The different between this intelligent green wave control and the foundational is intuitively presented signal cycle at section before arriving intersection. The signal lights along the road form the moving light zone. The computer program determines the signal lights move speed and length of the green or red lights zone, and make it coordinative of the signal cycle of intersection and the section light zone. The pre-formation could sure the efficiency of traffic to pass intersection, because the traffic flow could adjust the speed and position in order to join the green wave band, and it supply a possible of time and space to adjust running state.

The Basic Principle of The Formation

After a basic mathematical analysis for the characteristics of formation, we can get the following rules:
1 basic formula: set S is length of the formation, set L is the distance between the adjacent teams, should meet the formula:

$$L > S \tag{1}$$

So that can ensure the teams will not interfere each other, and there is enough space that the different direction teams to pass the intersection.
2 Symmetry principle: The traffic flow of counter lane should have same phase and same speed when pass the intersection. Set V_{a0} is the speed of direction a at the center of the intersection, V_{a0} is the speed of direction b at the center of the intersection, the formula is:

$$V_{aO} + V_{bO} = 0 \tag{2}$$

Speed is the vector data, the sum is 0. The direction of two speeds is opposite and the number is equal.
3 length of formation: set L is the distance between the two formations, according to the law of symmetry, so

$$L_a = V_a \cdot \Delta t_a \equiv L_b = V_b \cdot \Delta t_b . \tag{3}$$

The interval time of no traffic on the road Δt is

$$\Delta t = \frac{\Delta L_a}{V_a} = \frac{\Delta L_b}{V_b} \tag{4}$$

Set L is the length of the formation, set the pass intersection time of formation is t, then

$$t = \frac{l}{V} < \Delta t \tag{5}$$

This condition sure the formation will not conflict.

(4) Add crossing facilities: The pedestrian crosswalk facilities are always set in intersection. Because pedestrian crossing are easy controlled by the signal system. In this intelligent green wave control there is some red lights zone can be used at section. It could reduce the green time for pedestrian at intersection

(5) Speed stability requirement.

(6) Formation principle of right turn and left turn. There is three kinds of formation: straight formation, left turn formation, and right turn formation. In these three formations, straight and left turn flow will conflict, so that they need different phase to pass intersection. The right turn and light turn can use same phase, as Figure 4.

The Traffic Signal Phase Analysis

At the signal cycle, the plan is turn first and straight late. When the green signal begin, light turn and right turn flow pass the intersection ahead, and run after the green zone of crossed road. Then the straight traffic passes the intersection until the end of signal phase. At the graph c and d of figure 4, it is the two phase of turn flow and straight flow of crossed road. There need point out specially that it is beneficial for green zone formation that turn flow can accelerate pass.

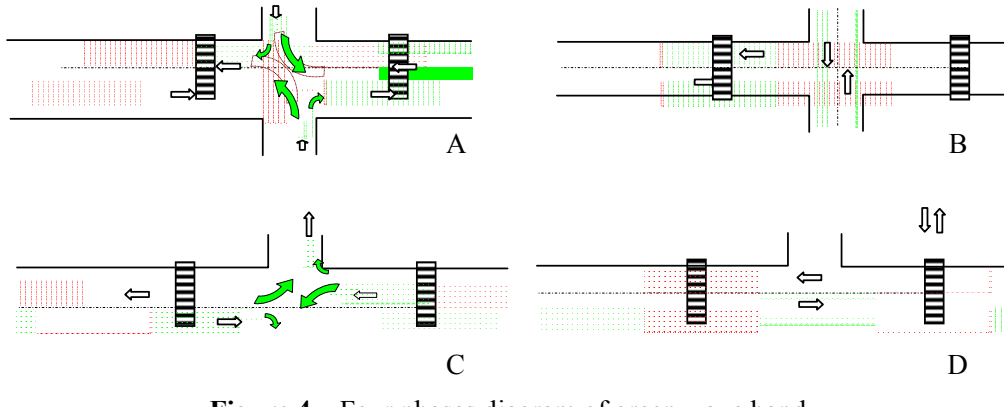


Figure 4 Four phases diagram of green wave band.

Conclusions

This intelligent green wave control system can improve capacity of intersection, especially improve the pass efficiency at green wave band control system where

traffic arriving is uneven distribution. The characteristic of pre-formation is a speed induced along section, and can make full use of green time. The plan is useful in the small and medium-sized cities, where traffic isn't saturated usually and the interval between intersections is short.

In this paper, the new traffic signal control mode is in the early stage of research study. The applied research is in the traffic simulation currently. Further research content include: the influence to driving behavior of pre-formation, the influence of traffic, and the influence to pre-formation of disturb factors. In the mathematical modeling, further research will focus on: to confirm the relation between the length of pre-formation and speed, and to sure the moving speed of light along the road, etc.

References

Ma, W. J., Nie, L., Yang, X. G. (2007). "Ootimization model and simulation analyses of pre-timed signals control at isolated intersections." *Journal of System Simulation*, 4543-4547.

Ma, W. J., Yang, Yang, X. G. (2008). "Coordination design of left movements of signalized intersections group." *Journal of tongji university*(*natural science*), 1507-1511.

Ma, Y. Y., Yang, X. G., Zeng, Y. (2009). "Multi-objective cycle length optimization model and solution." *Journal of tongji university*(*natural science*), 761-765.

Lao, Y.T., Yang, X. G., Yun, M. P., Liu, J. Y. (2006). "Evaluation of Traffic State Detection Method." *Transport and computer*, 74-77.

Zeng, Y., Yang, X. G., Ma, Y. (2009). "Study on synergy problem of dynamic lane-use assignment and signal control at intersections." *Journal of tongji university*(*natural science*), 907-908.

Optimizing Signalized Roundabout by Computer Simulation (Case study: Pelem Gurih Roundabout, Yogyakarta)

Ahmad Munawar [1]and Sigit Haryanto[2]

[1] Prof. in Transportation, Department of Civil and Environmental Engineering, Gadjah Mada University, Indonesia, Email: munawarugm@yahoo.com
[2] Head of Transport Division, Yogyakarta Provincial Government

Abstract

Pelem Gurih Junction is situated in Yogyakarta, Indonesia. This three arm signalized junction is very congested during peak hour. Therefore, it is planned to be modified into signalized roundabout. A computer simulation program: KREISIG (Kreisverkehrsplaetze mit Lichtsignalanlage=signalized roundabout), which was developed by Munawar (1994) in Germany has been employed to optimize the signal. This program can calculate the queues and delays at signalized roundabout. However, this KREISIG program is based on the German driver behavior, which is quite different from Indonesian driver behavior. Therefore, the sensitivity parameters and the yellow time decision of the drivers should be calibrated first. The saturation flow output has also been compared to 1997 Indonesian Highway Capacity Manual. Trial and error method has been employed to adjust the saturation flow. Turbine method has been used for signal setting. Then, the KREISIG program was run to optimize the signal setting. This program has been succeeded to optimize signal setting at this planned signalized roundabout.
Key Words: signalized roundabout; simulation; driver behavior

Introduction

Capacities, queues and delays at intersections may be predicted by the use of empirical, mathematical and simulation approaches. Empirical models attempt to predict highway capacity on the basis of pas observations, whereas in the mathematical approach, vehicular delay is calculated, and capacity or level of service defined in terms of the flow occuring when delays reach a certain value. In contrast, a simulation model can be flexible enough to cover a wide range of highway and traffic conditions. Input to the model can be specified to any distribution and the

form of traffic control and driver characteristics varied. Simulation time may be as long as desired and several figures of merit may be printed out during simulation process. A simulation model requires the formation of a model system which represents the real situation at the site being studied. This model has been developed by many researchers in the recent years, such as Hewage and Ruwanpura (2004).

KREISIG computer program (Kreisverkehrsplaetze mit Lichtsignalanlage=signalized roundabout) was firstly developed by Munawar (1994) in Germany, to calculate queues and delays at signalized roundabout. It has been implemented to optimize signal timing at signalized roundabouts in Hagen, Berlin, Hidelsheim and Frankfurt.

Methodology

Flow chart of the simulation program is shown in Figure 1. The program contains:

- Data inputting system
  a. Geometry data: diameter (D), and entry width for each arm.
  b. Traffic data: traffic counting data for each type of vehicle.
  c. Traffic fluctuation, to generate time varying flow.
- Random generator: to generate random numbers to be used for simulation, i.e.: time headway, vehicle type, vehicle direction, desirable speed, driver sensitivity, acceleration, distance at stop line, overtaking decision, and driver decision at yellow time.

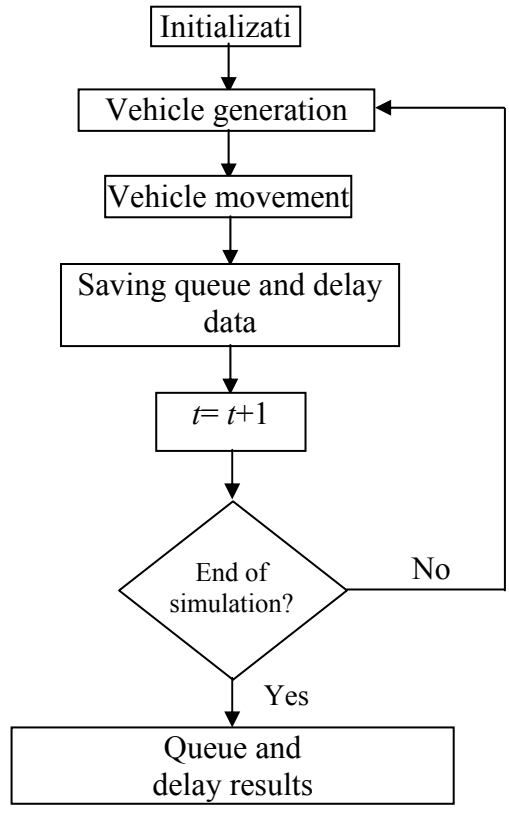


Figure 1. Simulation flow.

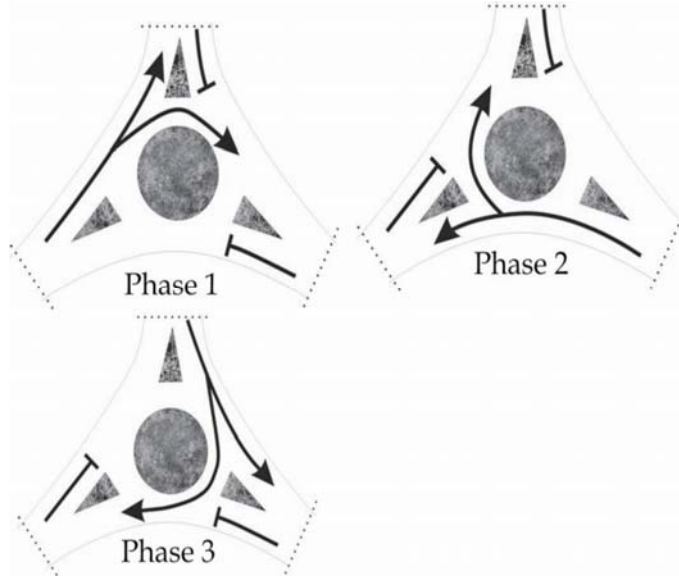


Figure 2. Turbine method.

Turbine method (anti clock wise green time phasing) has been employed to optimize the signal setting (Figure 2). This turbine method can minimize the lost time during intergreen period. For time headway, Erlang distribution has been employed for this simulation. When approaching the stop line, simulation of vehicle movement is based on car following theory (May, 1990). Basic equation is as follows:

$$\text{Reaction } (t+\Delta t) = \text{sensitivity x stimulant} \quad (1)$$

Two vehicle movement, i.e. vehicle (n) and vehicle (n+1), approaching the stop line, can be explained in Figure 3.

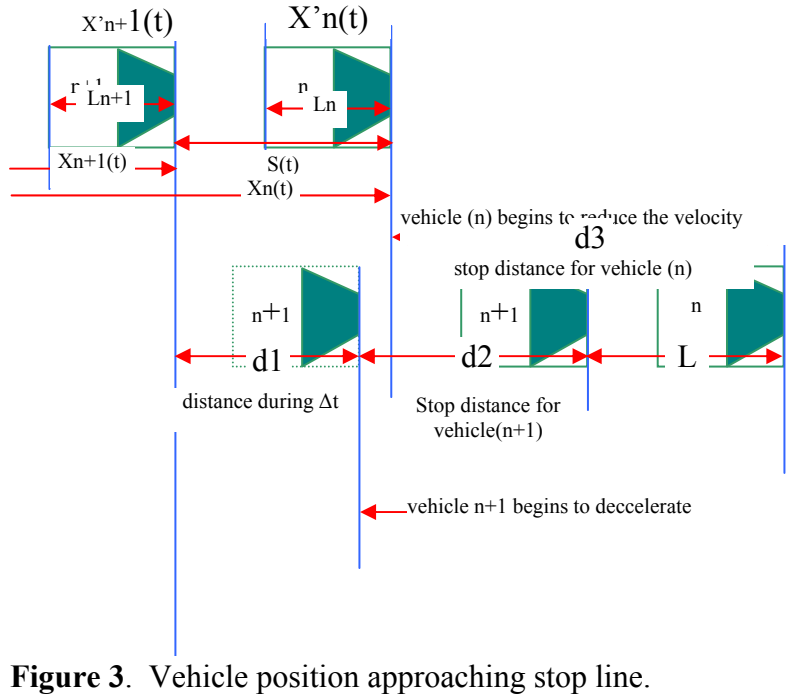


Figure 3. Vehicle position approaching stop line.

where,
$X'_n(t)$= velocity of vehicle n at time t (m/s)
$X_n(t)$ = position of vehicle n at time t (meter)
$X'_{n+1}(t)$ = velocity of vehicle n+1 at time t (m/s)
$X_{n+1}(t)$ = position of vehicle n+1 at time t (meter)
Ln= length of vehicle n (meter)
Ln+1= length of vehicle n+1 (meter)
S(t) = distance between vehicle n and n+1 at time
t = $X_n(t) - X_{n+1}(t)$ (meter)
d1= distance of vehicle (n+1) after moving during Δt
= $U_{n+1}(t)$
d2= distance of vehicle (n+1) after moving
= $[U_{n+1}(t+\Delta t)]2/2a_{n+1}(t+ \Delta t)$
L= distance between vehicle (meter)
$U_i(t)$ = velocity vehicle i at time t (m/sec)
$a_i(t)$ = acceleration of vehicle i at time t (m/sec^2)
Δt= reaction time (sec)

Acceleration of vehicle is shown in formula below:

$$X''_{n+1}(t+\Delta t) = \frac{\alpha o\left[X'_{n+1}(t+\Delta t)\right]^m}{\left[X_n(t) - X_{n+1}(t)\right]^l}\left[X'_n(t) - X'_{n+1}(t)\right] \quad (2)$$

where,
$X''_{n+1}(t+\Delta t)$ = acceleration of vehicle n+1 at time $(t+\Delta t)$ (m/sec^2)
$X'_{n+1}(t+\Delta t)]^m$= velocity of vehicle n+1 at time $(t+\Delta t)$ (m/s)
$X_{n+1}(t)$ = position of vehicle n+1 At time t (meter)
Δt= reaction time (sec)
α_o= sensitivity
l= distance component of sensitivity
m= velocity component of sensitivity

According to Hoefs (1972) and Koehler (1990): $\alpha o = 17$, m = 0,2 dan l = 0,5

According to the field survey, it is found that:

- Indonesian driver begins to deccelerate the speed when approaching the stop line during the red signal at a distance of 75 meters from the stop line. It can be compared to the existing KREISIG program, which is based on the German driver behavior, i.e. German driver begins to deccelerate the speed at a distance of 250 meters from the stop line. It is shown that Indonesian driver takes more risks than German driver.
- Most Indonesian drivers still pass the stop line during the yellow time, whether most German drivers decide to stop during the yellow time.
- According to the calibration (by trial and error), so that the saturation flow for vehicle crossing the stop line according to the simulation result should be the same as the actual saturation flow in the field and as calculated by using 1997 Indonesian Highway Capacity Manual (Munawar, 2006 and Directorate General of Highways, 1997). it is found that:

a. αo for decceleration = 140,
b. αo for acceleration = 0,2,
c. m (distance component of sensitivity) = 0,2,
d. l (velocity component of sensitivity)= 2,0
e. initial acceleration at the stop line (at the beginning of green time) for:
1) car = 2.2 x a_{max}
2) heavy vehicle = 1.5 x amax
3) very heavy vehicle = 1.0 x amax
4) motor cycle = 2.3 x amax

while a_{max} = maximum normal acceleration, which also depends on the type of vehicle.

Case Study: Pelem Gurih Junction

Pelem Gurih junction is situated in Yogyakarta, Indonesia. The geometry is shown in Figure 4 below. There are many conflicts and accidents because of the short intergreen period. However, increasing intergreen period will reduce the capacity.

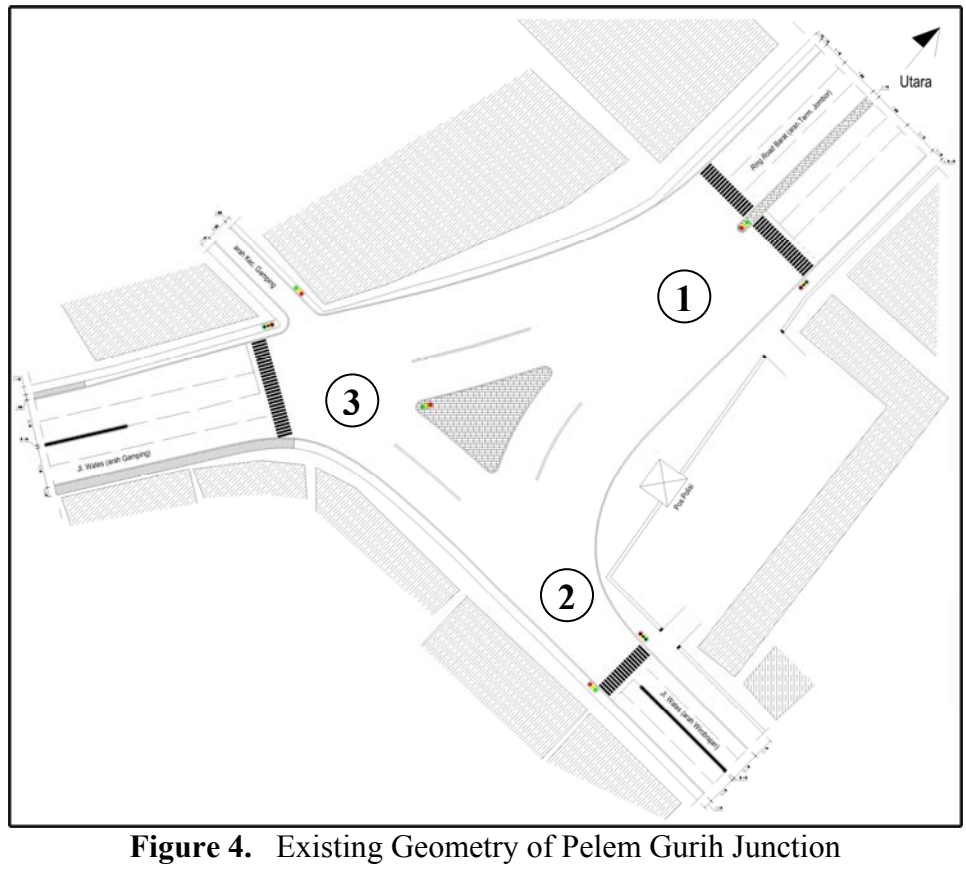


Figure 4. Existing Geometry of Pelem Gurih Junction

Table 1. Traffic flow for afternoon peak (in vehicles/hour and passenger car units/hour).

From arm	To arm					
	1		2		3	
	v/hr	pcu/hr	v/hr	pcu/hr	v/hr	Pcu/hr
1	-	-	343	120	396	240
2	143	64	-	-	959	387
3	660	302	2992	1006	-	-

The traffic flow for afternoon peak is shown in Table 1. Actually, this junction has 4 arms, but the fourth arm is very small, the road is very narrow and only a very few traffic flow on this road. Therefore, the signal for this arm has been set the same as arm 2.

Table 2. The existing green time and intergreen (for afternoon peak) cycle time 45 seconds.

Arm	Green time period	Intergreen (seconds)
1	41 – 6 (10)	+2
2	8 – 18 (10)	+2
3	20 – 39 (19)	+2

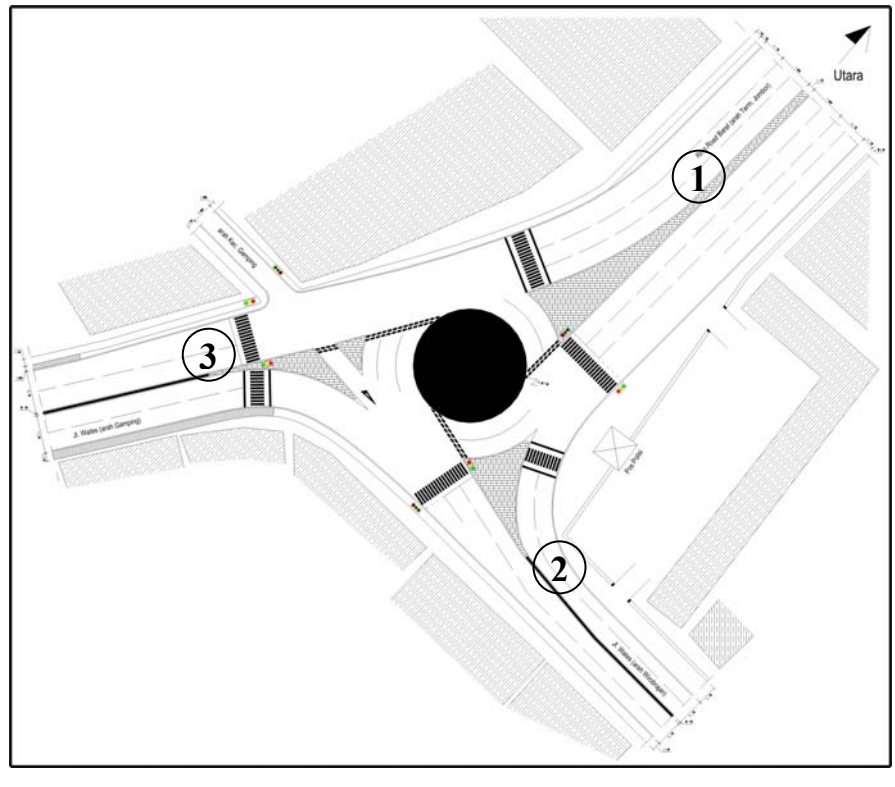


Figure 5. Proposed Pelem Gurih Roundabout

Table 3. Green time and intergreen as signalized roundabout for afternoon peak period (cycle time 45 seconds).

Arm	Green time period	Integreen (seconds)
1	41 – 8 (12)	0
2	29 – 41 (12)	0
3	8– 31 (23)	-2

Tabel 4. Average delay (s) for each arm.

Arm	Average delay (s)
1	12
2	9
3	13

KREISIG computer simulation program has been employed to optimize the cycle time. Various cycle times, from 40 seconds to 100 seconds have been simulated to find the minimum average delay. The optimized green time, intergreen and cycle time are shown in Table 3 below. The short (or negative) intergreen period will not give any conflict because of the turbine solution (anti clock wise green time phasing). The average delay for each arm is shown in Table 4. It is shown that there is no significant delay for each arm.

Conclusion

This modified KREISIG program has been successfully used to plan signalized roundabout at Pelemgurih Junction to increase the safety and to reduce the conflict without reducing the capacity.

References

Directorate General of Highways. (1997). Ministry of Public Works, Indonesia, *Manual Kapasitas Jalan Indonesia 1997,* (*1997 Indonesian Highway Capacity Manual*), Jakarta.

Hewage K. N. and Ruwanpura, J. Y. (2004). "Optimization of Traffic Signal Light Timing using Simulation," *Proceeding of the 2004 Winter Simulation Conference*, Monterey, Canada.

Hoefs, D. H. (1972). "*Untersuchung des Fahrverhaltens in Fahrzeugkolonnen*," Bonn, Bundesminister fuer Verkher, Abt. Strassenbau, 140.

Koehler, R. "*Verkhersablauf auf Binnenwasserstrassen – Untersuchungen zur Leistungsfaehigkeitsberechnung und Reisezeitverkuerzung*," Karlsruhe, Schrifenreihe des Instituts fuer Verkehrswesen, Universiaet Karlsruhe, 1968.

May, A. D. *Traffic Flow Fundamental*, New Jersey, Prentice Hall, 1990.

Munawar, A. (1994). "Entwicklung eines Simulationsmodells fuer Kreisverkehrsplaetze mit Lichtsignalanlagen." Schriftenreihe des Lehrstuhls fuer Verkehrswesen, Ruhr Universitaet Bochum.

Munawar, A. (2006). "Queue and delay at signalized intersection, Indonesian experience." *Proceeding of the Symposium on Highway Capacity Manual and Quality of Service*, Yokohama.

Dynamic Traffic Flow Equilibrium and Coordinated Signal Control for Ramps and Adjacent Intersections

Lixia Bao[1] , Xiang Gao[1], Weiquan Zhu[2] and Hong Chen

[1] Senior Engineer, Dr., Intelligent Transportation Systems Research and Development Center, Shanghai Urban Construction and Design and Research Institutes, Shanghai 200125, Email: jlublx@163.com
[2] Dr., College of Transportation Engineering, Tongji University, Shanghai, China

Abstract

Ramp congestion significantly hampered travel efficiency into urban expressways; about 30% of the total travel time is taken waiting at the expressway entrance during rush hours. This study is specifically concerned with ramps and an adjacent intersection, area traffic flow characteristics are analyzed, and the dynamic traffic flow equilibrium model is discussed with the objective to maximize travel capacity. Compared to a standalone approach, coordinated on-ramp and adjacent intersection control is presented, which has the advantage that cycle and phase information of upstream traffic signal can be used in the algorithms for ramp metering optimization. In analysis of the basic difficulties and limitations of theoretical models as well as continuous or interrupted traffic flow model during the practical application, innovative algorithm solving steps based on corridor traffic state is also discussed. With the model calculation results, a collaborative optimization control method to enhance capacity and alleviate congestion is applied to ramp metering and signal timing. Mostly, as a demonstration project of Shanghai expo ITS integrated system, a traffic flow model and coordinated control strategy are applied to the case of West Yan'an road with one adjacent on-ramp. A simulation study was conducted to compare on-ramp control performance between coordinated and a standalone approaches. The optimization control parameters are computed, and the simulation evaluation results show that with the coordinated control, traffic capacity increases by 18.63%, and the vehicle delay decreases by 21.3%.

Key Words: dynamic traffic flow equilibrium model; on-ramp metering; adjacent intersections signal timing; coordinated control

Introduction

Existing road network capacity can no longer meet the increasing traffic demand due to rapid socio-economic development and a dramatic increase in car ownership. Shanghai Beijing and Shenzhen as well as many mid-scale cities in China have built an urban expressway around the business district to ease traffic pressure. Other cities are also quickly constructing roads to reduce traffic congestion.

According to "Shanghai urban expressway operation monthly report," the capacity of urban expressway elevated upon the ground for the same length is 40% higher than the urban roads with several intersections during normal traffic situations because of fewer signal-controlled intersections. More and more drivers prefer the urban expressway especially in their trips through the city. Mainline and on-off ramps are primary parts of the expressway. The vehicles enter the expressway mainline through on ramps and leave by off ramps. So on-off ramps create a bottleneck on the total urban expressway system.

Recently, ramp congestion significantly hampered travel efficiency into urban expressway; about 30% of the total travel time is taken up going onto the expressway in rush hours. In Shanghai's 88 on-ramps, about 40% have a recurrent congestion phenomena during peak hours, more importantly, congestion will cause vehicles queuing extending to upstream approaches and the adjacent intersections. Regional congestion for an area of on and off entrance ramps around intersections is quite serious.

With the development of the Intelligent Transportation System in the world, more intelligent traffic control technologies are applied in traffic management based on traditional traffic flow theory. Dynamic Traffic Assignment (DTA) and area coordinated optimization control are effective methods in order to restrict the basis of supply and demand balance and optimize the traffic flow. A DTA model is any model that assigns a time-varying origin-destination demand to a transportation network in a time-varying way.

For the traffic problem in the area of on-ramps and intersections, DTA is considered as theory basis, while ramp metering and signal control are taken as effective traffic management strategies to maintain an efficient expressway system and keep the systems operating at or near capacity. Ramp metering helps regulate the mainline traffic flow and keep the traffic in equilibrium. Area intersection coordinated signal control is necessary to protect traffic safety and ensure that road capacity is at a maximum. This study is part of the work of the Shanghai Expo Priority Corridor Control (SEPCC) project, which hopes to enhance, develop, and implement new ramp- metering algorithms to increase the efficiency of local metering applications, network wide coordination, and the integration with local network management.

Literature Review

Research on dynamic traffic flow model for coordination and control has been carried out for many years; the basic laws of traffic flow theory as expressed in the

well-known fundamental relationship between traffic flow, speed and density, are often called the fundamental diagram of traffic. In journals, there is much research based on multiple linear regressions, genetic algorithms, etc. to improve the model and the actual validation studies.

Traffic flow in the collaborative research for traffic equilibrium and ramp control both domestic and international can mainly be divided into two categories:

Micro-model. Mohammad Ghanim (2009) proposed expressway roads and associated road network control methods. This method is called METANET microscopic traffic flow model and is used in expressways. The ground intersections adopt a model based on an earlier Kashani improved model to describe vehicles following station and the delay for each vehicle. From Shanghai University, Dr. Lei (2005) used hydrodynamic models and cellular automaton models for his research with urban expressways, ramps, and intersections. To carry out this research, a large number of measured data processing and analyses were obtained for sparse traffic flow and congested traffic flow in two-speed, one-density relation and accordingly speed and congestion relation degree are both used as the two important parameter values.

Macro-model. Papageorgiou (2001) coordinated the on-ramp and intersection of collaborative traffic flow characteristics in the study, through the definition of overcrowding, so that crowding on the ramp intersection with the ground balance of the congestion degree. Tian (2009) in his study presented the convergence of the entrance ramp intersection traffic flow model. It will be the state of traffic on the ramp that is divided into five levels, the queue length of ground intersection is divided into three levels, based on fuzzy theory between the two control strategy choices.

Dr. Zheng (2007) from Tongji University analyzed the traffic demand model. An optimization function is presented based on the traffic capacity model with the minimum goal being to optimize the control coordination, taking into account intersection signal control and expressway on ramp metering. The function is a constrained nonlinear optimal control problem, herein, three control parameters including intersection split, on ramp metering, and entrance metering are optimized for the corridor control. Literatures in dynamic traffic flow equilibrium and coordinated signal control show that there are much research has been conducted for on-ramp traffic problems with adjacent intersections. The optimized models or functions are always presented with some severe supposition conditions; traditional description of continuous traffic flow characteristics, interrupted traffic flow characteristics are using some classic parameters such as volume, speed, and density as well as occupancy. It should be noted that due to the inherent complexity introduced with this type of traffic model, these assignment algorithms are necessarily heuristic in nature, though they are often inspired from well-known solution algorithms for the static assignment problem.

Traffic Flow Equilibrium Model for Ramps and Intersections

Optimized Model. Traffic flow equilibrium models, which comprise a basic theory for assignment region vehicles, are becoming more and more common in the area of dynamic traffic assignment (DTA) modeling. In this paper, equilibrium modeling is the fundamental for traffic flow analysis of the mainline, on ramps, and within the intersection area, and the model solving results are for signal control parameters including ramp metering, cycle, and split.

Firstly, the optimized function goal should be ascertained according to the actual traffic condition. Function goal differs with the different traffic states of the corridor of ramps and intersections.

- In the case of mainline congestion or incident, the road network management hopes to recover expressway traffic; on ramps should be closed and vehicles forbidden to enter. Here, $r_{it} = 0$, to avoid the congestion of the intersection or ground links, the connected intersections signal control should be coordinated to realize maximum capacity.
- In the case of saturation, vehicles on the weaving area are exceeding its capacity, if entering traffic is not restricted, the weaving area will become disorderly and may affect mainline congestion or cause accidents. Then, the on-ramp queue is too long, the vehicles can not enter onto the mainline, and the ramp metering rate is too short or the intersection light is green for too long for inner approaches. For this example, the co-ordination objective is to reduce the queue length but not cause mainline congestion.
- For the normal situation, road network capacity can meet the demand, and the function objective is to reach the maximum traffic demand to effectively utilize the expressway resources. Optimal adjustment of ramp metering rate of convergence intersection signal timing, with the goal making the traffic capacity of the channel actually the largest. In the flow state, the optimization objectives, you can make the channel an average delay time for the smallest of vehicles.

In the actual traffic condition, the off ramp is always free flow without obstruction, so, in the model, only on-ramp metering is considered.

$$J_{1t} = \begin{cases} \min \dfrac{1}{n}\sum_{i=1}^{n} \dfrac{c_{it}(1-g_{it})^2}{2(1-y_{it})} & Q_{et} \geq C_{main} \\ \min\left(\sum_{i=1}^{n} g_{it}c_{it}C_{it} - \sum_{i=1} r_t * c * A\right) & Q_{et} \prec C_{main} \ \ \& \ \ q_{wi} \neq 0 \\ \max\left(Q_{et} + \sum_{i=1} r_t * c * A - \sum_{i=1}^{n} g_{it}c_{it}C_{it}\right) & Q_{et} \prec C_{main} \ \ \& \ \ q_{wi} = 0 \end{cases} \tag{1}$$

$$s.t. \quad \begin{cases} r_{c\,main} < \lambda \\ q_{wi} = \sum_{i=1}^{n} g_{it}c_{it}C_{it} - \sum_{i=1} r_t * c * A \\ \dfrac{Q_{et} + \sum_{i=1} r_t * c * A}{C_{main}} < 0.9 \\ \dfrac{g_{it}c_{it}C_{it}}{C_{i,t}} < 0.9 \end{cases}$$

where Q_{et} —traffic volume in expressway mainline;
r_t —ramp metering at t;
c —signal control cycle of the adjacent intersection;
A—lanes of on ramp;
g_{it} —green split of the direction i approach;
$C_{i,t}$ —capacity of i approach, and
d_s —average delay of on ramp.

Control Parameters Calculation.

- on-ramp queue

In order to avoid the creation of large queues, a queue control policy is employed in conjunction with either of the two local metering strategies. Let $w_{o,\max}$ denote the maximum allowed queue in origin o, then the queue control law takes the form:

$$q_o^w(k_c) = -\frac{1}{T_c}[w_{o,\max} - w(k_c)] + d_o(k_c - 1) \tag{2}$$

- coordinate parameter calculation for intersections:
  The Webster formula is used to calculate signal timing:

$$C_0 = \frac{1.5L + 5}{1 - Y} \tag{3}$$

C_0, the optimal cycle length, *Y*, sum of max congestion degree for each phase, $Y = \sum \max[y_{11},...y_{12},...y_{1m},...y_{n1},...y_{n2},...y_{nm}]$; y_{nm}-congestion degree for m phase at n intersection, L-lost time for a cycle(s).

- ramp metering rate calculation

Select a capacity-demand algorithm to calculate the initial ramp metering rate and suppose the cycle is 3 min. The formula becomes:

$$r(k) = (c_a - q_d(t-1)) * 180/3600 \tag{4}$$

where $r(k)$= ramp metering rate at $Tk \le t \le T(k+1)$ (veh/3min); $q_d(k-1)$= traffic demand of upstream during $(k-1)T \le t \le kT$ (veh/h), and c_a = capacity of downstream(veh/h).

Algorithm Test and Evaluation

This study is part of the work of the Shanghai Expo Priority Corridor Control project. In this paper, the congested area of the West Yan'an urban expressway, Kaixuan on-ramp, off-ramp and adjacent Kaixuan intersection is selected for demonstration of the coordinated algorithm and strategy. Where, recurrent congestion often appears in daytime. For the eastbound traffice when congestion occurs, vehicles are blocked on the mainline and can not leave the road due to Kaixuan intersection's jam. At the same time, Kaixuan's on-ramp metering is sufficient enough to allow a few vehicles to travel through the Kaixuan intersection.

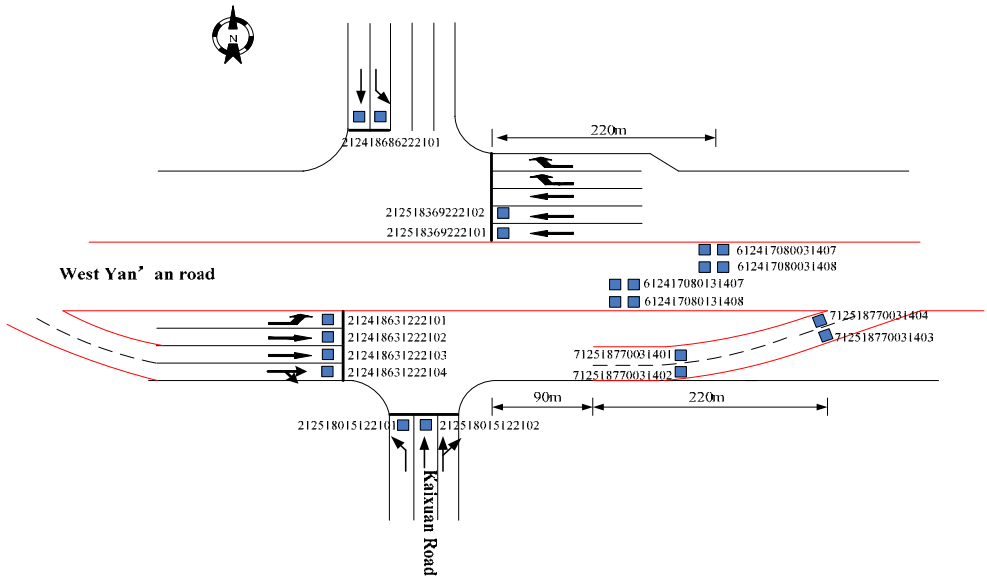


Figure 1. West Yan'an road on ramp and adjacent intersections.

For eastbound traffic, there are two lanes of expressway mainline, and three lanes for the weaving area as well as an on-ramp with two entrance lanes. The distance between the ramps is short; about 400 meters. For the adjacent intersections, there are two lanes for an approach and shorter distance between them. Therefore, a recurrent congestion phenomenon is very serious, and the intersection queue is extended to the ramp.

The actual area showstraffic flow data collected including mainline, off ramp, on ramp and the adjacent intersection for 5 days from 6:00 am to 2:00 pm. The processed data is listed in Table 1.

Table 1. Traffic data collected in the expressway and on ramp.

Mainline and ramps traffic data				
		Q_f(veh/3min)	Vt (km/h)	Delay(s)
Eastbound	mainline	101	57. 5	16. 5
	Ramp1	35	30. 1	39. 2
	Ramp2	39	20. 3	31. 3
Initial ramp metering (r) (veh/3min)				
r1=72				

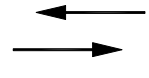
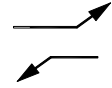
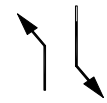
phase sequence	A	B	C	D
phase				

Figure 2. Phase control of the adjacent intersection.

Table 2. Traffic data and basic signal control for the intersection.

		Di(veh/3min)	Link Travel speed(km/h)	Average delay(s)
Kaixuan intersection	approachA	512	15. 6	35. 1
	approach B	496	14. 3	40. 2
	approach C	263	23. 1	29. 5
	approach D	242	26. 2	30. 6
Initial timing:C=120s, ga=35s,gb=30s,gc=25s,gd=15				

Table 3. Evaluation of coordinated against standalone control.

	index		Coordinated control	Standalone control
expressway	Mainline average speed		68. 3	57. 5
	Average delay for mainline		49	35
	Ramp delay		20. 1	31. 3
Adjacent intersection	Delay(s)	Link A	123	176
		Link B	134	182
		Link C	147	193
		Link D	116	145
	Traffic volume (veh/h)	Link A	3512	2963
		Link B	3785	3103
		Link C	2036	1752
		Link D	1853	1496

The on-ramp adjacent intersection uses a signal control in SCATS. The basic phase and the sequence is listed in Figure 2. The intersection is controlled with four phases.

For the adjacent intersection, east-west bound traffic has four lanes, and the north-south bound traffic has two lanes. Traffic data collected for each approach are analyzed along with the signal control parameters in Table 2.

In the test, the we select the time when $Q_{et} \prec C_{main}$ & $q_{wi} = 0$,using the optimization function goal, we set the objective is as follows:

$$J_{1t} = \max(\ Q_{et} + \sum_{i=1} r_t * c * A - \sum_{i=1}^{n} g_{it} c_{it} C_{it}) \quad Q_{et} \prec C_{main} \ \& \ q_{wi} = 0$$

$$s.t. \quad \begin{cases} r_{cmain} < 900\ veh\ /\ h \\ \dfrac{Q_{et} + \sum_{i=1} r_t * c * A}{C_{main}} < 0.9 \\ \dfrac{g_{it} c_{it} C_{it}}{C_{i,t}} < 0.9 \end{cases}$$

The optimization through the above function gets the coordinated control results as follows: (1) for the ramp control, the ramp metering is $r(t) = 626 veh/h$;(2)for the

adjacent intersection, the timing control parameters is: C=120s, ga=38s, gb=27s, gc=23s, and gd=17.

Before the control strategy is applied in a real traffic environment, traffic simulation is conducted, and the evaluation is completed for coordination against with the standalone control. The contrast results are shown in the Table 3.

Conclusions

This study is specifically concerned with ramps and an adjacent intersection. Area traffic flow characteristics are analyzed, and the dynamic traffic flow equilibrium model is discussed along with the objective to maximize travel capacity. In the analysis of the basic difficulties and limitations of theoretical models as well as continuous and interrupted traffic flow model during the practical application, innovative algorithm solving steps based on corridor traffic state is presented. With the model calculation results, the collaborative optimization control method to enhance capacity and alleviate congestion is applied to ramp metering and signal

timing. Mostly, as a demonstration project of the Shanghai expo ITS integrated system, a traffic flow model and coordinated control strategy are applied to the case of West Yan'an road with one adjacent on ramp. The optimization control parameters are computed, and the simulation evaluation results are shown that with the coordinated control, traffic capacity increases by 18. 63%, and the vehicle delay decreases by 21. 3%.

Acknowledgements

This paper is sponsored by the Shanghai Rising-Star Program (B type)(10QB1402400).

References

Ghanim, M., Reynolds Smith & Hills. (2009) "Integration of signal control and transit signal priority optimization in coordinated network using genetic algorithms and artificial neural networks." *88th Annual Meeting (09-3063)*, A12.

Li, L. (2005). *Investigation On Interaction and Characteristics of Traffic Flows in Urban Elevated Road-Ramp-Ground Road System*, Shanghai University, 55-65.

Kotsialos, K and Papageorgiou, M. (2001). "Optimal coordinated ramp metering with Advanced Motorway Optimal Control", *Transportation Research Record 1748*, 55-71.

Tian, Z. Z. (2009). "Safety of special time-of-day protected/permitted left-turn signal control display." *88th Annual Meeting*.

Zheng, J. H., and Chen, H. (2007). "Study on design and simulation of on-ramp fuzzy controller for urban expressway." *Journal of Highway and Transportation*

Research and Development, 55-57

Papageorgiou, M., and Apostolas, D. (2001). "Freeway ramp metering: an overview." *IEEE Intelligent Transportation Research Part C*, 337-352.

Owens, D., and Schofield, M. (1988). "Access control on the M6 motorway evaluation of Britain's first ramp metering scheme." *Traffic Engineering and Control,* 29(12), 616-623.

Ozbay, K., Yasar, I., and Kachroo, P. (2004). "Comprehensive evaluation of feedback-based freeway ramp-metering strategy by using microscopic simulation, taking ramp queues into account", *Transportation Research Record* 1867, Washington, D. C., 2004, 89-96.

Papageorgiou, M., Hadj-Salem, H., and Middelham, F, (1997). "ALINEA local ramp metering summary of field results", *Transportation Research Record,* No. 1603, 90-98.

Sultan, B., Piao, J., and McDonald, M. (2006). "Coordinating on-ramp queue control with adjacent local signal controller", *Proceedings of World Congress on Intelligent Transport System*, 8-12 October, London, UK.

Xin, W., Michalopoulos, P. G, Hourdakis, J., and Lau, D. (2004). "Minnesota's new ramp control strategy, design overview and preliminary assessment", *Transportation Research Record,* No. 1867, Washington, D.C., 2004, 69-79.

Design of Signal Timing Optimization Based on the Unsymmetrical Traffic Flow of Signalized Intersections

Zhongyi Zuo[1], Guangchuan Yang[2] and Chunfu Shao[3]

[1] Associate Prof.r, School of Traffic and Transportation Engineering, Dalian Jiaotong University; No. 794, Huanghe Road, Dalian 116028, China, and Ph.D. candidate, College of Transportation, Beijing Jiaotong University; Beijing 100044, China; Email: zuozy@djtu.edu.cn
[2] Bachelor's Candidate, majoring in Traffic Engineering, Dalian Jiaotong University, No.794, Huanghe Road, Dalian 116028, China; Email: yang-8628@163.com
[3] Prof., College of Transportation, Beijing Jiaotong University; Beijing 100044, China;Email: cfshao@center.njtu.edu.cn

Abstract

The operational condition of traffic flow and capacity of a signalized intersection to some extent depends on the setting of signals; the better the signal timing, the greater the capacity of intersections. This paper analyzed the survey data of Wuyi Road-Southwest Road intersection, Dalian, calculated the current capacity and delays of the intersection, then re-organized traffic flow according to the practical traffic demand and designed an unsymmetrical five-phase signal timing scheme based on the principle of overlap phase. The calculation showed that using the improved signal timing scheme, it will result in a moderate increase in traffic capacity (21.9%) and an significant decrease of delays (53.1%).
Key Words: overlap phase; flow ratio; traffic capacity; delay

Introduction

As the strategic passage of converging, turning, and passing through for various types of traffic, intersections are the control points of traffic management and organization as well as the bottleneck of traffic capacity and traffic safety in the road network systems. Among the numerous factors that influence the capacity of the intersection, the function of signal timing is to distribute the effective green time proportionally to the traffic flow in order to make full use of green time to increase the capacity of the intersection (Quan 1989). As early as the mid-20th century, British

engineer Webster designed a signal timing optimization algorithm that takes the minimum delay of vehicles passing through the intersection as an optimized object (Webster 1966). Most of the new development methods, for instance protected left-turn phase and ARRB method, are based on his theory (Al-Kaisy and Stewart 2001 and Wu and Li 2007). According to the repeated investigation and personal experience of the Wuyi Road-Southwest Road intersection, we found serious traffic congestion during peak hours, and this paper focuses on designing signal timing schemes based on the analysis of practical traffic volume of the intersection for the purpose of easing urban traffic bottlenecks.

Analysis of the Intersection

Geometry and control strategy of the intersection. Southwest Road is a dual three-lane artery of Dalian, with a 4 m wide medium strip and 300 m length in the southern part of the intersection. While Wuyi Road is a dual two-lane road, the right-turn of the east approach isn't necessary through the intersection. The geometry of the intersection is shown in Figure 1.

A four-phase signal controlling scheme is currently used, with alterable signal time associated with different time periods. Field observation shows that the length of signal cycle for morning peak is 183 seconds. Signal time scheme is shown in Figure 2 in accordance with the roads' name and direction in Figure 1.

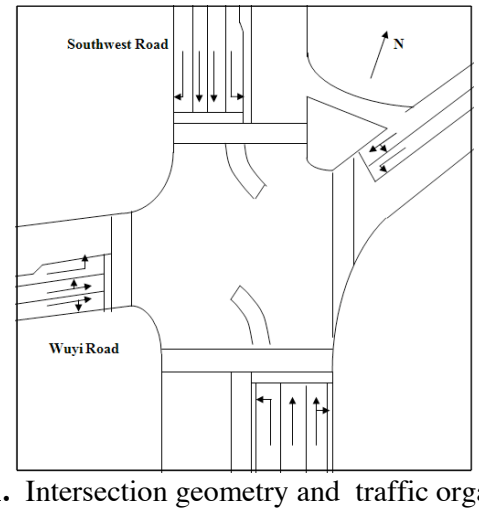


Figure 1. Intersection geometry and traffic organization.

	Phase 1	Phase 2	Phase 3	Phase 4
Signal Scheme				
Green Time	48s	10s	53s	53s

Figure 2. Current signal control scheme of the intersection.

Traffic Volume and Saturation Flow of the Intersection. **Investigation of the traffic volume.** Arrange the peak-hour traffic volume and directional volume numerated by investigators at all lanes of the intersection, in which traffic volume has been converted into equivalent units with the coefficient of signalized intersection.

Calculate saturation flow for each lane: Eliminating the influence of the opposed traffic flow, the saturation flow of a lane affected by various factors could be calculated by Eq. 1 (Wu and Li 2007):

$$S = \frac{2080 - 140 s_n - 428(\varphi \times G + 100(W_1 - 3.23))}{1 + \frac{G_{if}}{P}} \quad (1)$$

So we can obtain the directional volume of the approaches, traffic volume and calculated saturation flow of all lanes as listed in Figure 3 and Table 1.

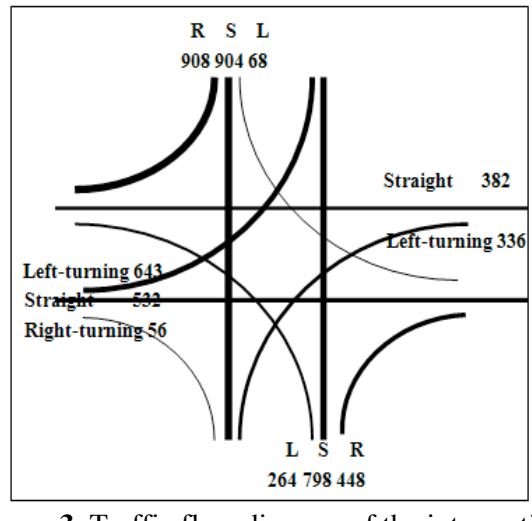


Figure 3. Traffic flow diagram of the intersection.

Table 1. Lane distribution and saturation flow of the intersection.

		Type	Volume	Turning Proportion	Saturation Flow
South Approach	Lane 1	Straight-right	746	60%	1971
	Lane 2	Straight	500	0	2130
	Lane 3	Left-turn	264*	1	2029
East Approach	Lane 1	Straight-left	412	7.3%	1961
	Lane 2	Left-turn	306	1	2073
North Approach	Lane 1	Right-turn	908	1	1885
	Lane 2	Straight	452	0	2103
	Lane 3	Straight	452	0	2103
	Lane 4	Left-turn	68	1	1967
West Approach	Lane 1	Straight-right	280	19.5%	1888
	Lane 2	Straight-left	440	30.6%	2083
	Lane 3	Left-turn	511	1	2037

Analyze the Traffic Performance of the Intersection Under Current Scheme. **Traffic capacity.** For a cross-shaped signalized intersection, the design traffic capacity is equal to the design capacity of all the approaches, while the design capacity of an approach depends on the sum of all the lanes. Using the method recommended by "Urban Road Designing Standards" to calculate the capacity of a lane (Zhang and Zhang 2002):

(i) The design capacity of a straight lane:

$$C_s = \frac{3600}{T_C}\left(\frac{t_g - t_l}{t_{si}} + 1\right)\delta_s \tag{2}$$

(ii) A straight-right lane's capacity:

$$C_{sr} = C_s \tag{3}$$

(iii) A straight-left lane's capacity:

$$C_{sl} = C_s\left(1 - \beta_1'/2\right) \tag{4}$$

(iv) And the capacity of a left-turning lane could be calculated by the "stop line transect" method:

$$C_l = \frac{3600}{t_c} \cdot \frac{t_g - v_l/2_s}{t_c} \tag{5}$$

Delay. The level-of-service(LOS) of signalized intersection is determined by the average delay caused by signal control. Average control delay per vehicle during the 15min analysis period is used to evaluate the LOS of the intersections in this paper.

Based on "interim function curve formula" to estimate the delay of the intersection (Wu 2007 and HCM2000).

(1) Delay of each lane:

$$d = d_1 + d_2 \tag{6}$$

$$d_1 = 0.5C\frac{(1-\lambda)^2}{1-\min[1,\ x]\lambda} \tag{7}$$

$$d_2 = 900T\left[(x-1) + \sqrt{(x-1)^2 + \frac{8ex}{CAPT}}\right] \tag{8}$$

(2) The average control delay of an approach is estimated by the weighted average delay of each lane included:

$$d_A = \frac{\sum_{n=1}^{i} d_i q_i}{\sum_{n=1}^{i} q_i} \tag{9}$$

(3) Total control delay of the intersection is estimated by the weighted average delay of each approach:

$$d_I = \frac{\sum_{A=1}^{A} d_A q_A}{\sum_{A=1}^{A} q_A} \tag{10}$$

Present traffic criterions of the intersection under current scheme. Based on the formulas listed above, we can calculate the related criterion of the intersection based on the current traffic volume and signal timing scheme; the results are shown in Table 2. Assuming $(PHF)_{mn} = 0.8$. (M for the approach and n for the directional traffic flow)

Table 2. Present traffic criterions of the intersection.

Criterion	South approach		West approach		North approach		East approach	
	Straight	Left-turn	Straight	Left-turn	Straight	Left-turn	Straight	Left-turn
Volume	798	264	532	643	904	68	382	336
Design volume	998	330	665	803	1130	85	478	420
Capacity	477	63	525	499	682	63	337	427
Green ratio	0.262	0.055	0.290	0.290	0.262	0.055	0.290	0.290
Saturation flow	2918	2029	2965	2674	4206	1967	1818	2216
Flow ratio	0.342	0.163	0.224	0.300	0.268	0.043	0.263	0.190
Saturation deg.	1.31	2.96*	0.77	1.03	1.02	0.78	0.91	0.66
Delays (in seconds)	221.5	1009	69	112	107.1	147.6	91.9	67
	417.3		92.5		110		79.2	
Total	Volume:3927		Capacity:3072		Saturation deg.:1.31		Delay:182.4	

It has been illustrated by the calculations that the LOS of the intersection is only in the F-class under current signal control scheme (Zhang 2002), and the average delay per vehicle is up to 3 min, which means each vehicle has to wait at least one signal cycle to pass the intersection after their arrival. What's more, according to the observation, actually those in the rear portion of the queue often need to wait two or three cycles to pass through the intersection. The delay of the major road (Southwest Road) is serious, with as much as 7 min of south approach and nearly 2 min of north approach, which leads to the enormous average delay of the intersection. Branch road (Wuyi Road) is relatively better than the major road, but the left-turning phase of the west approach suffers serious delays as well, therefore, rational and effective traffic organization and green time allocation should be taken for the purpose of better efficiency of the intersection.

Signal Timing Optimization Under the Current Scheme

One of the most popular theories for signal timing optimization is Webster's method; the formulas and calculating process are shown as follows (Wu and Li 2007):

$$C_0 = \frac{1.5L+5}{1-Y} \tag{11}$$

where,

$$L = \sum_{i=1}^{n}(l_i + I_i - A) \tag{12}$$

Generally speaking, it is unavailable using Webster's formula to calculate an appropriate cycle length when the value of Y is greater than 0.9. In this case, channelization of the approaches should be taken, then re-design the signal timing

scheme to calculate an optimum cycle length, and obtain the total effective green time G_e, effective green time of each phase g_{ej}, and the display green time g_j.

$$G_e = C_0 - L \tag{13}$$

$$g_{e1} = G_e \frac{\max[y_1, y_{1'}, \ldots]}{Y} \tag{14}$$

$$g_j = g_{ej} - A + I \tag{15}$$

Optimize the current signal timing scheme based on the field survey data:

The sum of critical flow ratios of each phase:

$$Y = \sum_{j=1}^{j} \max[y_j, y_{j'} \ldots] \tag{16}$$

Critical flow ratio of each phase: $y_1 = \max(0.342, 0.268) = 0.342$,
$y_2 = \max(0.163, 0.043) = 0.163$, $y_3 = \max(0.263, 0.190) = 0.263$,
$y_4 = \max(0.224, 0.300) = 0.300$. Reckoning: Y=1.068 >1.

The flow ratio Y is greater than 1 **, therefore it is unable to optimize the signal timing scheme under the current traffic conditions of the intersection.

As shown in Figure 1 and the investigation data, the entire width of the south approach is relatively greater, with a 4 m wide medium strip; besides, the left-turning angle is very small as a consequence of the geometric restrictions of the intersection; furthermore, the east section of Wuyi Road is narrow and the turning flow is obviously less than other directions, thus two scenarios with relatively lower cost for implementation are suggested here:

a) Channelize the south approach to the central of the road to add a lane, then install a dedicated right-turning lane as the volume of right-turning is 560 pcu/h.

b) Prohibit left- turns on the north approach, the former left-turning vehicles could go straight and make a U-turn at the end of the separator 300 m ahead then back to the intersection and turn right (Hong et al. 2001), or take other paths to arrive at their destinations.

Optimize the current signal timing scheme on conditions of the two suggested scenarios which are illustrated in Figure 4 (a) and (b), respectively, then calculate the flow ratio of each phase. The results are listed in Table 3:

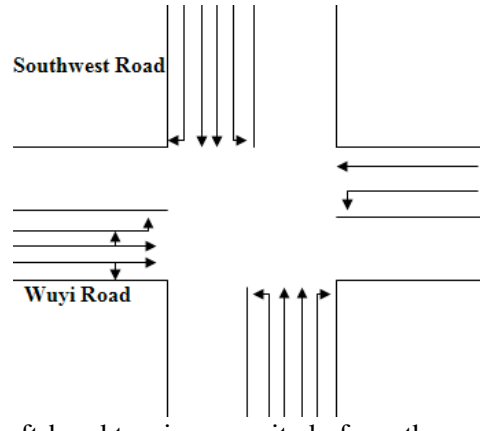
Notice: SB for southbound through and SBLT for southbound left turn, the same as others.

For Figure 4 (a):

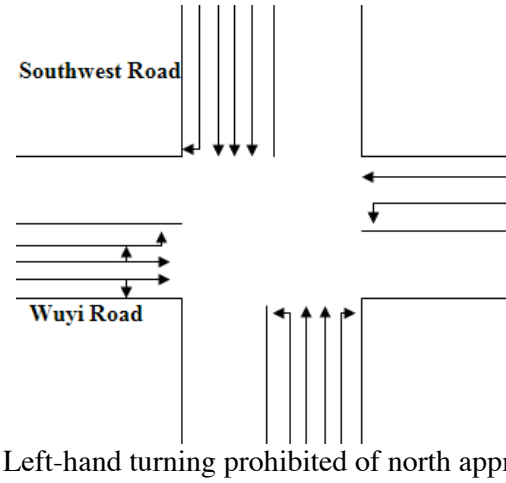
$Y = 0.269 + 0.163 + 0.300 + 0.244 = 0.976$ too large for optimization

For Figure 4 (b):

$Y = 0.240 + 0.163 + 0.300 + 0.244 = 0.947$ too large for optimization



(a) Left-hand turning permited of north approach



(b) Left-hand turning prohibited of north approach

Figure 4. Traffic organization afterwards the channelization of south approach.

Table 3. Flow ratio of two traffic organizations after the channelization of south approach.

	Design Volume	Saturation Flow 1	Flow ratio y1	Saturation Flow 2	Flow ratio y2
SB	998	4166	0.240	4166	0.240
SBLT	330	2029	0.163	2029	0.163
WB	665	2965	0.224	2965	0.224
WBLT	803	2674	0.300	2674	0.300
NB	1130	4206	0.269	6309	0.179
NBLT	85	1967	0.043	— —	— —
EB	478	1961	0.244	1961	0.244
EBLT	420	2073	0.203	2073	0.203

The results demonstrate that although the total flow ratio of the intersection can be slightly reduced by increasing the saturation flow after channelization, it still couldn't meet the requirement of signal timing optimization.

Design of Overlap Phase

Principle. As can be seen from the investigation that the directional flow may be unequal in identical phase, diversity is significant. However, many designers didn't pay much attention to the diversity, resulting in a waste of time and space resources (Peng and Zhang 2001). In some cases, it may be desirable to pair a left turn first with the opposing left turn and then with the through movement. This is sometimes referred to as "overlapping" the left turns and through phases, and the middle phase in the sequence is referred to as an "overlap phase" (Banks 2002 and Al-Kaisy and Stewart 2001). Use of overlap phases can shorten the cycle and make full use of effective green time if the dominant left turn (that is the left turn requiring the greater amount of time) and the dominant straight flow are from the same approach, so that the capacity of the intersection will be improved (Quan 1989).

Design. Section 2 demonstrates that current signal control scheme could not be optimized efficiently even after the channelization of the south approach and prohibited left-turn of north approach. After analyzing the directional traffic distribution of the intersection and influence to the saturation flow of each phase caused by traffic organization, we tested several schemes and design feasible traffic organization and signal control scheme taking the left-turning flow of the west approach as critical (see Figures 5 and 6). (For safety consideration, we suggest broadening the west entrance by channelization so as to install a merging lane for northbound right-turn traffic.)

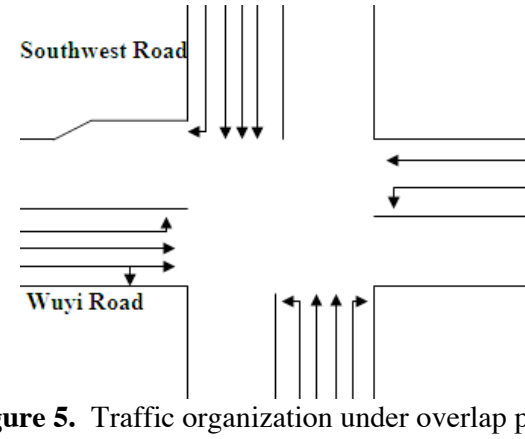


Figure 5. Traffic organization under overlap phase.

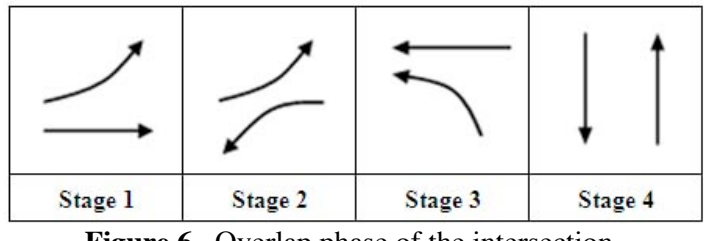


Figure 6. Overlap phase of the intersection.

Table 4. Flow ratio of overlap phase signal timing plan.

	Volume	Design Volume	Saturation Flow	Flow ratio
SB	798	998	4166	0.240
SBLT	264	330	2029	0.163
WB	532	665	3603	0.185
WBLT	643	803	2037	0.394
NB	904	1130	6309	0.179
NBLT	-	-	-	-
EB	382	478	1961	0.244
EBLT	336	420	2073	0.203

Under such signal control scheme and traffic organization, we can calculate the saturation flow and flow ratio of each phase as listed in Table 4:

Determine critical flow ratios of each phase:

$$y_1 = \max(y_{WBLT}, y_{WB} + y_{EBLT}) = 0.394 ,$$

$$y_2 = \max(y_{EB}, y_{SBLT}) = 0.244 , \quad y_3 = \max(y_{SB}, y_{NB}) = 0.240$$

The sum of critical flow ratio:

$$Y = y_1 + y_2 + y_3 = 0.878$$

Determine cycle length:

$$C = \frac{1.5L + 5}{1 - Y} = \frac{1.5(4 \times 3) + 5}{1 - 0.878} = 188s$$

The total effective green time:

$$G_e = C_0 - L = 188 - 4 \times 3 = 176s$$

It is obvious that during phase 1 and phase 2, y_{WBLT} is greater than the sum of y_{WB} and y_{EBLT}, thus the green time for left-turning on west approach should be divided into two parts proportionally with y_{WB} and y_{EBLT}:

$$y_{WB} = \frac{y_{WBLT} \times y_{WB}}{y_{WB} + y_{EBL}} , \quad y_{EBLT} = \frac{y_{WBLT} \times y_{EBLT}}{y_{WB} + y_{EBLT}} \tag{17}$$

Reckoning: $y_{WB} = 0.188$, $y_{EBLT} = 0.206$

Distribute effective green time for each phase:

$$g_{e1} = 38s, \quad g_{e2} = 41s, \quad g_{e2} = 49s, \quad g_{e2} = 48s$$

Green time of each phase displayed:

$$g_1 = 38s , \quad g_2 = 41s , \quad g_3 = 49s , \quad g_4 = 48s$$

Whereas the green time of left-turning phase for the west approach of Wuyi Road is 82 s, which is calculated by adding g_{e1} and g_{e2} and the interval of yellow time. So we can obtain the signal timing graphic as Figure 7.

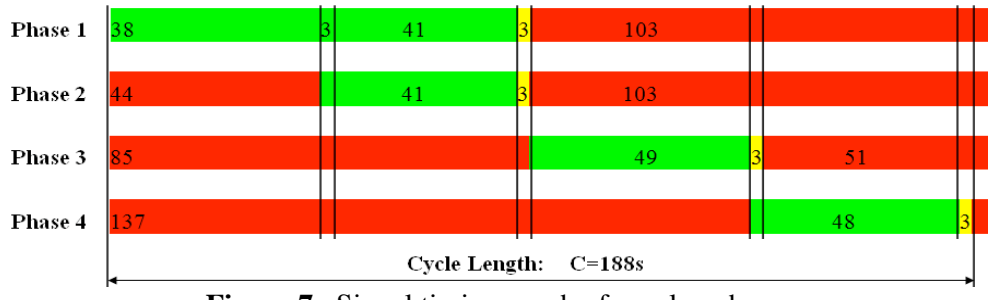


Figure 7. Signal timing graph of overlap phase.

Notes:

(1) Due to the width of south approach, drivers often squeeze the interval of two lanes, resulting in double left-turning flow; it means potential error exists in calculating the flow ratio, so reject it.
(2) Theoretically, the sum of flow ratio is less than 1; however, violation of traffic laws (for instance, early start and squeeze the interval) will influence the survey data and lead to the phenomenon.

Theoretical analysis. Calculate criterions of the intersection under the improved signal control scheme using the procedure listed in section 1.3.The results are shown in Table 5.

Table 5. Traffic criterions of the intersection within overlap phase signal timing scheme.

Criterion	South approach		West approach		North approach		East approach	
	Straight	Left-turn	Straight	Left-turn	Straight	Left-turn	Straight	Left-turn
Volume	798	264	532	643	904		382	336
Design Volume	998	330	665	803	1130		478	420
Capacity	664	360	475	613	996		339	299
Green ratio	0.255	0.261	0.202	0.436	0.255		0.261	0.218
Flow ratio	0.240	0.163	0.185	0.394	0.179		0.244	0.203
Saturation deg.	0.941	0.625	0.916	0.904	0.702		0.935	0.931
Delays (in seconds)	91.7	69.3	98.4	68.6	67.6		102.8	109.1
	86.1		82.2		67.6		106	
Total	Volume: 3859		Capacity: 3746		Saturation deg.:0.941		Delay: 84.3	

Comparison and Conclusions

In comparison with the criterions of current signal control schemes, we come to the conclusion that the advantages of improved signal control schemes can be demonstrated as follows:

(1) The degree of saturation of the intersection will decrease to 0.941 from 1.31, which means adjusted from over-saturation to the "quasi-saturation" state;

(2) Using overlap signal control scheme and related traffic organization, the capacity of the intersection will increase from 3072 pcu/h to 3746pcu/h, and the scope of increase is 21.9%;

(3) Average delay per vehicle will drop from 182.4s to 84.3s, a significant scope of 53.1%.

Obviously, considerable improvement of traffic capacity and average delay will take place with the use of channelization of approaches and overlap phasing. However, LOS of the intersection has not been improved, which means signal timing optimization can only alleviate the congestion to a certain extent, while the practical traffic volume and capacity of the intersection determines the congestion, which is beyond the ability of signal timing optimization.

Conclusions

The purpose of signal timing optimization is to make full use of effective green time; as for intersections with disproportionate traffic flow, using early cut-off (or late release) the opposed flow will reduce the loss of green time, consequently achieving the goal of improving the intersections' capacity. However, taking the diversity of different intersections into consideration, specific analysis and design is needed for the given case. However, we come to the conclusion from the calculation simultaneously that signal timing optimization can only alleviate the traffic congestion to a certain extent. In order to solve the problem radically, authorities related to the issue should consider the structure of road network synthetically to implement a traffic flow guidance project or widen the intersection outward to increase the capacity of the intersection when it is necessary.

References

Al-Kaisy. A. F., and Stewart. J. A. (2001). "New approach for developing warrants of protected left-turn phase signalized intersections." *Transportation Research part A*.

Banks. J. H. (2002). "Introduction to transportation engineering." *2nd ed. Library of Congress Cataloging-in-Publication Data*. 299-307.

Hong, S., Shin, E., Kim, D. N., and Kim, Y. (2003). "An optimization model for signal timings and alternate lane use at a signalized intersection." *Journal of the Eastern Asia Society for Transportation Studies*. 1109-1123.

Peng, G.-X, Zhang, Y. (2001). "Study of the way of setting ripple changes [J]." *China Journal of Highway and Transport*. 59-62.

Quan, Y. (1989). "Urban traffic control [M]." Beijing: *China Communication Press*. 95-128.

Transportation Research Board. Highway Capacity Manual 2000 [R]. *National Research Council,* Washington, D.C.

Webster, F. V., and Cobbe, B. M. (1966). "Traffic signals." *Road Research Technical Paper*,Vol.56, Her Majesty's Stationery Office, London.

Wu, B., and Li, Y. (2007). "Traffic management and control [M]." *3rd ed. Beijing: China Communication Press*. 147-162.

Zhang, Q., and Zhang, Y. (2002). "Road traffic capacity analysis [M]." *Beijing: China Communication Press*. 149-155.

A Study on the Improvement of T-intersection Signal Timing

Rui Li[1] and Wenquan Li [2]

[1]PH.D Student, [2]Prof., School of Transportation, Southeast University, No.2 Sipailou, Nanjing 210096, P. R. China, PH:15895985925; Email: lirui84517@163.com, PH:+86-25-83795647, Email: wenqli@seu.edu.cn

Abstract

In urban traffic systems, intersections are often to be capacity bottlenecks and high accident locations. T- intersections are generally less efficient than standard four-leg intersections. This paper addresses the traffic conflicts of T- intersections and introduces the traffic signal phases that could improve the efficiency of T-intersections. According to the general signal timing methodology for intersections, an improved method for T- intersections was introduced. Through a case study, the improved method proves to be more suitable for T- intersection signal settings and has a high practical value.

Key Words: T-intersection; signal timing; traffic conflicts; left-turn phase

Introduction

Intersection is a key element in urban traffic systems. Intersections are not only capacity bottlenecks, but also areas of high traffic accidents. T-intersection is an important form of urban intersections in China. With the increase of road traffic, T-junctions have gradually become urban road network in the traffic capacity of the "defile" and the traffic accidents "multi-origin,"(Chen Jifa, 2008) Now, the increasing number of vehicles in the roadway network, many T-intersections of the branch way switched from their original stop or yeild control to signal control. Therefore, this paper has a good practical value on the study of T-intersection signal timing.

From the sixties of last century, the foreign scholars began to address intersection traffic flow characteristics and the signal was studied.(Webster F V et al 1966, Kell J H et al 1982, Thomas F. Golob et al 1988, Avishai Ceder et al 2002) Since then, Chinese scholars have gradually begun to pay attention to intersection signal settings.(Zhang Li et al 2005, Li Yang et al 2007, Zhong Xiaoming et al 2009, Sun Chao et al 2009) Taking full account of the author on the basis of previous research results, combined with T-intersection traffic flow characteristics, a improved method about T-intersection signal timing is proposed. The effects of improved method are proposed and analyzed based on a case study.

T-intersection Signal Phase Setting Analysis

For the T-junction(shown in Figure 1): the major-minor intersection direction and crossing point、 diverging point、 merging point are defined based on Figure 1.

For the T-intersection, taking into account the traffic conflict between the left-turn vehicles of direction 2 and vehicles going straight of direction 3, therefore, left-turn phases should be set to diverge the traffic conflict. For non-motor vehicles and pedestrians crossing the street, this paper proposed to utilize successive crossings in the T-intersection (shown Figure 2), straight-turn and left-turn vehicles go through the street across the Crosswalk line, accompanied by the appropriate time for a green light signal.

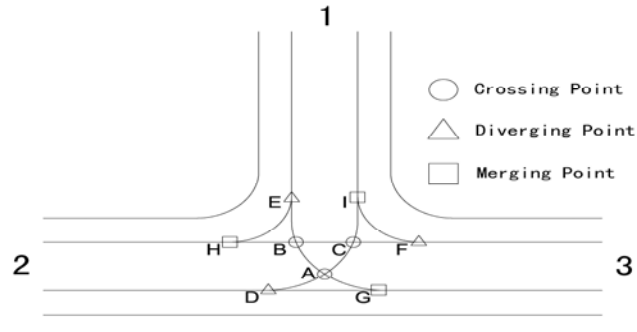


Figure 1. Traffic conflict of T-intersection.

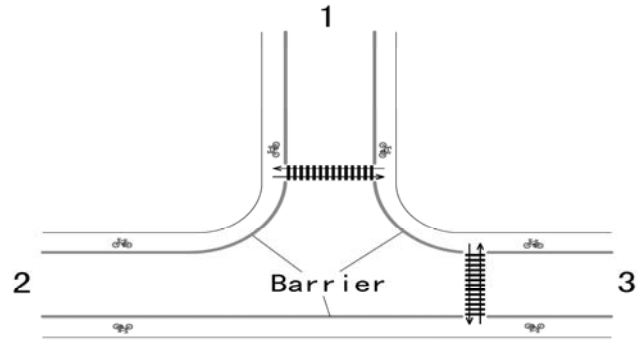


Figure 2. Crossing method of T-intersection.

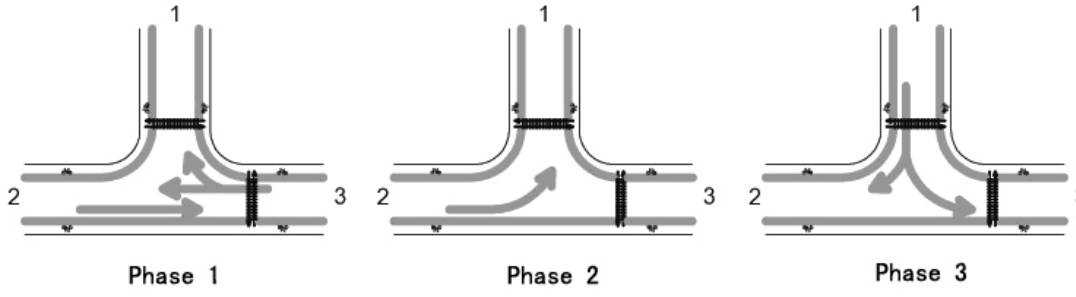


Figure 3. T-Intersection phasing design and sequence.

The signal phase of T-Intersection is set as follows(shown Figure 3): Phase 1, allowing the straight-turn vehicles of direction 2、 straight-turn and right-turn vehicles of direction 3、 non-motor vehicles and pedestrians of lane 1 to go through the intersection. Phase 2, allowing the left-turn vehicles of direction 2、 non-motor vehicles and pedestrians of direction 3 to go through the intersection. Phase 3, allowing the left-turn and right-turn vehicles of direction 1 to go through the intersection. The signal phase is set to fully guarantee the rights of left turn traffic movement, and motor vehicle access, while ensuring the maximum degree of satisfaction of non-motor vehicles and pedestrians acrossing the street needs. For each direction of motor vehicles, non-motor vehicle traffic more evenly spread is more applicable to the T-junction.

Model

Traditional intersection traffic signal timing methods were applied to a cross-intersection of the two-phase signal characteristics of signal timing design, such as the method offered by Wang Wei and others (2000)in “traffic engineering”. First of all, calculating the intersection of imports road traffic volume is equivalent to calculate the intersection signal cycle, and then the green light time of each phase of the distribution is calculated.

$$V_e = \frac{V + 0.5H + 0.6L}{n} \quad (1)$$

$$T = \frac{13330P}{1333 - V_e} \quad (2)$$

V_e = the equivalent traffic volume; V = the real traffic volume of intersection; H = the volume of bus and freight; L = the volume of left-turn vehicle; n = the number of import lane of some direction; T = the signal cycle; P = the number of traffic signal phase.

The improved method for T-intersection signal timing proposed in this paper is based on traditional intersection signal timing method. In addition, the paper also introduces the amendment of left-turn and right-turn vehicle at T-intersections. For the problem that signal cycle is too long, this paper also offers the solution.

As the three-phase T-intersection has separate left-turn phase, therefore, left-turn and opposite-straight-turn vehicles nearly have no traffic conflict, and the travel time of vehicles at the intersection is mainly related with travel distance and travel speed. For the speed of left-turn of direction 1 in T-junction, the left-turn and straight-turn vehicle have similar speeds, because that the intersections within the lower rate of vehicle start, and intersections within the speed is not high, therefore, left-turn vehicle's travel time in T-intersection mainly has relationship with travel distance. For the T-junction left-turn and right-turn vehicles of direction 1, the vehicles within the scope of their intersection can also be approximate speed of the straight-turn vehicle's speed, therefore, in this direction left-turn and right-turn vehicles at the intersection within the scope of travel time is mainly related to their distance from the traveling.

For the intersection within the left-turn and right-turn traffic to run the distance, you can run the track in accordance with their respective approximate obtained (shown in Figure 4:

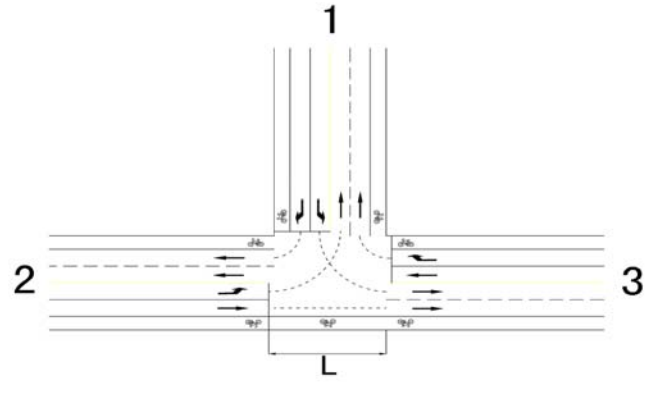


Figure 4. Track of vehicle operation in intersection.

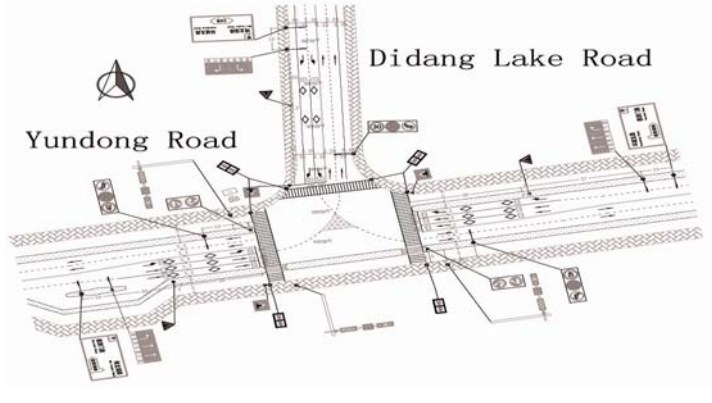


Figure 5. Design of intersection.

It is assumed that T-junction 3 entries are the same width (set to L); motor vehicle Rroads have four lanes on both sides; and each direction also have non-motor vehicle lane. In accordance with motor vehicle operation track, the distance of straight-turn vehicle travel at the intersection is approximately equal to the width of L. The distance of left-turn vehicles of direction 1 and direction 2 travel at the intersection is approximately equal to 1/4 arc of a radius of L/2. The distance of right-turn vehicle of direction 1 travel at the intersection is approximately equal to 1/4 arc of a radius of L/4. The use of mathematical knowledge to turn left and turn right onto the vehicle's distance derivation, can be drawn within the intersection left-turn vehicles running distance is $\pi L/4$; right-turn traffic at the intersection within the operating distance is $\pi L/8$. Considering the vehicle's trajectory and speed conditions, will turn left at the intersection of vehicles operating within the scope of the distance as L, right-turn traffic at the intersection within the set at a distance L/2.

Thus, for T-type three-phase intersection signal timing design: the parameter of equivalent traffic volume of left-turn vehicle is 1.0; the parameter of equivalent traffic volume of right-turn vehicle is 0.5; the equivalent traffic volume of T-type three-phase traffic intersection is $V_e = \dfrac{V - 0.5R}{n}$ (V is the equivalent traffic volume. Models are divided into standard car equivalent traffic volume). T-type three-phase traffic intersection is equivalent to the further refinement. The following equation can be developed.

$$V_e = \max(V_{1r}/2n_{1r},\ V_{1l}/n_{1l}) + \max[(V_{3r}/2 + V_{3s})/n_3,\ V_{2s}/n_{2s}] + V_{2l}/n_{2l} \qquad (3)$$

Where, V_e—equivalent traffic volume of the intersection

V_{3s}、V_{3r}—straight-turn and right-turn equivalent traffic volume of direction 3 at the T-intersection (other directions' parameters are the same)

n_{2s}、n_{2l}—the numbers of straight-turn and left-turn vehicle lanes of direction 2 at the T-intersection (other directions' parameters are the same)

n_3—the number of vehicle lanes of direction 3 at the T-intersection.

In accordance with formula (2) T-type three-phase intersection signal cycle is determined. As the T-type 3-phase intersection signal timing has been established, the signal cycle is likely to be too long (usually the signal cycle for amendments is more than 140 seconds), resulting in delay for the overall situation at the T-intersection. This can be through "compression" to reduce delays in the signal cycle. "Compression" of the signal period can calculate using the formulas: $G_0 = 2.1x + 3.7$ (Sun Chao et al 2009) (G_0 is a particular phase of the green time required for vehicle clearance (s); x is the cycle to car number). Cycle length can be obtained by adding up each phase time.

Table 1. Peak hours traffic volume of Yundong Road-Didang Lake Road intersection.

	Left Volume/pcu	Straight Volume/pcu	Right Volume/pcu	Total Volume/pcu
East Volume	−	673	178	851
West Volume	279	702	−	981
North Volume	275	−	309	584

Model Test

Taking the Yundong Road-Didang Lake Road intersection in a city of Zhejiang Province as an example, the phases of the T-intersection signal settings and optimization of practical application are described as below.

According to Yundong Road-Didang Lake Road intersection vehicle traffic flow (such as in Table 1 shown), the above T-intersection signal timing method is used to design the signal cycle of T-intersection, at the same time. "Traffic engineering" method(Wang Wei et al 2000) and Webster algorithm(Yang Peikun et al 1995) are also carried out to design the signal cycle of the same T-intersection, the three methods of calculation results in Table 2.

From the results above, the traffic signal cycle calculated by Webster algorithm is too short, which will inevitably lead to an increase in the number of vehicles stops. Thereby, increased delay of vehicles stops will seriously affect the pedestrian crossing signal timing. The traffic signal cycle calculated by traffic engineering method is too long, causing increased vehicle delays. Improved algorithm for cycle time is moderate, and the effective green time of each phase meets $G_0 = 2.1x + 3.7$. Therefore, this method for the T-type three-phase intersection signal timing is optimal.

Table 2. Comparison of algorithm optimization results.

	Signal Cycle/s	Signal Phase Ⅰ /s	Signal Phase Ⅱ /s	Signal Phase Ⅲ /s
Improved Algorithm	93	33	26	25
Webster Algorithm	45	14	10	12
"Traffic Engineering" Method	418	116	148	145

Conclusions

Based on analyzing the past traffic signal timing methods, this paper introduced an improved algorithm on traffic signal timing for T-intersections. Compared with the traditional methods (such as traffic engineering method and Webster algorithm), the improved method has a greater advantage in reducing the stop delays for T-intersection traffic signal timing , therefore, it has good applicability and practicality.

Acknowledgements

The work is supported by the National Key Technology Research and Development Program (No.2006BAJ18B03).

References

Avishai Ceder, Kobi Eldar. (2002). "Optimal distance between two branches of uncontrolled split intersection." *Transportation Research Part A: Policy and Practice*, 36(8), 699-724.

Chen Jifa. (2008). "Research on capacity of the urban unsignalized t-type intersection." Nanjing: Southeast University.

James H. Kell, Iris J. Fullerton. (1982). *Manual of traffic signal design*, Washington D C: Prentice-Hall, Inc.

Li Yang, Han Bingyuan, Wang Na. (2007). "Research on the timing of the phase position and traffic lights at the "T" type road intersection." *Forest Engineering*, 23(6),37-39.

Sun Chao, Xun Jianmin. (2009). "Traffic signal control and simulation for signal-point intersection." *Journal of Transport Information and Safety*, 27(2),35-38

Thomas F. Golob, Bolie Ruhl, Henk Meurs etc. (1988). "An ordinal multivariate analysis of accident counts as functions of traffic approach volumes at intersections." *Accident Analysis & Prevention*, 20(5), 335-355.

Wang Wei, Guo Xiucheng etc. (2000). *Traffic Engineering*, Nanjing: Southeast University Publishing House.

Webster F V, Cobbe B M. (1966). *Traffic signals*, London: Road Research Laboratory.
Yang Peikun, Zhang Shusheng. (1995). *Traffic Management and Control* , Beijing: People's Communications Publishing House.

Zhang Li, Ma Yan. (2005). "The description of traffic flow mathematical model of T intersection." *Journal of Northeast Forestry University*, 33(1):106-107.

Zhong Xiaoming, Wei Zhonghua, Zhang Li etc. (2009). "Traffic flow characteristics of T type unsignalized intersections." *Journal of Beijing University of Technology*, 35(2), 224-229.

Effects of the Spatial Distance between Two Adjacent Bus Stops on Traffic Flow

Tieqiao Tang[1], Haijun Huang[2], Yan Li[1] and Xin Bai[1]

[1,2]Associate Professor, School of Transportation Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing, China 100191, Email: tieqiaotang@buaa.edu.cn

Abstract

Bus-stop is one important compositions of urban traffic system and often produces some traffic interruptions since the buses should stop at the bus stop. In urban traffic system, there are often multi bus stops on a road, so the distance between bus stops will have great effects on traffic flow and produce some complex traffic phenomena. In this paper, we use the traffic flow model to further study the properties of the traffic flow on a ring road with two bus stops. The numerical tests show that our model can describe the effects of the distance between two adjacent bus stops on the stability of traffic flow and reproduce some complex traffic phenomena under the action of difference distances, which just show that our model is reasonable.
Key Words: traffic flow; bus stop; the spatial distance between two adjacent bus stops; stability.

Introduction

So far, traffic problems have become very serious and attracted scholars to develop many traffic flow models to describe various complex traffic phenomena (Chowdhury et al 2000; Jiang et al 2002; Lighthill and Whitham 1955; Pichards 1956; Zhang 1998). However, these models cannot directly be used to study bus stop because they did not consider the factor. In order to study bus stop, scholars later presented some cellular automaton models to investigate bus stop (Jia et al 2009; Qian et al 2008; Zhao et al 2008), but the CA models cannot describe the effects of bus stop on the dynamics property of traffic flow. In addition, Nagatani (2002) developed a coupled-map lattice model to study the dynamic behavior of a recurrent bus on a ring road with multi bus stops, but he did not consider the effects of bus stop on the dynamic equation. In the

real traffic system, bus will produce traffic interruption when it enters the bus stop, but the above models cannot be used to study the traffic interruption resulted by bus stop. Recently, Tang et al. (2008) presented a macro model with the consideration of the traffic interruption probability and pointed out that this model can be adopted to describe various traffic interruptions if we can properly define the traffic interruption probability. Later, Tang et al. (2009) used this model to describe the effects of the number of bus stops on traffic flow and found that this model can reproduce some complex phenomena resulted by bus stop, but they did not further study the effects of the spatial distance between two adjacent bus stops on traffic flow. In this paper, we adopt the model (Tang et al 2008) to study the effects of the spatial distance between two adjacent bus stops on traffic flow.

Model

Among the existing macro models, the first model is the Lighthill-Witham-Richards model (Lighthill and Witham 1955; Richards 1956), which is called as the LWR model and can be reduced as follows:

$$\frac{\partial \rho}{\partial t}+\frac{\partial q}{\partial x}=0 \tag{1}$$

where ρ is the density, and q is the flow and satisfies the following condition:

$$q=\rho v_{\mathrm{e}}(\rho) \tag{2}$$

where $v_{e(}\rho)$ is the equilibrium speed. The LWR model can describe the formation and propagation of shock (Wang 1995), but it cannot reproduce the non-equilibrium property of traffic flow because the speed cannot deviate from the equilibrium speed. In order to conquer the drawback, scholars developed many dynamic models. In fact, the existing dynamics models can be divided into the density-gradient (DG) models (Zhang 1998) and the speed-gradient (SG) model (Jiang et al 2002), where the DG model can be written as follows:

$$\frac{\partial v}{\partial t}+v\frac{\partial v}{\partial x}=\frac{v_{\mathrm{e}}-v}{\tau}-\frac{c^2(\rho)}{\rho}\frac{\partial \rho}{\partial x}+\mu\frac{\partial^2 v}{\partial x^2} \tag{3}$$

where τ is the reactive time,c(ρ)>0 the sonic speed, and μ the viscous coefficient. Daganzo (1995) found that backward motion will occur under some given conditions when the DG model is used to study the evolution of the density. In order to overcome this drawback, the speed-gradient (SG) model has been developed, i.e.

$$\frac{\partial v}{\partial t}+v\frac{\partial v}{\partial x}=\frac{v_e-v}{T}+c_0(\rho)\frac{\partial v}{\partial x} \tag{4}$$

where $c_0(\rho)>0$ is the propagating speed of small perturbation. It has been proved that the SG model can overcome the drawback of backward motion and can perfectly describe some complex traffic phenomena.

Although the above models can describe some complex traffic phenomena, they cannot be used to directly study traffic interruptions since the interruption factor is not explicitly considered. Based on the effects of the traffic interruption probability on the car-following behavior, Tang et al (2008) developed one macro model, i.e.

$$\begin{cases} \dfrac{\partial \rho}{\partial t}+\dfrac{\partial(\rho v)}{\partial x}=0 \\ \dfrac{\partial v}{\partial t}+v\dfrac{\partial v}{\partial x}=\dfrac{v_{\mathrm{e}}(\rho)-v}{T}+\dfrac{1}{\tau_1}p(-v)+c_0(1-p)\dfrac{\partial v}{\partial x} \end{cases} \tag{5}$$

where p is the traffic probability and T, τ_1 are two reactive times.

Zhang et al (2005) pointed out that the conservative scheme is a good choice for carrying out numerical tests if using dynamics model to describe traffic flow, so we should rewrite Eq.(5) as follows:

$$\frac{\partial u}{\partial t}+\frac{\partial f(u)}{\partial x}=s(u) \tag{6}$$

where $u=(\rho,v)^{\mathrm{T}}$, $f(u)=\left(\rho v, 0.5v^2-c_0(1-p)v\right)^{\mathrm{T}}$, $s(u)=\left(0,\dfrac{v_e-v}{T}-\dfrac{pv}{\tau_1}\right)^{\mathrm{T}}$. We use the following conservative scheme to discretize Eq. (6)

$$u_i^{n+1}=u_i^n-\frac{\Delta t}{\Delta x}(\hat{f}^n_{i+1/2}-\hat{f}^n_{i-1/2})+s(u_i^n)\Delta t \tag{7}$$

where $\hat{f}^n_{i+1/2}=\hat{f}(u_i^n,u_{i+1}^n)$ is the local Lax-Friedrichs flow, *i.e.*,

$$\hat{f}(u_i^n,u_{i+1}^n)=\frac{1}{2}\left(f(u_i^n)+f(u_{i+1}^n)-\alpha(u_{i+1}^n-u_i^n)\right) \tag{8}$$

where $i,n,\Delta x,\Delta t$ are the space and time indexes, the spatial and time steps, respectively; α the maximum characteristic speed of Eq.(6). For simplicity, we here set $\alpha=v_{\mathrm{f}}$.

Simulation

In this section, we discuss the effects of the spatial distance between two adjacent bus stops on traffic flow. Urban traffic flow is heterogeneous and there are multi lanes in the traffic system, so bus stop has great effects on traffic flow, which are related to the number of lanes, the components of the heterogeneous traffic flow and other factors. In this paper, we only consider a special lane for bus without overtaking and assume that the buses can move according to a preset schedule. Since overtaking is prohibited during the boarding/alighting period, traffic will come to a complete halt behind the stopped bus. Based on the above assumptions, we can define the traffic interruption probability of Eq.(5) as follows:

$$p(x,t)=\begin{cases} 1,\ x=x_0 \text{ and } t\in\left[t_0+(n-1)T_{\mathrm{h}},t_0+(n-1)T_{\mathrm{h}}+T_{\mathrm{d}}\right],n=1,2,\cdots,n_1 \\ 0,\ \text{otherwise} \end{cases} \tag{9}$$

where x_0 is the position of bus stop, t_0,n_1 respectively the initial time and the times that bus stays at bus stop, T_{h} The headway of buses and T_{d} the dwell time.

In this paper, we use the following equilibrium velocity function (Kerner 1993):

$$v_e = v_f\left(1/\left(1+\exp\left\{\frac{\rho/\rho_j - 0.25}{0.06}\right\}\right) - 3.72\times 10^{-6}\right) \tag{10}$$

where v_f is the free-flow velocity and ρ_j the jam density. For simplicity, we define the other parameters as follows:

$$T = 10\text{s}, v_f = 30\text{m/s}, \rho_j = 0.2\text{veh/m},$$
$$\tau_1 = 8\text{s}, \Delta x = 100\text{m}, \Delta t = 1\text{s}, t_0 = 0, T_h = 120\text{s}, T_d = 10\text{s}.$$

Bus stop will often destroy the continuity of the traffic interruption probability function in the model, so we use simulation to study the traffic phenomena resulted by bus stop. For simplicity, We suppose that there are only two bus stops on a ring road, where its length is equal to $L = 30\text{km}$. The numerical results are shown in Figs.1-3, where the initial densities in Figs.1-3 are respectively very low, mid and very high. From the three figures, we have:

(a) When the initial density is very low ($\rho_0 \le 0.03$), the spatial distance between two adjacent bus stops has little effect on the stability of traffic flow. At this time, two slight local clusters will respectively appear at the upstream of the two bus stops, but the two clusters will quickly be incorporated when the spatial distance between the two bus stops is enough small and become more independent with the increase of the spatial distance (Figure 1).

(b)When the initial density is mid($0.03 < \rho_0 < 0.08$), the spatial distance between two adjacent bus stops will have great effects on traffic flow and destroy the stability of traffic flow, i.e. stop-and-go will appear, but it will turn more serious with the increase of the spatial distance (Figure 2). The reasons are as follows: when the spatial distance between the two bus stops is very small, the two perturbations resulted by the two bus stops will be incorporated into one perturbation; the interactions between the two bus stops will become more independent with the increase of the spatial distance, so the two perturbations resulted by the two bus stops will become two dependant perturbations.

(c)When the initial density is very high ($\rho_0 \ge 0.08$), the spatial distance between two adjacent bus stops has little effect on traffic flow. At this time, slight local cluster will appear only at the upstream of the two bus stops (see Figure 3).

In the real traffic system, there are often more than two bus stops on a road and the number of bus stops will have great effects on traffic flow (Tang et al, 2009), the headway of buses T_h and the dwell time T_d are often related to many factors (the number of passengers, the bus density, etc.), so the above results are not completely accordant with the reality. However, the above factors have little qualitative effects on the results obtained in this paper. As for the quantitative effects of bus stop on traffic flow, we will in the future use many experimental data to further describe the complex phenomena resulted by bus stop.

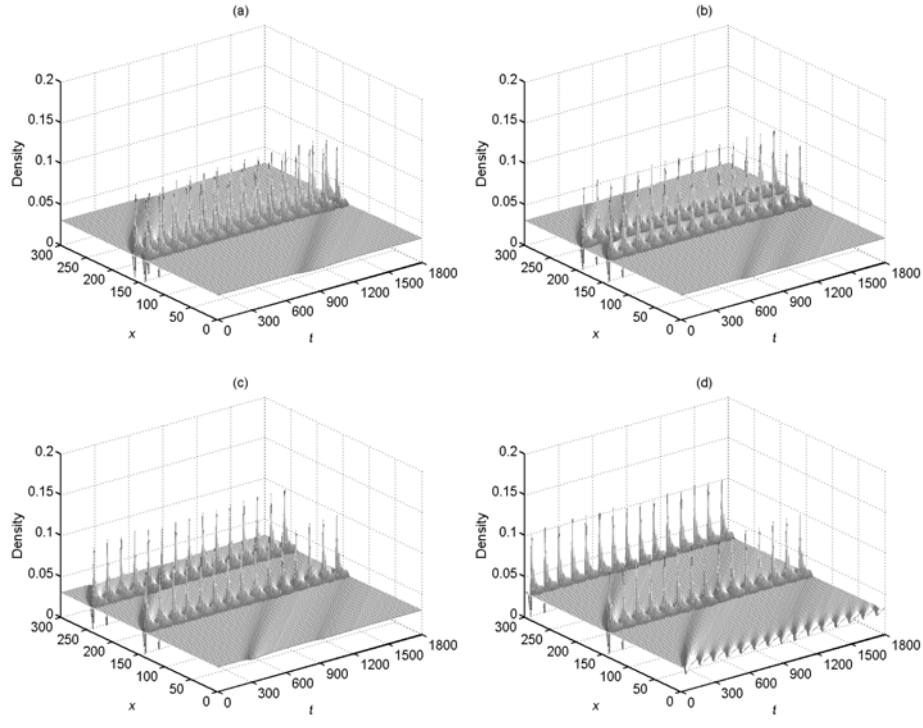


Figure 1. The evolution of the density, where the initial density is 0.03 and the distances in (a)-(d) are respectively 2 km, 5 km, 10 km and 15 km.

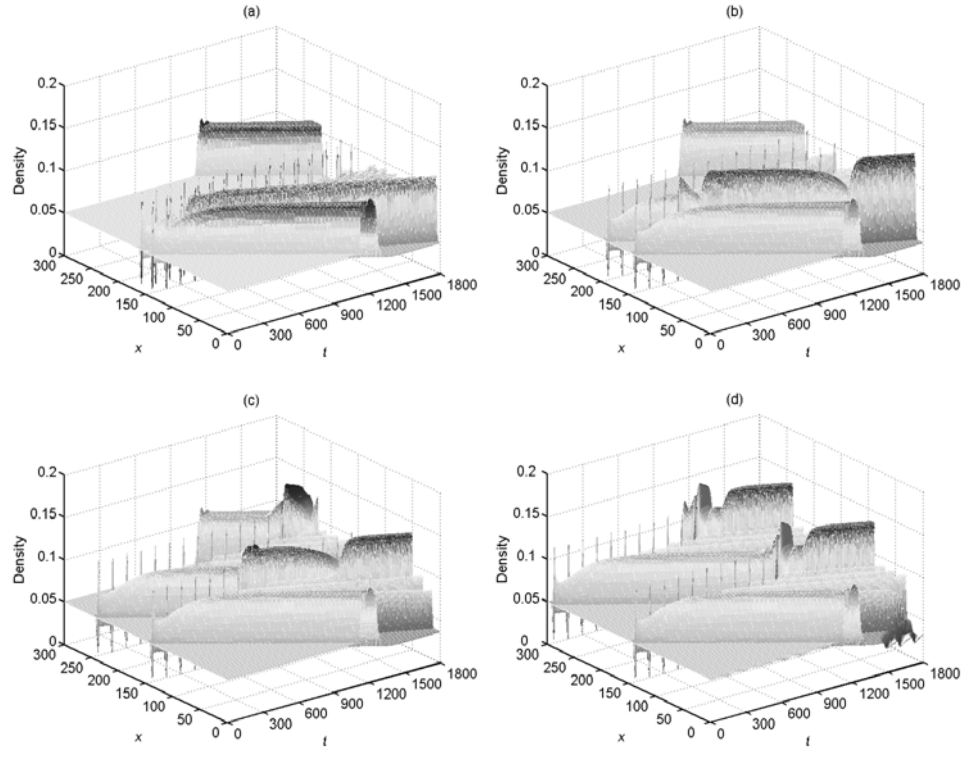


Figure 2. The evolution of the density, where the initial density is 0.05 and the distances in (a)-(d) are respectively 2 km, 5 km, 10 km and 15 km.

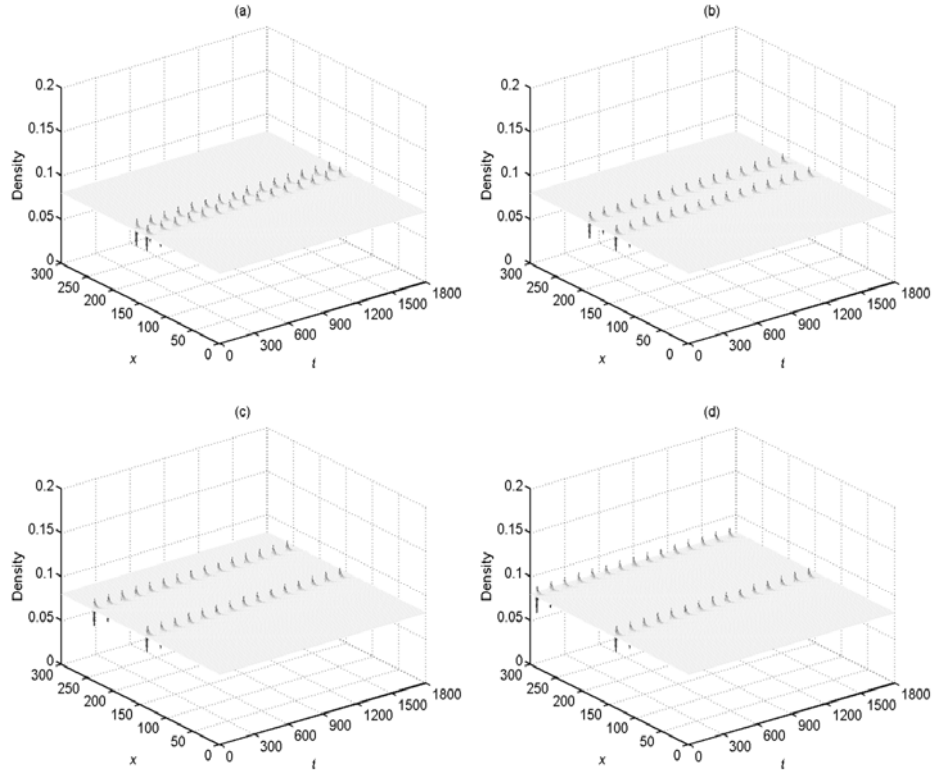


Figure 3. The evolution of the density, where the initial density is 0.08 and the distances in (a)-(d) are respectively 2 km, 5 km, 10 km and 15 km.

Conclusions

In this paper, we adopt the traffic flow model (Tang et al 2008) to study the effects of the spatial distance between two adjacent bus stops on traffic flow and find that this model can reproduce some complex phenomena under the action of different distances. However, we only study some qualitative effects that the spatial distance between two adjacent bus stops has on traffic flow, so we should further study some quantitative properties of the complex phenomena resulted by bus stop by incorporating many experimental data.

Acknowledgement

The work has been supposed by the Program for New Century Excellent Talents in University (NCET-08-0038), the National Natural Science Foundation of China (70701002, 70971007 and 70521001), and the National Basic Research Program of China (2006CB705503).

References

Chowdhury D., Santen L., and Schadschneider A. (2000). “Statistical physics of vehicular traffic and some related systems.” *Physics Report*, 199-329.

Daganzo C. F. (1995). "Requiem for second-order fluid approximations of traffic flow." *Transportation Research B,* 277-286.

Jia B., Li X. G., Jiang R., and Gao Z. Y. (2009). "The influence of bus stop on dynamics of traffic flow." *Acta Physics Sinica,* 6845-6851.

Jiang R., Wu Q. S., and Zhu Z. J. (2002). "A new continuum model for traffic flow and numerical tests." *Transportation Research B,* 405-419.

Kerner B. S. and Konhäuster P (1993). "Cluster effect in initially homogeneous traffic flow." *Physical Review E,* R2335-2338.

Lighthill M. J., and Whitham G. B. (1956). "On kinematic waves: Ⅱ. A theory of traffic flow on long crowded road." *Proceedings of Royal Society Series A,* 317-345.

Nagatani T. (2002). "Dynamical transition in a coupled-map lattice model of a recurrent bus." *Physica A,* 625-636.

Pichards P. I. (1956). "Shock waves on the highway." *Operational Research,* 42-51.

Qian Y. S., Wang H. L., and Wang C. L. (2008) "The study of a cellular automaton traffic flow model with public transit, harbor-shaped bus stop and mixed different-maximum-speed vehicles on single lane." *Acta Physics Sinica,* 2115-2121.

Tang T. Q., Huang H. J., and Xu G. (2008). "A new macro model with the consideration of the traffic interruption probability." *Physica A,* 6845-6856.

Tang T. Q., Li Y., and Huang H. J. (2009). "The effects of bus stop on traffic flow." *International Journal of Modern Physicc C,* 941-952.

Wang M. Q. (1995). "Progress in the study of traffic theory." *Advanced in Mechanics,* 343-356.

Zhang H. M. (1998). "A theory of nonequilibrium traffic flow." Transportation Research B, 485-498.

Zhang P., Liu R. X., and Wong S. C. (2005). "High-resolution numerical approximation of traffic flow problems with variable lanes and free-flow velocity." Physical Review E, 056704.

Zhao X. M., Gao Z. Y., and Li K. P. (2008). "The capacity of two neighbour intersections considering the influence of the bus stop." *Physica A,* 4649-4656.

Traffic Stream Modeling under Heterogeneous Traffic Conditions

Ajitha Thankappan[1], Yamem Tamut[2] and Lelitha Vanajakshi[3]

[1] Research Scholar, [2] Former Post Graduate Student, [3] Assistant Professor, Dept. of Civil Engineering, Indian Institute of Technology Madras, Chennai 600 036, India; Email: lelitha@iitm.ac.in

Abstract

Traffic stream models provide relationship among the three basic macroscopic traffic stream characteristics namely traffic speed, volume and density. Traffic stream models are the basic building blocks of traffic flow modeling, design of road systems and traffic management systems. A number of research papers are available proposing different traffic stream models. However, majority of them are based on homogeneous traffic conditions and they may not be directly suitable for the Indian traffic condition. Not much research has been done on this important aspect so far in India. The present study is an attempt to develop a traffic stream model suitable for the heterogeneous traffic flow condition, such as the one existing in India, taking Chennai as case study. Traffic data for the present study are collected using video graphic technique. Flow, speed and occupancy are manually extracted from collected videos and density estimated using the relation connecting occupancy and density. The results show an exponential speed-density relation and parabolic relations for the other two.

Key Words: traffic stream analysis; heterogeneous traffic; empirical models

Introduction

Traffic flow modeling is defined as the mathematical description of the complex traffic flow system for characterizing and predicting the behavior of traffic. Traffic stream models are the basic building blocks of many of the traffic flow models.

Traffic stream models identify the underlying relationships among the three basic traffic flow variables, speed, flow and density. Knowledge of traffic flow variables and their interrelationships and ways to model traffic are essential in the planning, design and operation of transportation systems. A number of research papers were published proposing different traffic stream models. Some of them began with

mathematical models and some others proposed purely empirical models without much attempt to generalize (Monograph on Traffic Flow Theory 1997). However, as emphasized in this report, the empirical models developed depend on the measurement method used to obtain the data and the location at which the measurements were carried out. Thus, a traffic stream model developed for a homogeneous traffic condition may not suit the heterogeneous traffic condition. As of now, majority of the studies which reported traffic stream models are based on homogeneous traffic conditions and they may not be suitable for the Indian traffic condition. Only limited number of studies is reported from India in this respect as can be seen from the literature review section. The present study is an attempt to develop traffic stream models suitable for Indian conditions.

Literature Review

A brief discussion on the fundamental traffic flow variables (May 1990) and the various stream models proposed by various researchers are given below.

Flow rate or volume is the easiest variable among the three basic variables, both concept wise and in terms of field measurement. It is the rate at which vehicles traverse a segment of road in vehicles/unit time period such as hours. It can be measured manually by standing at a location and noting down the number of vehicles crossing the observer or by using any of the automated vehicle detection technology such as loop detectors.

The second characteristic of traffic flow is the stream speed (in particular, space mean speed) which is defined as the arithmetic mean of speeds of vehicles occupying a given segment of highway. Because of the difficulty in measuring the speeds of all the vehicles occupying a segment of roadway at a given instant, other methods of calculating space mean speeds were adopted. For example, space mean speed can be calculated as harmonic mean of individual speeds that are recorded as vehicles pass an observation point over a selected time interval.

The third parameter, namely density is defined as the number of vehicles occupying a given length of roadway (vehicles/km). Even though density is a fundamental measure of traffic flow, its use in freeway traffic assessment and control has been limited due to the difficulty in measuring it from field (May 1990).
A direct measurement of density from field requires an elevated view of the road length and need techniques such as aerial photography/videography. Because of this practical difficulty, usually traffic density is estimated from other easily measured parameters such as speed, flow, or occupancy. Also, the concept of density was criticized by many especially when traffic is treated as a continuum (Darbha, Rajagopal and Tyagi, 2008).

In spite of all these drawbacks, density is treated as the fundamental macroscopic characteristic of traffic flow since it directly indicates the quality of traffic and ease with which one can drive. A brief discussion of the different methods used for the

estimation of density is given below. The available methods for estimating traffic density can be classified in to two broad categories: direct methods and indirect methods. Direct methods include photographic technique and input-output counts while indirect methods include estimation from other traffic parameters like volume, speed or occupancy. Photographic method involves taking overlapping aerial pictures and counting the number of vehicles in the section during each interval of time (Makigami *et al.* 1985). In the second method namely the input- output technique (May 1990), an initial count of the number of vehicles existing in the roadway between the two count stations is made. To this the number of vehicles entering the section is continuously added and the number of vehicles leaving the section is continuously subtracted to get the density inside the section. However, this technique is practical only if the counts are reliable and accurate.

Indirect methods estimate the density from other easily measurable parameters such as speed, volume or occupancy. The technique to calculate density from speed-flow measurement uses the fundamental traffic flow relation $q=ku$, where q is the flow in veh/h, u is the space mean speed in km/h and k is the density to be calculated in veh/km. One problem associated with this technique is the difficulty in measuring space mean speed from field.

The next technique coming under indirect method is based on the measurement of occupancy, which is the percent of time the detection area is occupied by vehicles. Calculation of density from occupancy uses the formula (May 1990):

$$k = \frac{OR}{(L_v + L_d)\,100}$$

where, k=density (veh/km), Lv=average vehicle length (km), L_d=detection zone length (km) and OR=occupancy ratio and is defined as the ratio of cumulative time taken by all vehicles to traverse through the detection zone to the time interval considered.

Also, under Indian traffic conditions, where traffic is heterogeneous and vehicles do not follow lane discipline, the standard definition of density/ occupancy cannot be used. Under such situations it is only possible to measure density/occupancy per unit length of roadway without considering the lanes and is adopted in this paper too.

Thus, it can be seen that one of the main challenges involved in developing traffic stream models is accurate measurement of the variables involved, mainly the space mean speed and density. In recent times, with the introduction of many of the automated ITS data collection techniques, it is becoming easier to measure these parameters. However, these ITS data collection techniques are mainly available for the lane disciplined and homogeneous traffic conditions, which is not the case in countries such as India. The present study uses video graphic technique to record the traffic and to manually extract the required data on speed, volume, and occupancy as detailed in the data collection section.

Many researchers have investigated the nature of speed-flow-density relationships under homogenous traffic conditions and are discussed below. The first reported study in this area is by Greenshield in 1935 using aerial photographic data collected at regular intervals. Based on the data, an empirical linear relationship between speed and density is assumed and a parabolic speed-flow and flow-density functions were derived from that. Since then a number of models have been suggested to account for the non-linearity between speed and density observed in the field. In general, these models may be classified into two categories: single regime models and multi regime models. Greenberg (1959) developed a logarithmic speed-density curve by treating traffic stream as a continuous fluid mainly based on congested condition of data. Underwood (1961) suggested a single regime model in which a single continuous and differentiable exponential function is assumed between speed and density to describe their relationship for the entire range of traffic conditions from zero to jam density. Drew (1968) proposed a generalized formulation by introducing a parameter *n*, resulting in linear model, parabolic model and exponential model between speed and density with varying values of *n*. Pipes (1967) reported another generalized model with three curves for speed-density, one a straight line and others exponential. Later, multi regime models were proposed with different relations suggested for different range of traffic. For example, a two regime model was suggested by Edie (1961) with different models for congested and un-congested conditions assuming Greenberg's model for the congested region and Underwood's model for the other regime. Drake *et al.* (1967) empirically tested the models proposed by Greenshield, Greenberg, Underwood and Edie along with a bell shaped speed-density relation and concluded that none of the models tested provide a good fit to the data. The reason for this was analyzed by Duncan (1976) in a study and noted that the consequence of taking a relatively good fitting speed-density function and transferring it to a speed-flow function (as done by Drake et al.) did not fit well with the original speed-flow data. In another study conducted by Duncan (1979) using field data proved that minor changes in the nature of speed-density function resulted in major changes in speed-flow function. In this study he fitted a third order polynomial regression equation for speed-density and a fourth order polynomial function for speed- flow function. Hall *et al.* (1986), using occupancy data directly as a measure of concentration than converting to density, developed an inverted V- shaped empirical relationship between flow and occupancy. Cassidy and Coifman (1997) investigated the relationship between occupancy and density and concluded that density and occupancy are linearly related to average vehicle length. This relationship was then empirically tested by Kim and Hall (2004) by considering average vehicle length as a variable and the results showed that if the proportion of long vehicles (trucks) is high, the average vehicle length has large variation. The relationship developed could be fitted with linear as well as quadratic functions. Recently a multiregime model was developed by Sun and Zhou (2005) by using cluster analysis as a natural tool for segmentation of speed-density data for different regimes After data segmentation, using regression analysis three functional forms namely linear, exponential and logarithmic functions are fitted to represent the three regimes of traffic flow namely free flow, congested flow and transition state.

Overall, from the literature it can be seen that most of the models were validated using the data collected from western countries with homogeneous and lane disciplined traffic conditions. Only limited amount of research has been reported from Indian traffic conditions in recent years. Kadiyali *et.al* studied the speed flow relationships on different categories of rural highways in India in 1982. They developed linear speed-flow relationships by regression analysis. In a study conducted by Kumar and Rao (1998), on selected stretches of NH 5 and NH 6, linear relation between speed and density and a non- linear relation connecting speed and flows were proposed. Later Khan and Maini (1998) made a review on available models of non-lane based, and heterogeneous traffic conditions and pointed out the need for larger data collection efforts to further validate these models

Thus it can be seen that, most of these reported studies on traffic stream modeling using data from Indian traffic conditions are limited to speed volume relation alone because of the difficulty in measuring density. The data collections in those studies were carried out manually in the field. Also, the traffic condition in India has changed tremendously in the last decade and the stream models developed in the last decade may not be suitable for the present traffic scenario. The present study uses advanced data collection techniques and obtains accurate measurement of the variables namely speed, flow and density and develops the stream models.

Data Collection and Extraction

Data collection for the present study was carried out using video recording technique mainly due to following reasons. Video recording is a widely used and cost effective means of traffic data collection which can provide a comprehensive and permanent record of traffic movements. One can extract various traffic parameters from the same recorded video. Also recorded video tapes can readily be used to observe the same traffic scenario repeatedly as many times as required during analysis. Data for the present study were collected from a 1 km stretch on the Rajiv Gandhi road, Chennai, India. The selected stretch is a 6 lane roadway, with 3 lanes in each direction. For the present study only one direction of traffic is considered. Video data were collected from pedestrian over bridges available at the entry and exit points during the one hour afternoon peak periods on 3 week days.

The collected videos were later analyzed in laboratory to manually extract the required input values for the development of stream models. The present study extracted flow, average time mean speed and average occupancy of all vehicles at every one minute interval from videos as explained below.

Traffic flow was determined by counting the number of vehicles passing the entry and exit sections of the study stretch from all the three lanes at every one minute interval and expressed on an hourly basis. To determine the occupancy time and average speed, two lines separated by a known distance are selected in the computer screen and measurement of the time taken by different categories of vehicles to pass this detection zone were carried out at every one minute interval. From these travel

times and known distance the average time mean speeds of vehicles are computed. The density and space mean speed required for stream modeling were then estimated as explained in the following section.

Analysis of Data

The information extracted from video are further analyzed to estimate density and space mean speed values required for the development of stream models.

Estimation of Density. One of the previous studies reported occupancy as the best representative of density under Indian traffic conditions (Padiath *et al*. 2010). Hence, in the present study, density was estimated from occupancy values extracted from videos. However, the occupancy measured in this study is slightly different from the classical definition of occupancy as available in literature. Here the time taken for a vehicle to cross two selected lines separated by a known distance is measured as the time interval between the front of the vehicle touching the entry line and the front of the same vehicle touching the exit line. Hence the length of the vehicle is not included in the occupancy time in this case and is represented as occupancy rate (*OR*). Hence, the relationship between occupancy and density in this case will differ from the classic one (May 1990) and is derived as follows.

Let t_0 be the time difference between front of an individual vehicle touching the entry and exit lines, which are at a distance L_d apart, then speed of this individual vehicle can be represented as,

$$u_i = \frac{L_d}{t_0} \tag{1}$$

Converting the relationship (1) to a macroscopic level for N number of vehicles crossing the detection zone with average value of occupancy times, $\overline{t_0}$ during a total time period T, the average space mean speed,

$$u_{sms} = \frac{L_d}{\overline{t_0}} \tag{2}$$

Then the cumulative time taken by N vehicles to traverse through the detection zone,

$$T_0 = N\overline{t_0} \tag{3}$$

Solving for $\overline{t_0}$ in equation (2) and substituting into equation (3) yields,

$$T_0 = N\frac{L_d}{u_{sms}}$$

Now occupancy rate (*OR*) which is defined as the ratio of cumulative time taken by N vehicles to traverse through the detection zone to the total time interval considered, can be represented as,

$$OR = \frac{T_0}{T} = N\frac{L_d}{u_{sms}}\frac{1}{T} = q\frac{L_d}{u_{sms}} \quad (\text{Since flow rate}, q = \frac{N}{T})$$

$$= kL_d \quad (\text{Since density}, k = \frac{q}{u_{sms}})$$

Thus, density,
$$k = \frac{OR}{L_d} \tag{4}$$

Hence density can be computed for each time step by using equation (4), if the occupancy rate and detection zone length is known. In the present study the above relation was used for density estimation at one minute intervals.

***Estimation of Space Mean Speed*:** The speed values obtained from video are time mean speeds and needs to be converted to space mean speeds before using these for stream modeling. This was carried out using equation (Wardrop 1952) given by,

$$u_{tms} = u_{sms} + \frac{\sigma_{sms}^{\ 2}}{u_{sms}}$$

which can be reduced to (May 1990),

$$u_{sms} = \frac{u_{tms}}{1 + c_v^{\ 2}}$$

where, u_{tms}, u_{sms}, σ_{sms} and c_v represent time mean speed, space mean speed, standard deviation of space mean speed and coefficient of variation of space mean speed respectively.

May (1990) has suggested that value of c_v lies between 0.18 and 0.17. In the present study, space mean speed values were estimated using the above relation using a coefficient of variation of 0.17.

Once the required data are extracted from the video, appropriate stream models such as speed-density, speed-flow and flow-density are developed and are discussed in section below.

Results and Discussion

Using the extracted data on speed, flow and density in the previous section speed-density, flow-density and flow-speed relationships were plotted and shown in Figures 1, 2 and 3, respectively. Various functions were fitted to this data and the best fitting ones were taken.

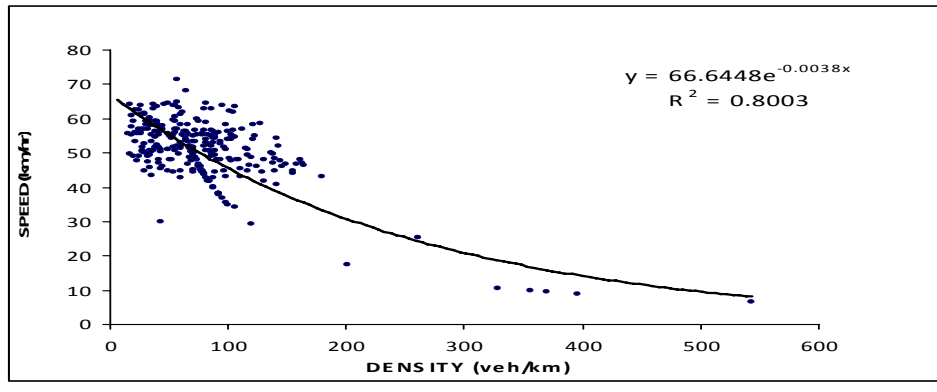


Figure 1. Speed-density relationship.

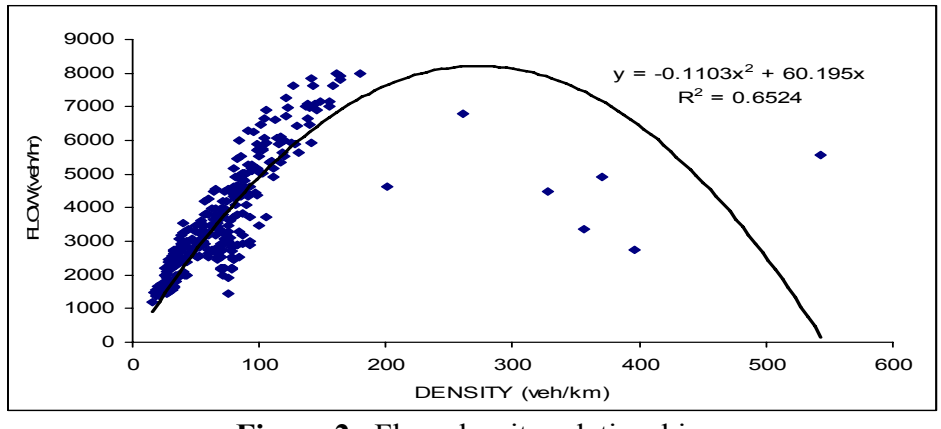


Figure 2. Flow-density relationship.

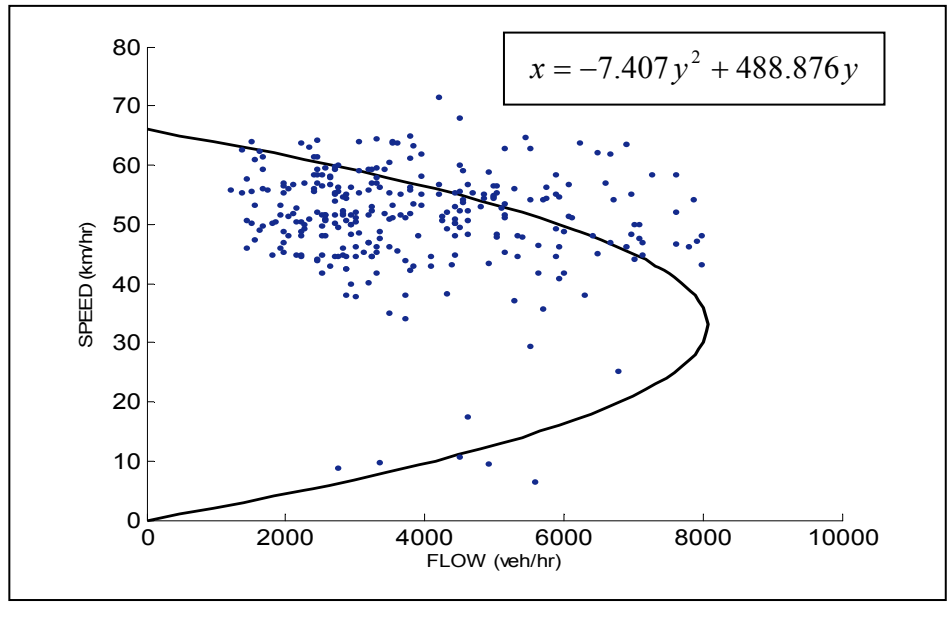


Figure 3. Speed-flow relationship.

The speed density data displayed a trend which is as expected with the density increasing as the speed decreases. The exponential relationship was found to be the best for speed-density data and was:

$$u = 66.64e^{-0.0038k}$$

This relationship is analogous to the Underwood's exponential speed-density relationship of the form,

$$u = u_f e^{-k/k_m}$$

where, u is the speed at density k, u_f the free speed and k_m represents the density at which flow experiences its maximum with respect to density. If we compare this with the observed relationship, the present study shows a free speed of 66.64 km/h, and a density at maximum flow (optimum density) as 265 veh/km.

Flow- density plot shows a very consistent data point pattern at free flow regime but the same is not so prominent in congested regime due to lack of sufficient data points. A parabolic relationship is fitted to this data and the maximum flow or capacity appears to be 8213 veh/hr and optimum density is observed to be 273 veh/km.

However, the flow speed data was not following any specific trends and hence was plotted using the obtained free flow speed, and capacity values. The actual data were found to be agreeing to this relation. However, more data points are needed to draw the final conclusion on this.

Conclusions

Modeling of traffic stream is important as stream models are the foundation for traffic flow modeling and for planning, design and effective management of roads systems. Models developed under homogenous and lane disciplined traffic conditions may not suit the heterogeneous and less lane disciplined traffic conditions as in India. Indian traffic is highly heterogeneous with vehicles having widely varying static and dynamic characteristics moving through the same roadway without any segregation. Lack of lane discipline adds to the complexity because vehicles can take any lateral position based on space availability. Hence, there is a need to develop stream models for Indian traffic conditions that can be used as building blocks for other areas of research. The present study attempted to develop empirical traffic stream models appropriate for Indian conditions. Three days traffic data collected by video graphic technique from a road stretch in Chennai, India were used for the study. The required traffic parameters were extracted from the video manually for better accuracy and are used for deriving the empirical models. All the traffic variables are measured without considering the traffic lanes. The results showed an exponential speed-density relation and parabolic speed-flow and flow-density relationships. The models indicated a free speed of about 66 km/h, optimum density of 265-273 vehicle/km and capacity of 8213 veh/hr.

Acknowledgement

The third author acknowledges the support provided by the Ministry of Urban Development, Government of India through grant No. N-11025/30/2008-UCD.

References

Cassidy, M. J., and Coifman, B. (1997). "Relation among average speed, flow and density and analogous relation between density and occupancy", *Transportation research record 1591*, Transportation Research Board. Washington, D.C.

Darbha, S, Rajagopal, K. R., and Tyagi, V. (2008). "A review of mathematical models for the flow of traffic and some recent results", *Nonlinear Analysis*, 69, 950-970.

Drake, J. S, Schoter, J. L., and May, A. D. (1967). "A statistical analysis of speed density hypothesis", *Highway research record*, 154, 53-87.

Drew, Donald. R. (1968). *Traffic Flow theory and control*, New York: McGraw-Hill.

Duncan, N. C. (1976). "A note on speed/flow/concentration relation", *Traffic Engineering and Control*, 17, 34-35.

Duncan, N. C. (1979). "A further look at speed/flow/concentration", *Traffic Engineering and Control*; 20, 482-483.

Edie, L. C. (1961). "Car following and steady state theory for non-congested traffic", *Operations Research*, 9, 66-76.

Greenberg, H. (1959). "An analysis of traffic flow", *Operations Research*, 7, 78-85.

Green shields, B. D. (1935). "A study of traffic capacity", *Highway research board proceedings*, 14, 448-477.

Hall, F. L, Allen, B. L., and Gunter, M. A. (1986). "Empirical Analysis of Freeway flow-density relationships", *Transportation Research*, 20A, 197-210.

Hall, F. L., and Gunter, M. A. (1886). "Further analysis of flow-concentration relationship", *Transportation research record 1091*, Transportation Research Board. Washington, D.C.

Kadiyali,L. R, Viswanathan, E, Guptha, R., and Ummat, S. K.(1982). "Speed volume studies on rural highways in India", *Indian Road Congress Journal*, Indian Road Congress, New Delhi, 43, 359-435.

Khan, S. I., and Maini, P. (1998). "Modeling heterogeneous traffic flow", *77th Annual meeting, Transportation Research Board*, Washington D.C.

Kim, Y., and Hall, F. L. (2004). "Relationships between occupancy and density reflecting average vehicle lengths", Presented at TRB 2004 annual meeting (CDROM).

Kumar, V. M. (1998). "Studies on speed-density-flow relationships on a few stretches of N.H 5 and N.H 6", *Indian Highways*, Indian Road Congress, New Delhi, 26, 33-41.

Makigami, Y, Sakamolo, H., and Hayashi, M. (1985). "An analytical method of traffic flow using aerial photography." *Journal of Transportation Engineering*, 4, 347-394.

May, A.D. (1990). *Traffic flow fundamentals*, Prentice Hall.

Padiath, A, Vanajakshi, L., and Subramanian, S. C. (2010). "Estimation of traffic density under Indian traffic conditions", Paper accepted for poster presentation in Transportation Research Record, Transportation Research Board, National Research Council,Washington, D.C.

Pipes, L. A. (1967). "Car-following models and the fundamental diagram of road traffic", *Transportation Research*, 1, 21-29.

Revised Monograph on Traffic Flow Theory: A State-of-the-art Report. (1997). Transportation Research Board, National Research Council, Washington, D.C.

Sun, L., and Zhou, J. (2005). "Development of multiregime speed-density relationships by cluster analysis", *Transportation research record 1934*, Transportation Research Board. Washington, D.C.

Underwood, R. T. (1961). "Speed, volume and density relationships: Quality and theory of traffic flow", *Yale bureau of highway traffic*, 141-188.

Wardrop, J. G. (1952). "Some Theoretical Aspects of Road Traffic Research", *Proceedings of the Institution of Civil Engineers*, Part II, 1, 325-362.

A Novel Method for Traffic Bottlenecks Identification in Urban Traffic Network

Yu Zhou, Jianming Hu, Yi Zhang and Huasheng Yang

Department of Automation, Tsinghua University, Beijing 100084, China; Email: yu-zhou07@mails.tsinghua.edu.cn

Abstract

In the urban traffic network, concurrent traffic congestions are serious problems. They usually occur in fixed positions called bottlenecks. They are sources of the traffic congestions which influence the efficiency of the traffic network. Therefore the first step to improve the transport performance is identifying the traffic bottlenecks. This paper presents a novel two-phase method to identify the traffic bottlenecks in urban road network and analyzes their effect. The method presented in this paper integrates both urban planning and actual traffic flow. In 1st phase, the initial possible areas containing bottlenecks are determined. We use urban planning data to determine the reliability of the traffic network and the scope of possible traffic bottlenecks based on fuzzy c-means (FCM) algorithm. Then, in 2nd phase, we use the bottleneck index denoting the match degree between actual traffic data and the network physical attributes to identify the exact bottleneck positions. Finally, we test the effectiveness of two-phase method using an actual urban traffic network. The chosen road network based on network in Jilin, China is built up by Paramics Software according to the actual data obtained from Jilin Traffic Management bureau and Jilin Construction Commission. The simulation results perform in accordance with the practical situation, which can verify the applicability and feasibility of the proposed method.

Key Words: traffic bottlenecks; urban network; two-phase method; traffic planning

Introduction

Traffic congestions are severe problems in urban traffic network. They usually occur in some fixed locations called bottlenecks, which seriously decrease the efficiency and effectiveness of the whole traffic network. Thus, identifying the traffic

bottlenecks is the first step to relieve the traffic congestion and improve the traffic system's performance, and it is also essential for traffic planning.

In 1999 and 2004, American Highway Users Alliance published two study reports respectively. In 1999's report (American Highway Users Alliance, 1999) they emphasize the influence of traffic bottlenecks and in 2004's report (American Highway Users Alliance 2004) they analyzed the effect of improving the traffic bottlenecks. During 1991, Banks, J. H. studied the characteristics of capacity of the highway bottlenecks. Part of the theory (Banks, 1991) was used as the theoretical basis for ramp metering. Since 1995, Cassidy and Daganzo committed themselves to the study of highway bottlenecks and made a series of achievements (Cassidy, 1995). And then in 1997, Daganzo, C. studied the queue length and location in upstream, using the curve of vehicle arrival (Lawason et al., 1997). In 2000, Huapu Lu et al. analyzed the traffic congestions in urban network, and proposed some relief methods (Lu, 2000). In 2002, Qingdong Da and Guowu Zhangstudied the relationship between flow and occupancy of traffic congestions (Da, 2002). In 2005, Yizhou Wu pointed the contradiction of traffic supply and demand was an important issue leading to traffic congestions (Wu, 2005). In 2007, Haixian Jiao proposed a method for bottleneck identification based on network capacity equilibrium model (Jiao, 2007).

This paper presents a novel two-phase method to identify the traffic bottlenecks in urban road network and analyzes their effect. The existed approaches of judging the traffic bottlenecks are mainly based on evaluating traffic flow parameters such as flow, velocity and occupancy. A novel method presented in this paper integrates both urban planning and actual traffic flow, which gradually narrows the search space for traffic bottlenecks according to the physical attributes and traffic flow conditions of urban traffic network. In 1st phase, the initial possible scopes containing bottlenecks are determined by network hierarchical judgment and road sections clustering analysis by fuzzy c-means (FCM) algorithm. In 2nd phase, we obtain the bottleneck index of road sections to identify the exact positions of bottlenecks. The bottleneck index denotes the match degree between actual traffic data and the network physical attributes. Finally, we establish a visual simulation environment to determine traffic conditions with Paramics software based on actual urban network data.

This paper is organized as follows. Section 2 sums up related works. Section 3 explains the distinguishing features of traffic bottlenecks. Section 4 provides the two-phase method of identifying bottlenecks. In section 5 we applies the two-phase method to the Changyi District in Jilin City of Jilin Province in China. Finally, we conclude our work.

Features of Traffic Bottlenecks

Traffic bottlenecks are the vulnerable points of the traffic network which severely influence the transport efficiency. When the traffic flow in one road section exceeds the capacity, the vehicles passing through this section will slow down. Consequently,

the congestion will be formed. Such a road section is called a traffic bottleneck. They usually appear in ramps and intersections. Obviously, traffic bottlenecks are sources of the traffic congestions.

Traffic bottlenecks are caused by multiple factors including inappropriate road design, behaviour of disobey the traffic rules, etc. Nevertheless, the essential reason for traffic bottlenecks arises from the incompatibility between limited traffic resources and increasing traffic demands. Once road network was completely constructed, its capacity is restricted. However, traffic demands are continuously increasing. This contradiction between traffic supplies and demands leads to serious traffic congestions over the whole traffic network. As the most vulnerable points of the whole network, the bottlenecks have even worse situation. Since the excessive traffic flows from upstream cannot pass the bottlenecks, waiting queues will gradually increase to the upstream intersections, which will lead to a wide range of traffic jams. This is a key issue which should be settled down urgently.

Two-Phase Method

The method we proposed here integrates both from urban planning and actual traffic flow viewpoints as follows. In 1st phase, we analyze the road network hierarchical structure and use FCM to locating the initial scopes of bottlenecks. In 2nd phase, we determine the bottleneck index of road sections in the initial scopes of bottlenecks to indicate the possibility of traffic bottlenecks. Through this method, we can provide technical support to city planning authorities.

The 1st Phase: Locate Initial Scopes of Bottlenecks. In 1st phase, we aim to find the initial scopes of bottlenecks. First, we analyze the traffic structure hierarchically based on the planning and construction data. A well constructed traffic network should have a reasonable ratio between the expressways, trunk roads, sub-trunk roads and branches. Assuming that the ratio of the length between the expressways (L_1), trunk roads (L_2), sub-trunk roads (L_3) and branches (L_4) is expressed as

$$L_1 : L_2 : L_3 : L_4 = a_1 : a_2 : a_3 : a_4 \tag{1}$$

Then the hierarchical structure, denoted by the ratio of all kinds of roads' length, should be

$$l_1 : l_2 : l_3 : l_4 = 1 : \frac{a_2}{a_1} : \frac{a_3}{a_1} : \frac{a_4}{a_1} \tag{2}$$

Additionally, the trunk roads, sub-trunk roads and branches should follow the rules

$$L_2 \le L_3 \le L_4 \tag{3}$$

$$L_2 + L_3 \le L_4 \tag{4}$$

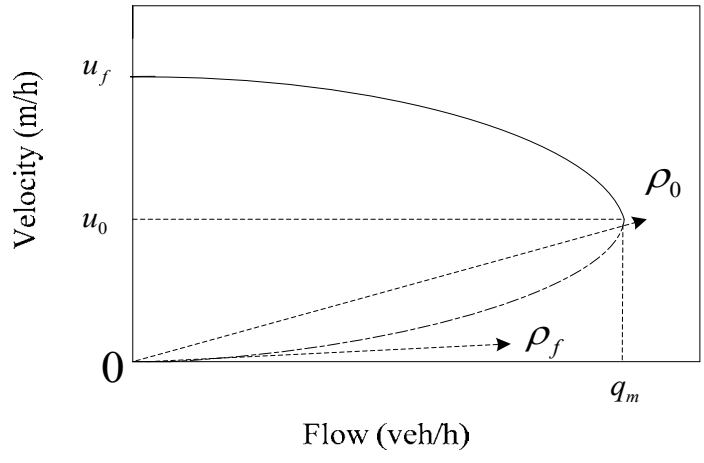


Figure 1. The relationship between flow rate and velocity. $\rho_0\, u_0, q_m$ are the parameters in the critical state.

A feasible hierarchical structure should be able to ensure that the traffic flows are assembled from low-level roads to high-level roads, and also evacuated from high-level roads to low-level roads. Therefore the hierarchical structure of the traffic tends to have a pyramid shape (Shi, 2007).

Second, we separate the road sections which are most vulnerable to traffic congestion according to the traffic flow data. These sections are regarded as the most likely locations of traffic bottlenecks. Furthermore, a proper hierarchical structure is a necessary condition to guarantee the efficiency of traffic network, but not the sufficient condition. It only shows the preliminary structure of the traffic network. Thus, we have to analyze the traffic flow data to judge the traffic status. Generally, the characteristics of traffic flow can be described with parameters of flow rate (q), density (ρ) and velocity (u). Figure 1 shows the relationship between flow rate and velocity, in which the dash-and-dot line stands for traffic congestions since both the flow rate and the velocity are decreasing. Therefore, the road sections which have the feature displayed as the dash-and-dot line in Figure 1 are the possible bottleneck locations. With respect to the uncertainty and ambiguity of traffic status, we used FCM to separate the initial bottlenecks scopes according to the flow rate and velocity of different road sections at different time. FCM developed by Dunn in 1973 and improved by Bezdek in 1981 is a clustering algorithm to define degree of each data to every cluster using the membership function which allows one piece of data to belong to two or more clusters. Its feature is suitable for analyzing the urban traffic situation.

The 2nd Phase: Determine the Bottleneck Index. In 2nd phase, the match degree between the designed capacity and the actual traffic volume is calculated to identify the exact positions of the traffic bottlenecks, specifically for the road sections in the initial bottlenecks scopes obtained by the first phase. We define this match degree as bottleneck index B to indicate the possibility of traffic bottlenecks. In urban traffic

network, the vehicles with different origins and destinations turn to different directions when they pass through an intersection. Turning rate T can reflect the traffic demands which can be calculated based on real-time traffic data. Since the road capacity of every direction is fixed for each intersection in the planning period, we take road capacity rate R of intersections as a substitution of the traffic supply. Then we compare the turning rate and capacity rate of intersections to find the relationship between traffic demands and supplies, which is bottleneck index.

According to road design data, we calculate the capacity rate R of intersections by Equation (5)

$$R = \frac{n_{turn} l_{turn} d_{turn}}{\sum_i n_i l_i d_i} \tag{5}$$

where, n_{turn} is the designed density of turning direction, l_{turn} is the road length of turning direction, d_{turn} denotes the lane width of turning direction and $\sum_i n_i l_i d_i$ stands for the road capacity of total turning directions respectively. With the flow rate of every direction of each intersection, we calculate the turning rate T by

$$T = \frac{q_{turn}}{\sum_i q_i} \tag{6}$$

Where, q_{turn} is the vehicle number of turning direction, and $\sum_i q_i$ is total number of turning vehicle. Compare the measured turning rate T and the carrying capacity rate R, we define the bottleneck index B as

$$B = \frac{T}{R} \tag{7}$$

The closer B is to 1, the higher match degree of the road section we get. If $B<1$, it can be inferred that the road section is not a bottleneck since the designed traffic capacity of the section is adequate to the traffic demand. If $B>1$, it represents that the designed traffic capacity cannot sustain the traffic flows. The greater the bottleneck index B is, the more vulnerable to traffic congestion the section is. Therefore, the road sections with high bottleneck index are judged as the traffic bottlenecks.

Experiment and Analysis

We proved the effectiveness of two-phase method using an actual traffic network. The chosen network is the Changyi District of Jilin City, Jilin Province, China. We established the simulation model by Paramics Software. We calibrate the model with

actual data which obtained from traffic investigation and Jilin government. After running the simulation model, we observed the traffic congestions in certain roads which matched the fact.

Road Network Model. The chosen network lies by Songhua River as shown in Figure 2(a). This area is a typical urban traffic network which is mainly constituted by residential districts, business districts, government offices and schools. We established a visual simulation environment of the chosen network based on Paramics software to obtain more detailed data to determine traffic conditions, which used both planning and practical traffic data including infrastructures, GIS information and O/D data. Descriptions of the infrastructures, including signal locations and settings, are obtained from Jilin Traffic management bureau. The O/D data with an interval of 20 minutes is obtained from OD survey by Jilin Construction Commission. This network contains 6 trunk roads, 16 sub-trunk roads and 39 intersections (all of them are signalized). The detail information of the network is shown in Table 1 and the simulation network is shown in Figure 2(b).

Table 1. Road information of Changyi, Jilin, Jilin, China.

No	Name	Type	Orientation	Length(m)	Lane
1	Jilin Street (middle)	trunk	North-South	2200	8
2	Jilin Street (north)	trunk	North-South	2700	6
3	Jiefang Road(east)	trunk	East-West	2000	6
4	Jiefang Road (north)	trunk	North-South	4300	6
5	Tianjin Street	trunk	North-South	2200	4
6	Tongtan Road	trunk	East-West	1850	4
7	Sonegjiang Road (east)	sub-trunk	East-West	6850	2
8	Songjiang Road (north)	sub-trunk	North-South	3800	2
9	Tongjiang Road	sub-trunk	East-West	2255	2
10	Zhongxing Street	sub-trunk	North-South	3700	4
11	Hadawan Street	sub-trunk	East-West	2800	2
12	Chongqing Road	sub-trunk	North-South	2300	2
13	Zhongkang Road	sub-trunk	East-West	600	6
14	Weichang Road	sub-trunk	North-South	730	4
15	Liaoning Road	sub-trunk	East-West	750	4
16	Kangzhuang Road	sub-trunk	North-South	400	2
17	Shanghai Road (west)	sub-trunk	East-West	854	2
18	Shanghai Road (east)	sub-trunk	East-West	859	4
19	Xinchang Road	sub-trunk	North-South	950	4
20	Yan'an Road	sub-trunk	East-West	1704	4
21	Qingnian Road	sub-trunk	East-West	1430	2
22	Jiangwan Road	sub-trunk	East-West	2235	4


(a) Picture from Google Earth


(b) Paramics project

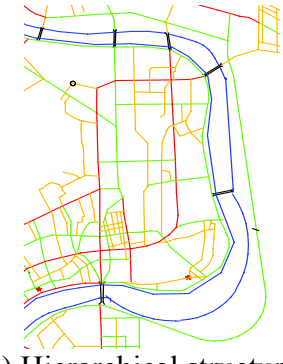

(c) Hierarchical structure

Figure 2. Experiment network.

Table 2. Statistical results of hierarchical structure of experimental network.

Type	Num	Length (m)
Trunk road	6	15250
Sub trunk road	16	32217
Branch	45	71649

Implementation and Analysis. We implemented the two-phase method on the experimental network to identify the exact locations of traffic bottlenecks.

In 1st phase, firstly we judged the hierarchical structure of experimental network, as shown in Figure 2(c). The red lines represented the trunk roads, green lines represented the sub-trunk roads, yellow lines represented the branches and blue lines represented Songhua River. Combining with the data in Table 1, we obtained the hierarchical structure of experimental network, as displayed in Table 2. Because there are no expressways in experimental network, applying Equation (1) and (2), we have

$$L_2 : L_3 : L_4 = 15250 : 32217 : 71649 \tag{8}$$

$$l_2 : l_3 : l_4 = 1 : 2.1 : 4.7 \tag{9}$$

According to Eq. (9), we noted that the hierarchical structure is basically conforming to the pyramid structure but the trunk roads are not enough. The existed trunk roads were not enough to take the traffic flows from sub trunk roads and branches.

Then, we used FCM to separate the initial bottlenecks scopes according to traffic flow rate and velocity of different road sections in the morning. According to Paramics Software modeling setting, the 22 roads (trunk roads and sub trunk roads) were divided into 583 sections and obtain the flow rate and velocity of each section per minutes from 6:00am to 10:00am. We clustered the flow rate and velocity of 583 sections into four clusters corresponding to block, jam, regular and smooth traffic statuses. The results are shown in Figure 3 and Table 3.

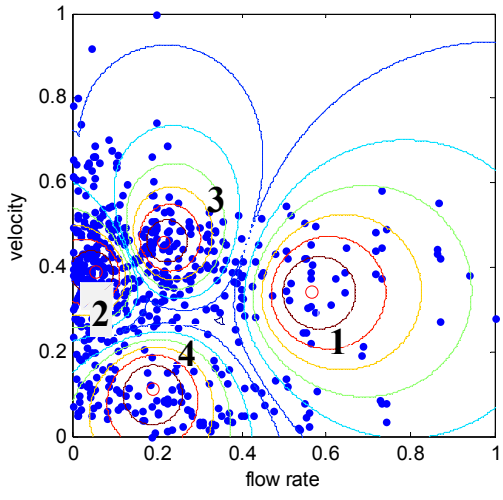


Figure 3. The clustering result of the flow and velocity data from 6 a.m to 10 a.m. (The data was normalized before we performed the FCM)

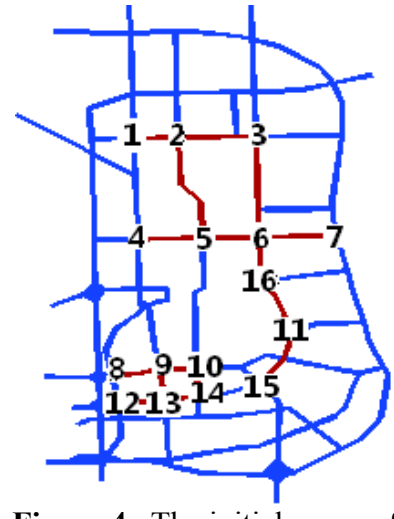


Figure 4. The initial scope of bottlenecks and adjacent intersections. (The red lines represent the possible bottlenecks locations and the numbers are adjacent intersections)

Table 3. The clustering centers of 583 links.

ID	Flow	Velocity	Section Num	Status
1	0.5674	0.3397	78	Block
2	0.0546	0.3901	230	Smooth
3	0.2158	0.4573	154	Regular
4	0.1891	0.1101	121	Jam

Compare Figure 1 with Figure 3, we concluded that Cluster 1 and Cluster 4 represent block and jam status, and Cluster 2 and Cluster 3 represent smooth and regular status. Consequently, the initial bottlenecks scopes are the sections in Cluster 1 and Cluster 4, including Tongjiang Road, Yan'an Road, Zhongxing Street, Shanghai Road and several sections of Jilin Street and Jiefang Road as shown in Figure 4. Applying Eqs. (6) and (5), we calculated turning rate (T) and capacity rate (R) of 16 intersections as shown in Tables 4 and 5, respectively.

From Figure 5, we can see the possible bottleneck links adjoin more than one intersection, so they influenced by different turning rates of different intersections. Due to the signal control system in urban traffic network, the link traffic flows is from one direction in one phase cycle. Therefore we choose the direction which has most contribution to the link to calculate the bottleneck index B by Eq. (7). The calculated results of possible bottleneck links are illustrated in Table 6.

Table 4. The turning rate (*T*) of 16 intersections.

No	From	North (N)			East (E)			South (S)			West (W)		
	To	E	S	W	N	S	W	N	E	W	N	E	S
1		0.00	1.00	0.00	0.00	1.00	0.00	0.99	0.01	0.00	0.00	1.00	0.00
2		0.13	0.83	0.04	0.03	0.91	0.06	0.22	0.78	0.00	0.01	0.99	0.00
3		0.00	0.84	0.16	0.00	0.00	1.00	0.62	0.00	0.38	0.57	0.28	0.15
4		0.75	0.09	0.16	0.24	0.00	0.76	1.00	0.00	0.00	0.13	0.87	0.00
5		0.74	0.26	0.00	0.41	0.22	0.37	0.70	0.25	0.05	0.02	0.93	0.05
6		0.00	0.44	0.56	0.75	0.00	0.25	0.83	0.08	0.09	0.13	0.84	0.03
7			0.61	0.39				1.00		0.00	1.00		0.00
8		0.38	0.33	0.28	0.56	0.00	0.44	0.81	0.00	0.19	0.37	0.57	0.06
9		0.22	0.27	0.51	0.04	0.02	0.93	0.89	0.11	0.00	0.21	0.79	0.00
10		0.16	0.84	0.00	0.64	0.36	0.00	0.59	0.41	0.00	0.32	0.34	0.34
11		0.00	1.00		1.00	0.00		0.34	0.66				
12		0.04	0.96	0.00	0.36	0.00	0.64	0.01	0.99	0.00	0.00	1.00	0.00
13		0.18	0.20	0.62	0.25	0.00	0.75	0.80	0.20	0.00	0.62	0.38	0.00
14		0.00	0.92	0.08	0.00	0.65	0.35	0.32	0.67	0.01	0.48	0.05	0.48
15		0.01	0.88	0.10	0.12	0.00	0.88	0.94	0.01	0.05	0.55	0.12	0.32
16		0.05	0.95		0.95	0.05		0.98	0.02				

Table 5. The capacity rate (*R*) of every intersection.

No	From	North (N)			East (E)			South (S)			West (W)		
	To	E	S	W	N	S	W	N	E	W	N	E	S
1		0.21	0.58	0.21	0.46	0.39	0.15	0.62	0.19	0.19	0.47	0.14	0.39
2		0.38	0.42	0.21	0.52	0.32	0.16	0.54	0.30	0.16	0.46	0.25	0.28
3		0.15	0.71	0.14	0.34	0.55	0.11	0.60	0.21	0.19	0.33	0.12	0.55
4		0.34	0.44	0.23	0.46	0.36	0.19	0.50	0.30	0.20	0.42	0.25	0.33
5		0.28	0.41	0.32	0.29	0.40	0.31	0.33	0.31	0.36	0.30	0.28	0.41
6		0.41	0.38	0.20	0.30	0.46	0.24	0.29	0.48	0.23	0.24	0.40	0.37
7			0.17	0.83				0.14		0.86	0.44		0.56
8		0.42	0.32	0.43	0.35	0.37	0.28	0.31	0.45	0.25	0.28	0.42	0.30
9		0.23	0.46	0.31	0.57	0.26	0.17	0.65	0.15	0.20	0.59	0.14	0.27
10		0.42	0.40	0.18	0.35	0.44	0.20	0.34	0.46	0.20	0.28	0.37	0.35
11		0.50	0.50		0.75	0.25		0.74	0.26				
12		0.65	0.07	0.28	0.29	0.15	0.56	0.14	0.60	0.26	0.17	0.75	0.09
13		0.36	0.14	0.50	0.27	0.16	0.57	0.22	0.33	0.46	0.32	0.49	0.19
14		0.57	0.15	0.28	0.32	0.24	0.44	0.19	0.54	0.27	0.22	0.61	0.16
15		0.28	0.30	0.43	0.18	0.33	0.48	0.19	0.32	0.49	0.22	0.38	0.40
16		0.36	0.64		0.57	0.43		0.71	0.29				

Table 6. The bottleneck index of each link.

ID	1	2	3	4	5	6	7	8
Section	$L_{1,2}$	$L_{2,3}$	$L_{2,5}$	$L_{3,6}$	$L_{4,5}$	$L_{5,6}$	$L_{6,7}$	$L_{6,16}$
T	1	0.99	0.91	0.83	0.87	0.56	0.84	7.07
R	0.14	0.35	0.31	0.18	0.10	0.20	0.46	2.85
B	7.07	2.85	2.87	4.64	8.40	2.79	1.83	2.87
ID	9	10	11	12	13	14	15	16
Section	$L_{8,9}$	$L_{9,10}$	$L_{12,13}$	$L_{13,14}$	$L_{11,15}$	$L_{16,11}$	$L_{10,14}$	$L_{9,13}$
T	0.44	0.93	0.66	0.99	0.48	0.95	0.80	0.36
R	0.17	0.20	0.16	0.18	0.14	0.64	0.19	0.10
B	2.58	4.59	4.23	5.34	3.37	1.48	4.09	3.56

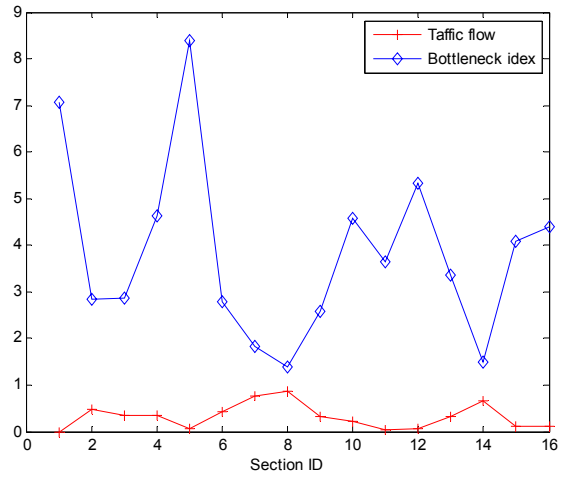


Figure 5. Compare of bottleneck index and traffic flow of the sixteen sections. (The traffic flow data was normalized before)

Comparing the bottleneck indices of the 16 road links with their average traffic flow rate, we find the variation tendencies of their bottleneck indices and average traffic flow rate are opposite, as shown in Figure 5. The traffic flow rate is normalized before. The higher bottleneck index is companied with lower traffic flow, which is in accordance with practical situation.

Conclusions

In our research, we focus on approaches for indentifying bottlenecks based on urban planning data and traffic flow data. Compared with other method，the two-phase method measures the accordance between traffic demand and supply by quantitative analysis . We evaluate the road sustainability through the bottleneck index. The higher index means more serious problem. The bottlenecks location derived from the two-phase method were in accordance with the practical situation, which confirm the validity and feasibility of the two-phased method to identify the bottlenecks, especially in large-scale traffic network.

Acknowledge

The work described in the paper is partially supported by National Natural Science Foundation of China (NSFC) 5078054 and 60834001, Hi-Tech Research and Development Program of China (863Project) 2007AA11Z222, National Basic Research Program of China (973Project) 2006CB705506 and National Key Technology Re-search and Development Program 2006BAJ18B02.

References

American Highway Users Alliance. (1999). "Unclogging America's Arteries: Presriptions for Healthier Highways". *Availiable: http"//www.highways.org.*

American Highway Users Alliance. (2004). "Unclogging America's Arteries: Effective Relief for Highway Bottleneck". *Availiable: http"//www.highways.org.*

Banks, J.H. (1991). "The two-capacity phenomenon: some theoretical issues". *Transportation Research Record 1320*, 234-241.

Cassidy, M.J. (1995). "Methodology for assessing the dynamics of freeway traffic flow." *Transportation Research Record 1484*, 73-79.

Daganzo, C.F. (1997). "A continuum theory of traffic dynamics for freeways with special lanes," *Transportation Research B*, 83-102.

Qingdong Da and Guowu Zhang. (2002) "Quantitative methods of traffic congestion analysis," *Communication and Transportation systems engineering and information*, 45-48.

Haixian Jiao, Yingpeng Hu and Yizhou Liu. (2007) "A model and method for bottleneck identification and capacity integration equilibrium distribution in trunk road network," *Journal of Transportation Systems Engineering and Information Technology*, 63-67.

Lawson, T., Lovell,D., and Daganzo, C. (1997). "Using the input-output diagram to determine the spatial and temporal extents of a queue upstream of a bottleneck." *Transportation Research Record 1572*, 140-147.

Lin, W. and Daganzo, C. (1997). "A simple detection scheme for delay-inducing freeway incidents." *Transportation Research A*, 141-155.

Huapu Lu, Hong Yuan and Jianwei Wang. (2000) "The study on the traffic congestion countermeasures of Beijing", *Journal of Tsinghua University (Philosophy and Social Sciences)*, 87-92.

Fei Shi and Wei Wang. (2007) "Urban Road Network Structure Analysis", *City Planning Review*, 68-73.

Yizhou Wu. (2005) "Countermeasure study of urban traffic congestion on TDM," *Sci,& Tech, information of Water Transportation*, 78-80.

The Area Traffic Control System Evaluation Method Based on Travel Time Reliability

Zhiyong Guo[1] and Wei Wang[2]

[1] Transportation College, WuHan University of Technology, Wuhan, China, 430064; PH: (86) 027-86565741; Transportation College, Southeast University, Nanjing, China, 210096; Email: guozyy1972@163.com
[2] Transportation College, Southeast University, Nanjing, China, 210096; PH (86) 025-83794101; Email: wangwei@seu.edu.cn

Abstract: Travel time reliability is one of the parameters that evaluating the quality of traffic service in urban road network. The paper suggests a new method of area traffic control evaluation based on the travel time reliability. First, the concept of travel time reliability is defined and the travel time reliability model is built. Subsequently, the paper emphasizes on establishing the area traffic control evaluation model, which is based on the travel time reliability. The demonstration and utilization of evaluating the area traffic control system illustrates that the new method proposed in this paper is feasible and efficient.
Key Word: area traffic control, threshold, travel time reliability, evaluation

Introduction

Urban traffic control system is one of the most important component parts of modern urban traffic management system and also the component parts of intelligent transportation system (ITS). It can help improve traffic order, strengthen traffic safety, reduce traffic pollution and enhance capacity by controlling and directing traffic flow reasonably on urban roadway. Hence, it plays a significant role in decreasing traffic congestion, guaranteeing traffic unobstructed and keeping traffic flow in orders.

Along with the development of ITS technology, more and more real-time optimizing adaptive Area Traffic Control (ATC) are established and put into use. Accurate and objective evaluation of ATC performance is a vital part of construction. Now adopting travel time and delay as indicators to evaluate ATC performance

quantitatively is widely used. However, with the improvement in living quality and the aggravation in contradiction between traffic demand and traffic supply, travelers expect a travel with a smaller travel time variation to get destination timely, not just with a shorter travel time. The research (Bogers, 2004) indicates that, if travel time ranges with large deviation, travelers prefer to a steady travel with small time variation rather than an unpredictable "fast" travel, which is defined as travel time reliability.

Travel time reliability is defined to describe the probability that travel can be accomplished within a certain time range. It indicates the stability of travel time and the possibility of traveler finishing the trip within the expecting time. Travel time reliability is not only the evaluating indicator that travelers concern but also the evaluating indicator that used by traffic managing departments to judge the service standard. On the one hand, travelers hope to reach the destination within the expecting time range, reducing the unnecessary reserving time. On the other hand, traffic managers also expect to decrease travel time variation in order to provide more steady and reliable service and avoid to abate social welfares. Therefore, travel time reliability can be measured from both "quality" and "quantity" to evaluate the efficiency of traffic control system.

Asakura and Kashiwadani (1991) first introduce the travel time reliability concept in literature and believe that the variation of travel time can reflect roadway network reliability. Travel time reliability is clearly defined as the probability of getting destination in a specified time range (Iida, 1999). Bell (1999) develops a computing method for travel time reliability—estimating variation of travel time due to variation of traffic demand through sensitivity analysis. The SUE assignment model based on Logit is applied to estimate route traffic flow. Bhat (2006) uses revealed preference and stated preference data collected from a web-based commuter survey in Austin, Texas, to estimate a commute mode choice model. Lyman (2008) examines the uses of measured travel time reliability indices for improving real-time transportation management and traveler information with the use of archived intelligent transportation system data. Xiong (2004) analyzes the travel time reliability from the aspect of interaction between roads in networks. Chen (2007) studies the time distribution of travel time reliability among Beijing road network using the IC data collected from taxi. Liu (2009) uses travel time reliability for evaluating one-way traffic organization. Zhang (2009) demonstrates the distribution of travel time and travel time reliability of signalized trunk roads in daytime.

This paper adopts the method of travel time reliability, selecting travel time reliability as evaluation indicator, to establish the evaluation model for ATC system, and evaluates an ATC system of a certain city.

Model of Travel Time Reliability

Definition f Travel Time Reliability. Travel time reliability defined by Iida (1999) is as follows:

$$R = P\{x \le T\} \tag{1}$$

where R is travel time reliability; x is random variable of travel time; T is threshold of specific travel time; P is probability of that travel time x is less than or equal to threshold T.

There are two methods for determining threshold of specific travel time: one is due to travelers' acceptable travel time, and the other is due to traffic managers' acceptable travel time which does not take travelers' types into account. This paper adopts the latter method to determine threshold. Figure 1 shows the relationship among reliability, travel time and travel time threshold. The area of shadows is reliability. Considering the congestion effect, the threshold for traffic managers' acceptable travel time in this paper is under a certain Level of Service (LOS), which refers to α times of travel time T_f under the free flow condition. Then Formula (1) can be represented as:

$$R = P\{x \le \alpha T_f\} \tag{2}$$

The definition of Formula (1) and Formula (2) are not sufficient for two reasons (Zhang, 2009): first, travel time varies at different time during the 24 hours in a day. Second, there is also variation of travel time at certain time in differ day. To determine travel time reliability objectively, it is more meaningful to study travel time reliability at different time. Hence, the reliability should be defined dynamically as the probability of travelling from start to destination in a specific time range. Then the travel time reliability at time t can be represented as:

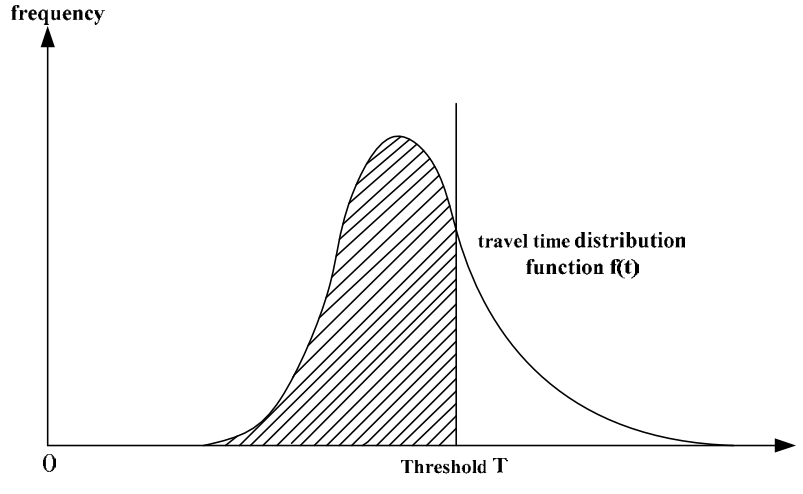


Figure 1. The relationship among reliability, threshold and travel time distribution.

$$R(t) = P\{x(t) \le \alpha T_f(t)\} \tag{3}$$

Where, $x(t)$ is random variable of travel time at time t; $T_f(t)$ is the travel time at time t under the free flow condition.

Route Travel Time Reliability. Assume that a specific route l consists of n independent sections, travel time of each section is known, and the sample is large enough. The travel time reliability for section i in route l at time t can be computed by following formula:

$$R^l(i,t) = P\{x_i^l(t) \le \alpha T_f^i(t)\} \tag{4}$$

Where, $R^l(i,t)$ is the travel time reliability of section i in route l at time t; $x_i^l(t)$ is travel time random variable of section i in route l at time t; $T_f^i(t)$ is travel time of section i under the free flow condition.

Since route consists of sections in series, according to the theory of series system, the route travel time reliability for specific route l at time t is as follows:

$$\begin{aligned} R(l,t) &= P(x_1^l(t) \le \alpha T_f^1(t), x_2^l(t) \le \alpha T_f^2(t), \ldots, x_n^l(t) \le \alpha T_f^n(t)) \\ &= \prod_{i=1}^{n} P(x_i^l(t) \le \alpha T_f^i(t)) = \prod_{i=1}^{n} R^l(i,t) \end{aligned} \tag{5}$$

Where, $R(l,t)$ is route travel time reliability of route l at time t; $R^l(i,t)$ is travel time reliability of section i on route l at time t; n is the total number of sections of route l.

Evaluation Model for ATC System Based on Travel Time Reliability

ATC system consists of intersection control, arterial control and district control. The coordination of several signalized intersection control forms arterial control subsystem (or subdistrict in ATC). The coordination of arterial control subsystems (or subdistricts) forms district control. Therefore, the evaluation procedure for ATC system is as follows:

(1) An arterial control subsystem can be regarded as a route. The route travel time reliability before and after operation of ATC system at time t can be calculated by using travel time data from floating car according to Formula (4) and Formula (5)

(2) The travel time reliability of the evaluation area can be evaluated by the method of weighted average. The traffic flow at time t is the weight. The formula is as follows:

$$R(t) = \frac{\sum_{l=1}^{m} Q(l,t) \bullet R(l,t)}{\sum_{l=1}^{m} Q(l,t)} \tag{6}$$

Where, $R(t)$ is the travel time reliability of ATC system at time t; $Q(l,t)$ is the traffic flow on route l at time t;

(3) The operating efficiency of ATC system can be calculated as follows:

$$E = \frac{R_h(t) - R_q(t)}{R_q(t)} \times 100\% \tag{7}$$

where $R_q(t)$ and $R_h(t)$ are the travel time reliability before and after operation of ATC system. They can be calculated according to Formula (6). E is the evaluation value of ATC system. The larger the E is, the more improvement the ATC brings in travel time reliability. Meanwhile travelers can reach destination timelier and ATC can provide more steady and efficient service.

Application Example

Data Collection. This paper selects 15 roads consist of trunk roads and subsidiary roads considering road distribution, road type, and traffic flow characteristic. Every route is a subsystem or consists of several subsystems. In all, there are about 140 signal controlled intersections, covering the whole ATC control area. The data was collected in April, 2009. The travel time and traffic flow data was collected for two periods of 7 days (7:00 to 19:00 every day) before and after operation of ATC.

Travel time data was collected by floating car as follows: GPS receiver equipped on floating car recorded date, time, position and speed every certain time interval. Then data was imported to computer and matched with electronic map to calculate travel time. Similarly, road average speed, delay data can be obtained by this method.

In terms of traffic flow assignment, the SUE model is widely used under the circumstance that the OD in evaluating area is known. According to user optimum principle, the traffic demand is distributed among routes, thus the traffic flow can be worked out. However, there two shortcomings prevent this method into practice: on one hand, it is difficult to obtain dynamic OD. On the other hand, it is impossible to solve SUE model for an entire urban road network. The traffic flow was got from loop installed on arterial roads. The traffic flow detecting data and corresponding signal timing data can be obtained from ATC control center. This method can get dynamic traffic flow data easily with high accuracy.

Calculation of Travel Time Reliability. There is no uniform criterion of travel time threshold in China. The research indicates that the travel time threshold can be 1.1~1.2 times of travel time under the free flow condition (Bell, 1997). The standard in Netherlands is: the travel time threshold is 1.2 times of average travel time for distance less than 50 kilometers and the target of this policy require 90 of 100 travels are reliable (AVV, 2004). To study reliability of route, the value can be 1~2 according to the detail level of research. Considering the influence of Level of Service (LOS), this paper takes the 1.2 times of free travel time at LOS B as threshold. The distance of every road section, exported from GIS system, was divided by free speed to get free travel time T_f of a certain road section. And 1.2 times of free travel time T_f is the travel time threshold.

Put the travel time and threshold at time t into Formula (4) to calculate travel time reliability of every section; then the above results is put into Formula (5) to calculate travel time reliability on route l at time t. Figure 2 shows the changes of travel time reliability of route 5 before and after operation of ATC system.

According to the daily traffic flow data detected by ATC system, there is clear morning peak (7：00~10：00) and evening peak (16：00~19：00) , while there is few variation during 10：00~19：00. To conveniently compare the travel time reliability before and after operation of ATC system, data was collect in three different time ranges: morning peak, flat hump time and evening peak. The average travel time reliability in three time ranges is shown in Table 1.

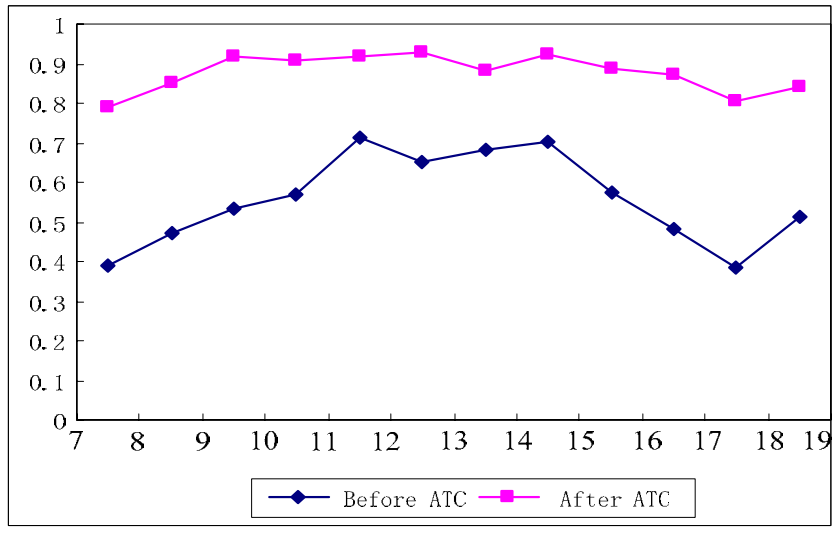


Figure2. The variation of travel time reliability before and after operation of ATC of route 5.

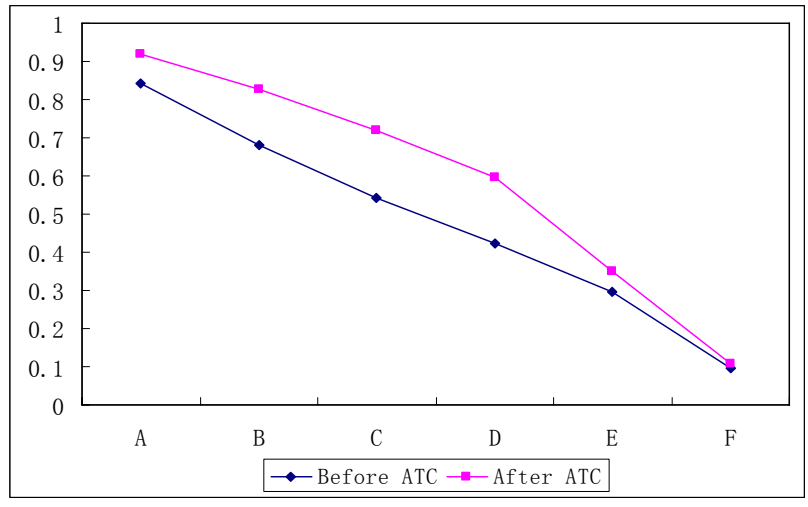


Figure3. The variation of travel time reliability at different LOS before and after operation of ATC.

Table 1. 15 route travel time reliability.

Route	Before ATC			After ATC		
	Morning peak	Flat hump time	Evening peak	Morning peak	Flat hump time	Evening peak
Route 1	0.351	0.292	0.286	0.786	0.728	0.721
Route 2	0.169	0.219	0.193	0.651	0.719	0.568
Route 3	0.475	0.641	0.441	0. 879	0. 905	0.851
Route 4	0.404	0.330	0.376	0.697	0.826	0.774
Route 5	0.525	0.718	0.516	0.852	0.898	0.863
Route 6	0.541	0.592	0.370	0.928	0.941	0.920
Route 7	0.324	0.450	0.405	0.656	0.768	0.766
Route 8	0.682	0.617	0.673	0.772	0.689	0.714
Route 9	0.607	0.651	0.629	0.850	0.863	0.832
Route 10	0.690	0.695	0.433	0.870	0.833	0.750
Route 11	0.815	0.841	0.833	0.914	0.941	0.909
Route 13	0.682	0.709	0.627	0.875	0.921	0.880
Route 13	0.465	0.620	0.437	0.844	0.857	0.850
Route 14	0.478	0.626	0.515	0.818	0.877	0.750
Route 15	0.625	0.765	0.620	0.873	0.882	0.879

Evaluation of ATC Operating Performance. Take the travel time reliability in Table 1 and corresponding traffic flow data detected by loop into Formula (6), the travel time reliability before and after operation of ATC can be calculated. Furthermore, the improvement in travel time reliability can be obtained by Formula (7), just as shown in Table 2.

Table 2. Improvement in travel time reliability of road network before and after operation of ATC.

Time	Morning peak	Flat hump time	Evening peak
Before ATC $R_q(t)$	0.519	0.578	0.485
After ATC $R_h(t)$	0.809	0.843	0.797
Improvement E (%)	53.87	45.80	63.29

Table 3. Route travel time standard deviation before and after operation of ATC.

Time	Morning peak	Flat hump time	Evening peak
Before ATC	0.166	0.363	0.159
After ATC	0.085	0.079	0.091

Table 4. The level of service for large and middle-sized cities.

LOS	A	B	C	D	E	F
Saturation (V/C)	＜0.4	0.4~0.6	0.6~0.75	0.75~0.9	0.9~1.0	＞1.0

Result Analysis. (1) Data in Table 2 indicates that the route travel time reliability is enormously improved after operation of ATC system during the three different time ranges and the operation has a high efficiency. The improvement value in morning and evening peak time is higher than that in flat hump time, which suggests that, in peak time, by coordinating signal timing of intersections, ATC system can assign network traffic flow in reason, mitigate traffic congestion, promote the traffic system operating efficiency and enhance the level of service.

(2) It can be seen from Figure 2 that route travel time reliability fluctuates along with the change of traffic flow at different time during 7:00~19:00. And it decreases at morning and evening peak with the increase of traffic saturation, which is lower than that in flat hump. After ATC system, the deviation is smaller than before. The decrease in deviation of route travel time reliability means travelers can get to destination more punctually, by shortening reserving time, to promote efficiency.

(3) According to the data in Table 1, the 15 routes' travel time reliability standard deviation can be calculated, shown in Table 3. It shows, after operation of ATC, the 15 routes' standard deviation of travel time reliability obviously decrease, which means dispersion degree of the 15 routes' travel time reliability abates and the whole operating performance of urban traffic system is improved.

(4)This paper also analyzes the reliability at different LOS, and LOS is divided to 6 levels according to traffic saturation. From Figure 3, along with the increase of traffic

saturation, traffic flow turn into free flow, steady flow and finally saturating flow and route travel time reliability decreases. At LOS A, travel time reliability do not change much before and after operation of ATC. With the decrease in LOS, the decreasing rate of travel time reliability after operation of ATC is less than that before it. The difference at same LOS expands from 0.05 to 0.17. Inflexion turns up at LOS D and travel time reliability decrease sharply even after operation of ATC. At LOS F, the travel time reliability is almost the same before and after operation of ATC and is close to 0.1. The data in Figure 3 also indicates that the travel time is more reliable than that before the operation of ATC at same LOS and reserving time decreases in steady flow (LOS B, C, and D).

Conclusions

Along with the improvement in living quality, travelers expect a travel with a smaller travel time variation to get destination timely, not just with a shorter travel time. Travel time reliability reflects the service quality of urban traffic system for travelers. This paper introduces the travel time reliability to evaluation for ATC operating performance. The travel time reliability is first defined, and it is calculated by data collected by floating car method. Then, the traffic flow collected by loop is used as weight and the evaluation model based on travel time reliability is established. The example of urban ATC system evaluation shows the maneuverability, sensitivity and applicability of this method. The collection method in this paper is quite costing, and cannot obtain dynamic travel time. Hence, the data collection method needs to be improved. In addition, it is also necessary to have a further research in travel time distribution function, the inter-connection between variation of travel time reliability and the traffic flow.

Acknowledgements

This research was supported by the Major Program of the National Natural Science Foundation of China (Grant No.50738001) and National the Key Basic Research Program (973 Program) (Grant No.2006CB705505).This support is gratefully acknowledged.

References

Asakura Y, M. (1991). “Kashiwadani. Road network reliability caused by daily fluctuation of traffic flow.” *The 19th PTRC Summer Annual Meeting*, Proceedings Seminar G, 73-84.

AVV Transport Research Centren (2004). Dutch Ministry of Transport, Public Works

and Water Management. Nota Mobility [EB/OL]. http: // www.notamobiliteit.nl.

Bell, M. G. H, Shield, C. M., and Busch, F. K. (1997). "A stochastic user equilibrium path flow estimator." *Transportation Research*, 5 C, 197-210.

Bell, M.G.H, Classir C, Iida, Y., and Lam, W. H. K. (1999). "A sensitivity-based approach to network reliability assessment." *Proceedings of the 14th International Symposium on Transportation and Traffic Theory (ISTTT)*, Jerusalem Israel, 283-300.

Bhat CR, Sardesai R. The impact of stop-making and travel time reliability on commute mode choice. *Transportation Research Part B,* 2006, 40(9), 709-730.

Bogers E A I., and Zuylen van H J. (2004). "The Importance of reliability in route choice in freight transport for various actors on various levels." [C/CD], *Proceedings of the European Transport Conference. Strasbourg*: [s.n.].

Chen K, L Yu, J. F. Guo, and H.M.Wen. (2007). "Characteristics analysis of road network reliability in Beijing based on the data logs from taxis."[EB/CD], *The 79th Transportation Research Board Annual Meeting*, Washington D.C.

Iida, Y. (1999). "Basic concepts and future directions of road network reliability analysis." *Joumal of Advanced Transportation*, 33(2), 125-34.

Liu Wei, and Zhang Jian-xu (2009). "Application of travel time reliability in one-way traffic organization." *Journal of Highway and Transportation Research and Development*, 26(2), 120-123.

Lyman Kate, and Bertini Robert L. (2008). "Sing travel time reliability measures to improve regional transportation planning and operations." *Transportation Research Record,* n2046, 1-10.

Xiong Zhi-hua, Yao Zhi-sheng, and Shao Chun-fu. (2004). "Travel time reliability in road network associated with road section." *China Safety Science Journal*, 14(10), 81-84.

Zhang Yong, Wang Shiming, and Yang Xiao-guang. (2009). "Computing and demonstration of travel time reliability on signalized Road." *Journal of Tongji University (Natural Science)*, 37(6), 772-776.

Evaluation on Vehicle Restriction Measure in Beijing

Shuling Wang[1,2], Chunfu Shao[1], Fang Wang[2] and Jianping Sun[2]

[1] MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University. Beijing 100044, China
[2] Beijing Transportation Research Center. No. 9 Beibinhe Road, Xuanwu District, Beijing 100055, China; E-mail: wangsl.bj@gmail.com

Abstract

Aimed at the hot topic "vehicle restriction in Beijing," this paper set up a systematic monitoring and evaluation system. Firstly, this paper reviewed the experience and effectiveness of the implementation of TDM in the world, particularly Mexico City's implementation of a 10-year-old motor vehicle restriction measure. On this basis, according to Beijing's vehicle restriction multiple effects, this paper established a set of comprehensive evaluation systems including traffic, environmental, and social impact. Through the methods of floating car systems, congestion evaluation indices, public transportation IC cards, public opinion surveys, and web-based surveys, systematic data were analyzed. The data shows that the traffic and air quality has improved, especially the traffic congestion The public strongly supports all improvement within the traffic system. This paper will provide an important reference for decision-making and other cities' traffic management research and implementation.
Key Words: TDM; evaluation; vehicle restriction; congestion index; policy suggestion

Introduction

Transportation Demand Management (TDM) aims to maximize the efficiency of the urban transport system by discouraging unnecessary private vehicle usage and promoting more effective, healthy, and environmentally friendly modes of transport such as subway, buses, bicycle, and walking. TDM can provide many benefits, not only alleviating congestion and pollution emissions, but also parking cost savings and safety improvement. When all impacts (benefits and costs) are considered, TDM is often the most cost effective overall transportation improvement strategy.

There are two basic types of measures that can achieve the goal of improving transport system efficiency: "Pull" measures that make other modes of transport more attractive, and "Push" measures that make private vehicles less attractive. Economic measures can affect driver behavior and reduce car usage. Such measures include fuel taxes, road tolls, parking fees, and low emission zone. Many countries and cities have adopted these measures to reduce car use (Wagner 2008; Breithaupt 2008; China Daily 2005; Transfer for London 2007; http://www.lowemission zones.eu/content/view/45/61,http://www.comune.milano.it/dseserver/ecopass/index.html).

Another example of a TDM policy that can be used in certain situations is vehicle license number restrictions. This is a policy where vehicles are restricted from driving in an area on certain days of the week, based on registration plates, with the aim of reducing the overall number of cars in use. International experience includes many cities, which have implemented such measures with varying degrees of effectiveness. In most cases, the plate restriction applies to certain types of vehicles, certain zones, or certain times of day; in some cities, it may apply for the full day.

Example of Restricting Car Use with Plate Restrictions

Bogota uses a scheme in which 40% of private vehicles cannot operate in the city from 7:00--9:00 and from 17:30–19:30 in accordance with designated number plates.

Santiago de Chile uses a scheme, which is in effect only on days where atmospheric pollution reaches emergency levels. All vehicles except buses, taxis, and emergency vehicles are prohibited from circulation in morning and evening peaks on the six principal roads connecting the outer areas and city center.

Sao Paulo uses a scheme over a wide central area (within the Inner Ring, 15 km diameter) in which 20% of vehicles (1 and 2 on Mondays etc.) are prohibited from 7:00--8:00 and 17:00--20:00 for weekdays.

Manila uses a scheme, which prohibits certain vehicles, identified by number plates, from operating on the main traffic arteries during peak periods.

Singapore has an off-peak car plate restriction scheme, which complements its transportation policies. These are normal cars with red number plates that are allowed to be on the road only during the night-time hours and weekends, i.e., 18:00--17:00 on weekdays, after 15:00 on Saturdays, and whole day on Sundays and public holidays. The scheme offers drivers the option to save on car registration and road taxes in return for reduced usage of the car. In 2005, off-peak cars accounted for about 2% of the total car population (Briaddys et al. 2009).

Mexico City introduced a program, *Hoy No Circular*, that bans most drivers from using their vehicles one weekday per week based on the last digit of the vehicle's license plate. For example, vehicles with a license plate ending in 5 and 6 may not be

used on Mondays. The restrictions are in place during weekdays between 5:00 a.m. and 10:00 p.m. and affect the vast majority of residential and commercial vehicles. Taxis, buses, police cars, ambulances, fire trucks, commercial vehicles operating with liquid propane gas, and commercial vehicles transporting perishable goods are exempt. The restrictions apply to the entire Mexico City metropolitan area, hereafter "Mexico City," which includes Mexico City and municipalities in neighboring states.

Researchers conducted a test regarding driving restrictions:
(1) Across pollutants and specifications, there is no evidence that the restrictions have improved air quality. In the contrary, the air pollution in the weekend has more polluted than before.
(2) Evidence from additional sources indicates that the restrictions led to an increase in the total number of vehicles in circulation as well as a change in composition toward high-emissions vehicles.
(a)Motor Vehicles: the motor vehicle count increased 325,000 during the year of 1989-2005.
(b) Oil: From 1980 to 2007, the sales of crude oil has not diminished but rather increased. The average sale of all types of gasoline annual growth rate was 1.7% (Davis 2008).

Vehicle Restriction Measure in Beijing

Basic Information of Beijing. With the population and motor vehicle ownership increasing rapidly, the traffic congestion became worse over time and the congestion index continued to increase compared to the previous year. By the end of 2008, Beijing resident population was 16.95 million people. Between 2003 and 2008, the average annual increase of automobiles in Beijing was 11.0% of which: average annual growth rate of private cars is as high as 20%, which means 37.4 million private cars increased in 2008. By the end of 2008, the city's vehicle use reached 3.504 million (Beijing transportation research center 2009).

Beijing government departments have adopted a series of practical and effective management measures to guide and regulate the public to use the health, environmental, and sustainable transport tools instead of cars. In particular during the Beijing Olympic Games, motor vehicles were prohibited from use according to registered license plate number. The public generally felt the smooth flow of traffic and the fresh air. Public opinion poll during the Olympic Games showed 76% of residents suggested that the motor vehicle use should be restricted after the Olympic Games (Wang et al. 2008).

Vehicle Restriction Measure. In response to the traffic and environmental pressures after Olympic Games, Beijing Municipal Government issued "Beijing Municipal Government notice on the implementation of traffic management measures," from October 11, 2008, to April 10, 2009, Beijing motor vehicle use is restricted one weekday per week according to the last digit of the vehicle's license plate. Group No.1 (vehicle license plate ending in 1 and 6) could not be used on Mondays; group

No. 2 (vehicle license plate ending in 2 and 7) could not be used on Tuesdays; group No. 3 (vehicle license plate ending in 3 and 8) could not be used on Wednesdays; group No. 4(vehicle license plate ending in 4 and 9) could not be used on Thursdays; and group No. 5 (vehicle license plate ending in 5 and 0) could not be used on Fridays. The restriction weekday will change once a mouth. For example, in the second mouth, group No. 1 will be restricted on Tuesdays.

The restrictions time is 6:00 a.m. to 9:00 p.m. and affect the vast majority of residential and commercial vehicles. Taxis, buses, police cars, ambulances, fire trucks, and other vehicles with special permitted licenses are exempt. The restrictions apply to the fifth ring road metropolitan area (http://www.bjjtgl.gov.cn/publish/portal0 2008) (Figure 1).

Table 1. Public activities to encourage green traffic.

Time	Activities(Environmental Defense Fund 2008)
May 16, 2006	Public service "car free day monthly for more blue sky"
August 11, 2007	"Drive less, Enjoy Olympic Games"
September 22, 2007	"China Urban Public Transport Week and Car Free Day"
June 23, 2008	"Green Olympics, Green Travel"
July 20--September 20, 2008	Half vehicle driving restricted
August 22, 2008	"riding in Beijing, a low-carbon action"

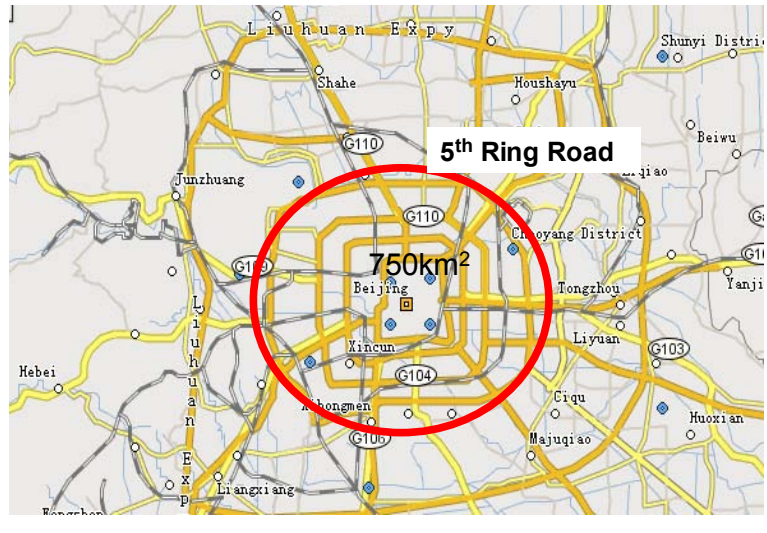


Figure1. Vehicle restriction areas.

Evaluation Systems

It is estimated that about 9 million car trips are made in Beijing every day. The vehicle restriction measures are bound to influence these 9 million passenger trips. A positive effect may include reduction of moving vehicles and pollutant emission,

improvement of running speed and air quality, and many other beneficial aspects. This measure will inevitably cause inconvenience to car users.

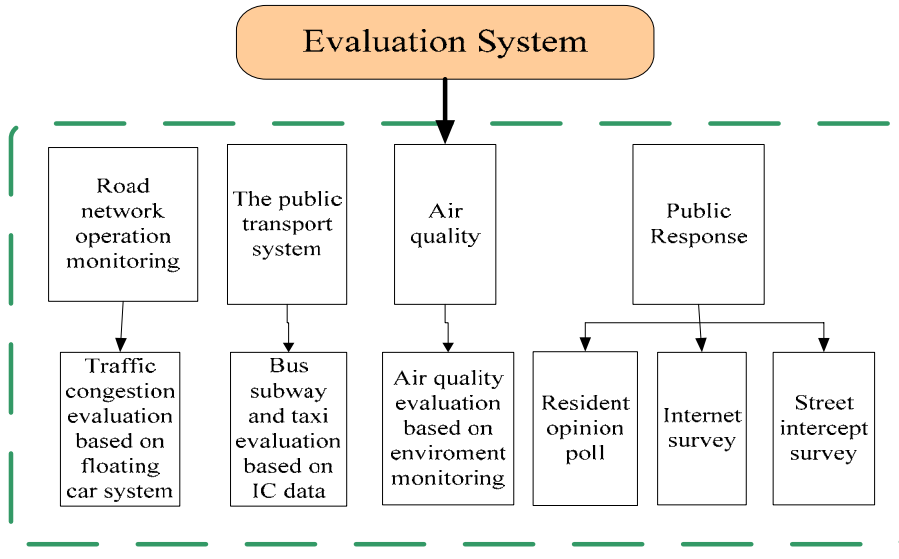


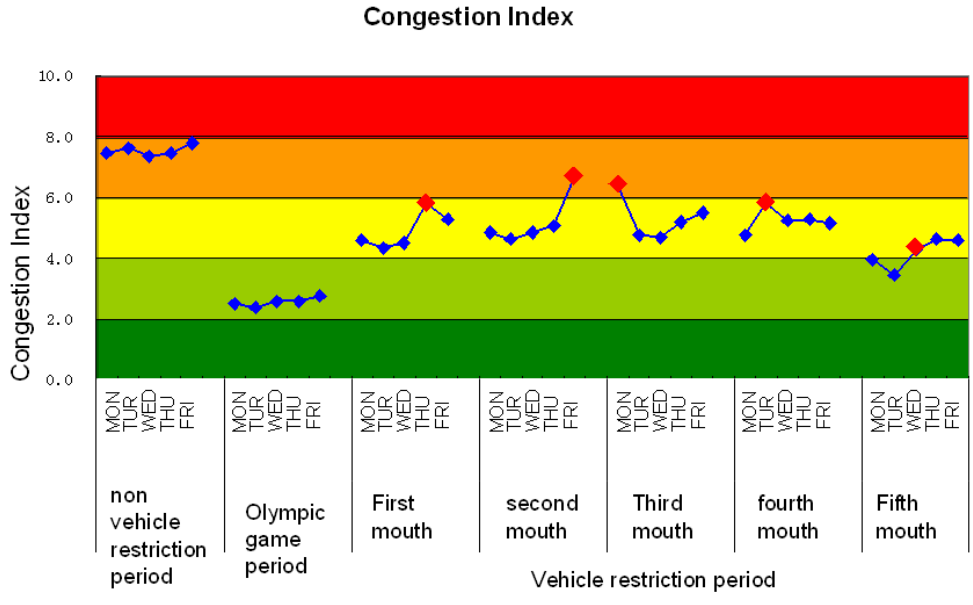
Figure 2. Evaluation systems fragment.

The evaluation of vehicle restriction measures requires a comprehensive index system. Based on the scope and impact of the measure, the evaluation index system includes the road network operation monitoring, public transportation system monitoring, and air quality monitoring and public reaction.

Monitored Data

Road Network Operation Monitoring. In this paper, the traffic congestion evaluation index system, which is used as the basis for performance analysis, is used. It consists of comprehensive indices and characteristic indices, that reflect the degree of congestion, congestion duration, congestion area (congested road network, road length, percentage), and congestion frequency. According to congestion levels and feelings of the drivers, congestion index is divided into five levels: very smooth (congestion index between 0 and 2), smooth (congestion index between 2 and 4), mild congestion (congestion index between 4 and 6), moderate congestion (congestion index between 6 and 8), and severe congestion (congestion index between 8 and 10) (Wen et al., 2008).

October and November of 2007 were selected as “non vehicle restriction measure period” of baseline data. Based Floating Car System, which data can cover 90% of road network within the fifth ring road and the data can update with an interval of 60-90 seconds, and congestion evaluation system, traffic congestion index, congestion areas (the road sections appear “serious congestion” at least four days of each week. The morning peak hour and evening peak hour are accounted separately.), congestion duration time and the road network speed were chosen to evaluate measure implementation.



Colors					
Congestion level	Very smooth	smooth	Mild congestion	Moderate congestion	Serious congestion

Figure 3. Congestion index.

Congestion Index. In the "non vehicle restriction measure period," the traffic congestion index was 7.54, road network was running at a "moderate congestion" level. During vehicle restriction period, the traffic congestion index decreased to 5.15, at the "mild congestion" level. As the vehicle with license plate tail number of 4 is relatively less than other number vehicles, the data show that the tail number 4 and 9 of the motor vehicles prohibited day, its traffic congestion index is maximum value of the month.

Congested Areas. The congestion areas decreased from 99 km (422 road sections) to 50 km (249 road sections), the decreasing ratio reached 49.4% (40.8%) during the morning peak hour. In evening peak hour, the congestion areas decreased from 177 km (818 road sections) to 110 km (555 road sections) with the ratio of 37.5% (32.1%).

In the morning peak hour, the congestion areas concentrated on the second ring road and Zhongguancun area, the traffic flow is toward the center of the city. In the evening peak hour, congestion occurred within the Second Ring road, CBD, Zhongguancun and the West Third Ring Road areas.

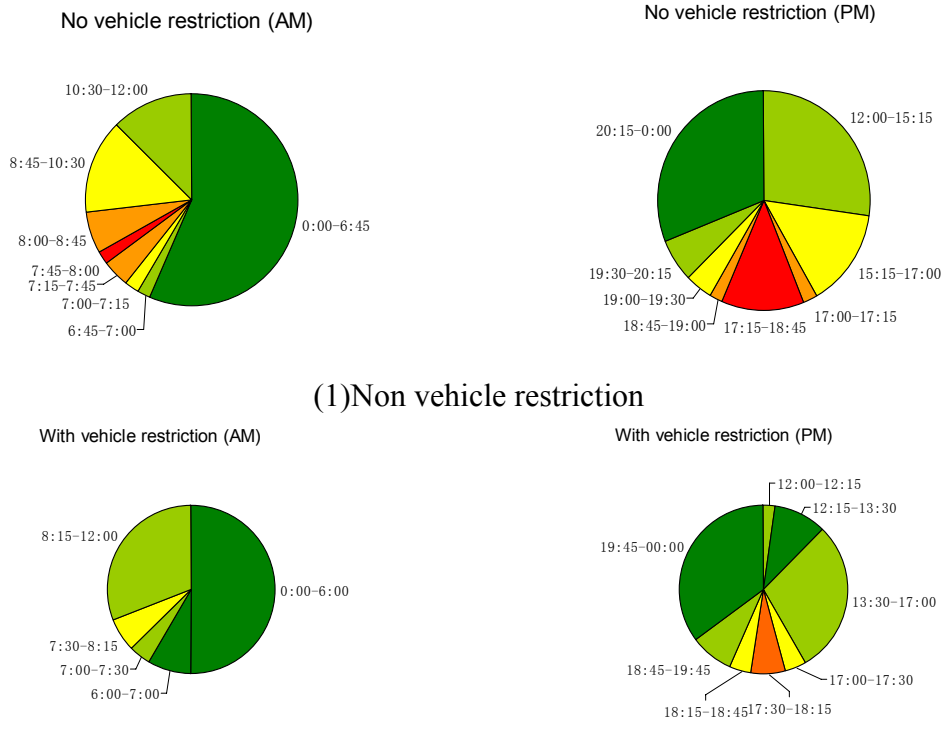
Congestion Duration Time. Compared with "non vehicle restriction measure period," the congestion duration time (including severe congestion, moderate congestion, and mild congestion) reduced from 7 hours and 45 minutes to 2 hours and 30 minutes. Among them, "severe congestion" time was reduced by 1 hour and

45 minutes; “moderate congestion” time was reduced by 1 hour; “mild congestion” time was reduced by 3 hours 30 minutes.



(1) non vehicle restriction (2) with vehicle restriction

Figure 4. Congested areas distribution.



(1)Non vehicle restriction

(2)with vehicle restriction

colors					
Congestion level	Very smooth	smooth	Mild congestion	Moderate congestion	Serious congestion

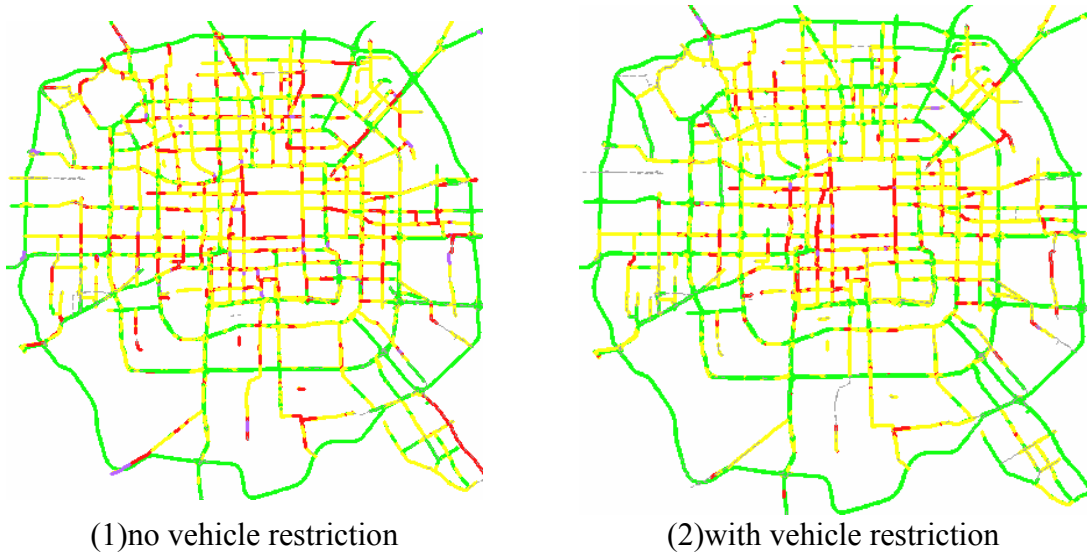
Figure 5. Congestion duration time distribution.

Meanwhile, the morning peak hour congestion beginning time lagged from 7:00 a.m. to 7:30 a.m., the evening peak hour beginning time lagged from 3:15 p.m. to 5:00 p.m. It indicated that citizens adjusted their trip time. For the more smooth traffic than before, they would spend less time in traffic journey, and so they could leave home a little later.

Road Network Speed. Compared with "non vehicle restriction period," the vehicle running speed has increased obviously. The running speed increased by 3.4 km/h (15.6%) in morning peak hour and 2.6 km/h (13.8%) in evening peak hour. Among them, the expressway running speed changed significantly, the morning peak hour speed increased 6.6 km/h (21.3%), the evening peak hour speed increased 5.8 km/h (23.0%).

Public Transport System. The monthly average daily passenger volume increased 2.87 million passengers, increased 20.8% compared with "non vehicle restriction period." About 60% of the drivers chose to take public transport when their cars were limited. It is generally agreed that urban public transport has improved, but improvement is still needed.

Air Quality. According to the municipal environmental protection department's data, the air quality is improving in the overall trend. Daily motor vehicle emissions reduced by 375 tons, motor vehicle emissions of hydrocarbons, carbon monoxide, nitrogen oxides and reparable particulate matter were separately reduced by 10%, 10%, 8%, and 8%, respectively.



(1)no vehicle restriction (2)with vehicle restriction

colors				
speed	N/A	0-20 km/h	20-40 km/h	>40 km/h

Figure 6. Speed contrast of morning peak hour.

Public Response

Degree of Support. Vehicle restriction measure has relationship with the close interests of all the residents; the public's attitude toward the measure also has a direct impact on future decision-making. Aimed at this measure, a poll was carried out with 3500 samples, which included 2000 car users (drivers) and 1500 people without vehicles (public passengers). The survey showed 89.3% of the respondents support vehicle restriction, 4.1% of the respondents were neutral, and 6.6% of the respondents did not support the measure. It is indicated that vehicle restriction measures generally received support from the public.

Evaluation of Vehicle Restriction. In fact, the majority of Beijing residents felt the benefit of vehicle restriction measures on the improvement of traffic and air conditions. Surveys showed that 80.2% of the respondents agreed the improvement of traffic operation and 73.9% of the respondents believe that the air quality has improved since the implementation of vehicle restriction measures.

Suggestion for Future Traffic Measures. Survey data showed that 86.6% of respondents support vehicle restriction measures in the future, and 10.1% of respondents disagreed to implement this measure in a long-term. The supporting ratio of car owners is obviously lower than the ratio of non-car people. Thus when considering relevant policies in the future, the interest of car owners, in particular, should be take into account.

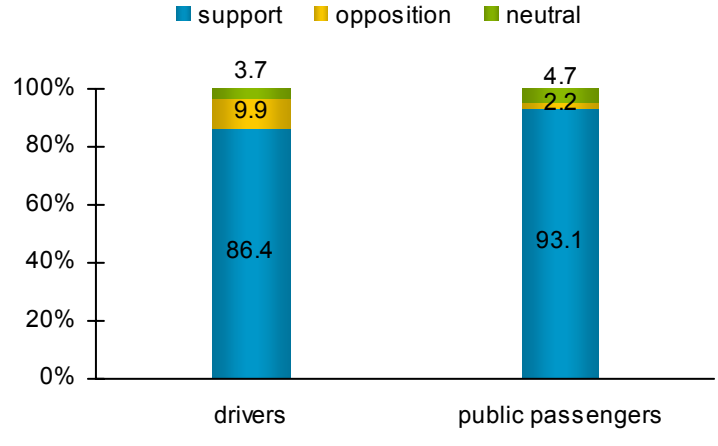


Figure 7. Attitudes to vehicle restriction of different group.

Internet Opinion. Local and international media pay much attention to Beijing vehicle restriction measures. “Newsweek” magazine, published in the United States, published an article expressing appreciation for the Beijing vehicles management measure in October 20, 2008. People’s Daily, China Youth Daily, Legal Daily, Xinhua net, People net, and Sina net all comment on the media articles to participate vehicle restriction discussion. Sina net survey showed the public supporting rate was

31.9% to this vehicle management measure (http://survey.news.sina.com.cn/voteresult.php?pid=27447 2008). The hot topic mainly focused on the measure's legal basis, the rotation of limited vehicles, and limited scope (whether it contains the fifth ring road), etc.

Conclusions

After the implementation of the vehicle restriction measure, both the traffic congestion and the air quality improved, public transport passenger has grown stably, taxis operation has increased, and people have shown their concern and support. Vehicle restriction measures pay an important role to alleviate urban traffic congestion.

(1) Alleviated congestion. By the end of February 2009, the vehicle number had increased to 454,000 vehicles compared to the number at the end of November 2007. The peak hour speed increased 14.7% with the implementation of a vehicle restriction measure.

(2) Improvement in air quality. With the implementation of a vehicle restriction measure, the number of days reaching air quality standards are better than the ever.

(3).Increase in public transport passengers. There is steady increase in passenger volumes and taxi operational efficiency. The passenger volumes increased 2.87 million per day, up 20.8%, and taxi delivery capacity increased by about 6%.

(4) People concern and support. The poll data show that 86.4% of a car owners support vehicle restriction measures, non-car people show a higher support rate of 93.1% and 81.9% of car owners support the long-term implementation of the vehicle using control measure.

(5) Faced with the higher speed of vehicle number increase (increasing more than 1,600 cars per day), it is proposed to take both vehicle use control measures and "Public Transport Priority" strategy, to promote and create conditions for residents to use green travel tools, thus achieving the goal of smooth traffic and clean air quality.

References

Beijing Municipal Government notice on the implementation of traffic management measure. (2008). http://www.bjjtgl.gov.cn/publish/portal0/

Beijing Transportation Research Center. (2009). "*Beijing Transport Annual Report 2009*." *Beijing.*

Breithaupt, M. (2008). "Environmental vehicle taxation: international experiences." *Presented on the International Workshop on Integrated Transport for Sustainable Urban Development in China*, 15-17, December 2008.

Broaddus, A., Litman, T., and Menon, G. (2008). "*Transport demand management training document*." Deutsche Gesellschaft fürTechnische Zusammenarbeit (GTZ) GmbH P. O. Box 518065726 Eschborn, Germany, http://www.gtz.de. http://www.sutp.org

China Daily, 30 December 2005.

Davis, L. W. (2008). "*The effect of driving restrictions on air quality in Mexico City*." University of Michigan, Journal of Political Economy.

Environmental Defense Fund. (2008). "Green trip report of Beijing 2008." Beijing.

http://www.comune.milano.it/dseserver/ecopass/index.html

http://www.lowemissionzones.eu/content/view/45/61

Sina net, (2009). http://survey.news.sina.com.cn/voteresult.php?pid=27447

Transfer for London. (2007, 2008). "*Annual Report and Statement of Accounts.*" London.

Wagner, A. (2008). "*Fuel taxation as an economic instrument to tackle climate change*" presentation Bangkok http://www.gtz.de/de/dokumente/gtz2008-en-fuel-taxation.pdf

Wang, S., Chen, J., Guo, J., and Li, C. (2008). "Application and evaluation of traffic demand management policies during Beijing Olympic Games." *Journal of Transportation Systems Engineering and Information Technology.*

Wen, H., Hu, Z., Guo, J., Zhu, L., and Sun, J. (2008). "Operational analysis on Beijing road network during the Olympic games." *Journal of Transportation Systems Engineering and Information Technology*

Traffic Flow Characteristics of Urban Expressway in the Period of Ice and Snow of Cold Areas

Enhui Xing[1] and Rui Wang[2]

[1] Associate Professor, School of Mechanical Engineering, Jiamusi University, 148 Xuefu Street, Jiamusi, China 154007; Email: xingeh@126.com
[2] Associate Professor, School of Art, Jiamusi University, 148 Xuefu Street, Jiamusi, China; Email: wangrui_jmsu@126.com

Abstract

Urban expressway is a rapid arterial that connects main areas of the city, and meets the great-amount, long-distance, and high-speed demand of urban traffic. The friction coefficient of roads covered with snow and ice is reduced, and the traffic operation risk is increased in regions with high latitudes, whose winter is long and cold, and those conditions led to the reduction of road capacity and the increase of traffic delays. Using the video detection device, one of the traffic flow information collecting devices, a large number of data in the period of ice and snow of cold area are collected, which is the research base of this paper. The macro and micro traffic flow characteristic of urban expressway under conditions of the snow and ice environment are analyzed, such as the instant traffic flow, the distribution of headway time, and the properties of the speed distribution. The relation models of speed-volume are built. At the same time, the traffic flow characteristic of the urban expressway in the period of ice and snow of cold areas are compared with that of non-snow periods. The outcomes studied above are the foundation for analyzing traffic operation and capacity of urban expressway in the period of ice and snow.
Key Words: traffic flow characteristics; headway time distribution; snow and ice environment; urban expressway

Introduction

With the consistent and swift economic development of China, private passenger cars have become essential forms of transportation in many families. Traffic demand is increasing so rapidly that serious traffic congestions have appeared in some developed cities. In order to enhance road traffic capacity and lighten urban traffic burden, many large cities such as Beijing, Shanghai, and Guangzhou in China have built their expressway systems one after another. The urban expressway is a rapid arterial that connects main areas of the city and meets the long-distance and high-

speed demand of urban traffic. It plays an important role in resolving the increasing urban traffic pressure. The majority of the northeast and the northwest areas in our country experience snowy winters, with this climate often producing snow for 3~4 months. The visibility is reduced in these regions, which are influenced by the snow, snow cover, and the ice, sometimes even blocking the traffic and causing traffic accidents. The urban expressway, as the main undertaker in the city's long distance and rapid traffic, is vulnerable to this weather. Therefore, it's very necessary to study systematically on traffic flow and capacity of urban expressway under the snowy conditions.

Literature Review

The influence of snow to the road traffic flow mainly embodies two aspects: 1) the snow reduces visibility and impacts driving sight, and 2) the snow on the road reduces the road adherence coefficient and influences the vehicle's travel stability. The research of traffic flow characteristics is always one of the most basic and important contents in traffic engineering, and it is also the main basis of the road traffic management countermeasure formulation and the effect appraisal. In recent years, domestic and foreign researchers have studied the traffic flow characteristics under the condition of the snow-ice climate. Research on traffic flow models is still an important problem in traffic engineering fields. Traffic flow models based on traditional traffic flow theory such as Greenshields linear model, Underwood exponential model, Drew model, and May model are continual traffic flow models of two-dimensions. But the traditional research of traffic flow characteristics is based on the result under the condition of the normal weather, and there is no obvious effect under the heavy weather condition. The traffic flow volume would be reduced 7 %-56 % because of the snow in the winter (Hanbali and Kuemmel, 1993). The reduced rate of the road traffic flow volume is proportional to the amount of snow and wind (Keith and Leland, 2000). Research shows that the saturation flow rates reduce by 20%, the average speeds reduce by 30%, and starting loss increases by 23% at signal intersection in inclement weather (Perrin and Hansen, 2001). The influence of snow climate to the urban road traffic flow velocity and capacity and the running characteristics of road intersection have been analyzed (Ren, 2008). This paper examines the traffic flow characteristics of the urban expressway during icy and snowy weather conditions.

The Influence of Snow on Road Safety

When driving on a snowy road, the tires slip on the road, making the temperature rise, thus changing the property of the snowy road. The snow is compacted and becomes smooth, producing a film of water due to the tire's friction heat melting the snow or ice. The adhesion coefficient is lower than that of the dry roads, increasing the probability of collision or scraping in case of emergency braking. The braking distance will extend greatly, 2-4 times than on a dry road. The driver will increase the headway distance between the front and the back vehicles, usually 2-3 times compared to a dry road.

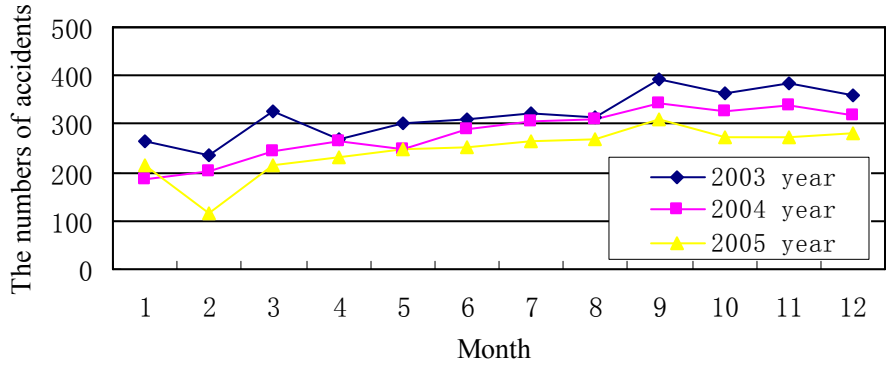


Figure 1. Accident distributions according to month accident numbers.

Figure 1 shows the number of accidents in the Heilongjiang province of China during 2003-2005. From the figure, we can see that in the final three months, the number of road accidents is higher than any other months (except in September) mainly because of the snowy roads. The numbers of accidents from January to March are less than that of other months, due to the bad road condition, significantly reduced vehicle travel. Therefore, the snow climate of road traffic safety is very significant.

Data Collection of Traffic Flow

Choosing Hexing road, one section of the 2nd Urban Ring Road of Harbin(u_f=70km/h) , as the observation road section, the digital camera to record short-term traffic condition is mounted on a pedestrian footbridge. Figure 2 shows the roadway condition on the investigated section.

The original data are collected by the American video detecting instrument called VT Edge2 video. The equipment can detect traffic volume and speed directly. The basic data of actual measurement flow include the type of vehicle, the composition of traffic, volume, speed, headway, and density/occupancy.

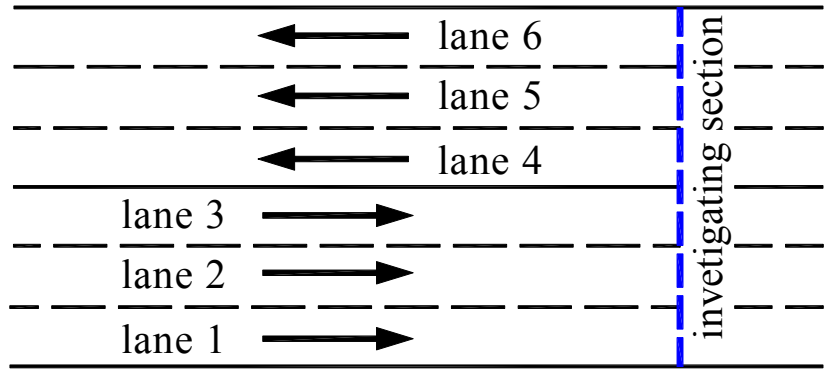


Figure 2. Investigating section.

Table 1. Change of traffic flow parameters.

	Snow	Decrease percentage of capacity (%)	Decrease percentage of speeds (%)
HCM2000	<=1.27mm/h	4%	4%
	1.524-2.54mm/h	9%	8%
	2.55-12.7mm/h	11%	9%
	>12.7mm/h	22%	13%
Paper	Snowy road		20%
	Icy road		30%

Impact Analysis of Snow-icy Road on the Speeds of Flow

The direct impaction of snow-icy roads on traffic flow is speed of traffic flow. The speed is the most important data to show the state of traffic flow, analysis of vehicles running characteristic, manage, or control the traffic. HCM 2000 (Highway Capacity Manual 2000) listed the average influence of snow to the highway capacity and speeds, as shown in Table 1.

Icy and snowy roads have the most influences on the flow speed, and the flow speed is the most important data to show the state of traffic flow, analysis of vehicles running characteristic, and manage or control the transportation. HCM 2000 listed the average influence of the snow to the highway about capacity and speed, as is shown in Table 1.

From Table 1, the comparison result under the snow-icy road between HCM 2000 and this paper researches show that the influence of snow-icy road to free flow speed on Hexing road is more serious than that in HCM 2000.

Impact Analysis of Snow-icy Road on the Time Headway

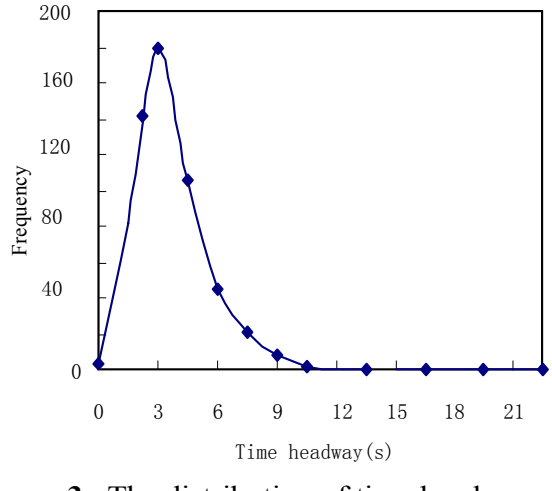


Figure 3. The distribution of time headway

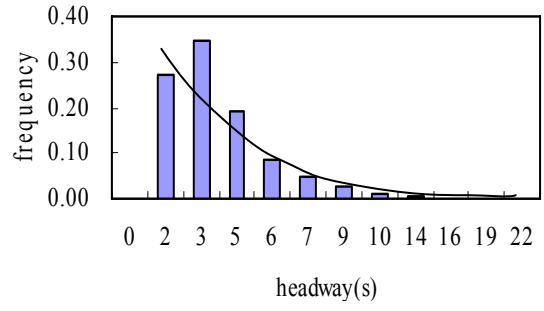


Figure 4. The frequency of headway.

The time headway means that the time interval between the passing of the front ends of two consecutive vehicles or trains moving along the same lane or track in the same direction. Headway is a key input in calculating the overall capacity of any transit system. Figure 3 shows the distribution of the time headway on Hexing road.

Through the analysis of traffic data, we can see that the frequency of time headway under the snow-ice condition is close to the negative exponential distribution. Figure 4 shows the frequency of time headway. As the impact of snow-icy road conditions, the minimum time headway becomes larger.

The Relationship Analysis of Speed and Volume Analysis

Traffic volume Q, traffic flow density K, and the driving speed are three basic parameters to describe traffic flow characteristics. Traffic volume equals traffic flow density k multiplied by driving speed. During this study of traffic flow characteristics, both the speed and the volume can be obtained directly from the actual observation. They are used to measure road traffic conditions.

The relationship curve between parameters can be obtained by regression analysis according to statistical principle. Using 5 minutes as a statistical interval, the traffic flow data such as flow volume and speeds are collected under the non-snow-icy and the snow-icy road surface at the same site. Figure 5 shows the scatter plots of flow-speed on the snow-icy road surface on Hexing road, and Figure 6 shows that on the non-snow-icy road surface.

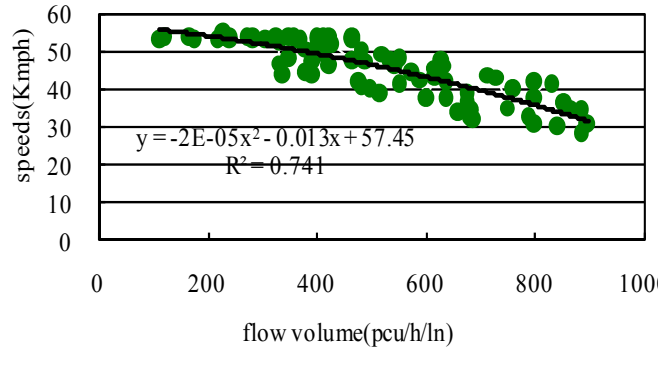


Figure 5. Speed-flow volume scatter plots on the snow-icy road.

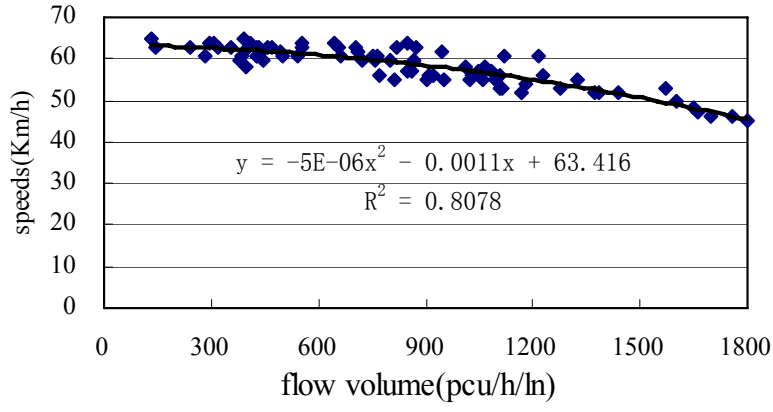


Figure 6. Speed-flow volume scatter plots on the non-snow-icy road.

Table 2. Fitting model of speed-volume.

	Fitting formula	R^2
Snow-icy road	$V= -2\times10^{-5}q^2 -0.0135q + 57.459$	0.74
Non-snow-icy road	$V= -5\times10^{-6}q^2 - 0.0011q + 63.416$	0.81

Through the comparison of Figure 5 and Figure 6, we can see that:

- On the snow-icy road, when the flow volume is less than 350 pcuphpln, the speed basically stays at 53 km/h, and then the car runs at the free flow speed. When the volume is less than 600 pcuphpln, the speed stays at the steady data; the free flow speed will decrease to 10 km/h on the snow road than that on the non-snow-icy road.
- On the snow-icy road, with the increase of flow volume, the flow speeds decelerated faster than that on the non-snow-icy road. When the volume is increasing to 900 pcuphpln, the speed's decrease of driving on the snow-icy road is nearly 20 kmph. On the non-snow-icy road, the speed appears scattered, and the speed's decrease is only slight.

The relationships between speed and volume are analyzed according to data regression analysis. Taking the second-conics fitting, the relationship models of speed and volume are shown in Table 2.

Conclusions

Conclusions from the analysis above are as follows:

- The speeds of urban expressway traffic flow reduce by 20% on the snow-icy road than that on the non-snow-icy road.
- The speed of free flow reduces 10 km/h on the snow-icy road than that on the non-snow-icy road.
- With the volume increasing, the speed on the snow-icy road decreases faster than that on the non-snow-icy road.

Acknowledgment

This work has been supported by the National Natural Science Foundation of China (Grant No. 50908101), and we acknowledge their fund warm support and detailed helps.

References

Hanbali, R. M., and Kuemmel, D. A. (1993). "Traffic volume reductions due to winter storm conditions." *Transportation Research Board*, National Researches Council, Washington, D.C.

Keith, K. K., and Leland, D. S. (2000). "Winter storm event volume impact analysis using multiple sources archived monitoring data." *Transportation Research Board,* 79th Annual Meeting, Washington, D.C.

Perrin, J. P., and Hansen, B. G. (2001). "Modifying signal timing during inclement weather." *Transportation Research Record 1748*, TRB, National Research Council, Washington, D.C.

Ren, Y. Y. (2008). "Study on the characteristics of traffic stream and management countermeasure at urban road under the condition of snow and ice." *Master's degree thesis of Jilin University*.

Transportation Research Board.(2000). "Highway capacity manual." *Transportation Research Board*, National Researches Council, Washington, D.C.

Performance Evaluation Model for Traffic Networks

Yang Jiang[1], Jianjun Wu[2] and Huijun Sun[1]

[1] School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China; Email: Hjsun1@bjtu.edu.cn
[2] State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

Abstract

The traffic congestion can aggravate the system function and the network efficiency, which leads to the decrease of traffic system performance. Therefore, the performance evaluation for traffic networks (e.g., railway and road) has extremely important practical significance in the transportation planning and management. Here, we analyze the performance in the scale-free network generated by Molloy-Reed algorithm. We compare two important network performance evaluation methods with the consideration of traffic congestion and structural efficiency respectively. A significant finding is that only congestion or efficiency is not accurate to evaluate the performance reasonably. In order to avoid shortcomings and find a remarkably compromise between the congestion and efficiency, we propose a combined model with the consideration of them. Additionally, the effect of weighted parameter on the performance of traffic system is investigated.
Key Words: Traffic congestion; structural efficiency; performance

Introduction

The growing traffic congestion has led to excessive energy consumption and increased level of pollution. The traffic congestion can aggravate the system function and the network efficiency which leads to the decrease of traffic system performance. Therefore, the performance evaluation for urban traffic network has extremely important practical significance in the transportation planning and management.

As regards the modeling of traffic networks, there is an urgent need of traffic models which are able to explicitly represent congestion phenomena, and that can be applied to large-scale networks, that is, to metropolitan traffic networks. Traffic networks models can be divided into three categories: microscopic models, where the

dynamics of each vehicle is explicitly represented by taking into account its own behavior and those of neighboring vehicles (Chandler et al., 1958); mesoscopic models, which consider single vehicles but do not model physical phenomena (Prigogine, 1961); and macroscopic models, where vehicles are considered as a whole fluid entity, and thus the models are characterized by flows and densities (Daganzo, 1994). The methods above just take the road conditions into count, ignoring the effect of the network structure.

The structure property on network has attracted a tremendous amount of recent interest since the finding of scale-free networks with the power law degree distribution form: $p^k \sim k^{-\lambda}$, where k is the number of links of a randomly chosen node in the network and γ is the scaling exponent (Barabási, 1999; Albert, 2002). The traffic and transportation network is very complex and such a traffic network transports flows. Links with large capacity can transport heavier flows. Recently, Gao et al.(2007) investigated the urban road network and found a scale-free network with small-world characteristic based on the *GIS* technology. Wu et al. (2004; 2006) studied the topology complexity of Beijing public transport network and found that it had scale-free and small-world properties.

There are some works on the performance evaluation for traffic networks from the viewpoint of network structure. For example, Latora and Marchiori (2001; 2003; 2004) discussed the network performance issued by measuring the "global efficiency" in a weighted network as compared to that of the simple non-weighted small-world network. In a weighted network, the network is not only characterized by the edges that connect different nodes, but also by the weights associated with different edges in order to capture the relationship between different nodes. There are very few papers to-date that consider network flows in assessing network performance. The results in Zhu et al.(2006) are notable since they demonstrate empirically through an application to the airline network of China how a measure with flows and costs outperforms existing measures in yielding more realistic results in terms of, for example, which cities are critical and their rankings in the network. Qiang and Nagurney (2008) proposed the network performance measure that can be used to assess the efficiency of a network in the case of either fixed or elastic demands. Recently, Jenelius et al. (2006) proposed several link importance indicators and applied them to the road transportation network in northern Sweden. Murray-Tuite and Mahmassani (2004) also focused on identifying indices for the determination of vulnerable links in transportation networks. But these works mentioned above did not consider the effects of traffic congestion on the performance.

In fact, only congestion or efficiency is not accurate to be used to evaluate the performance. This paper proposes an evaluation method with the consideration of traffic congestion and structural efficiency. Additionally, we investigate the effects of weighted parameter on the performance in the scale-free traffic network.

The paper is organized as follows. Two definitions, e.g., the congestion factor and

efficiency, are described in section 2. Section 3 is the model and simulation. Finally, the conclusion is given.

Congestion Factor and Efficiency

Congestion and efficiency are two key measures to evaluate the network performance. The former shows the total cost of traffic system, while the latter is to describe the structural efficiency.

Wu et al., (2006b) gave a definition of congestion factor *Cong* based on the numbers of congested links, which is written as follow:

$$Cong = \frac{TC}{TN} \tag{1}$$

where, *TN* and *TC* are the number of total links and total congested links in the traffic system. Apparently, *Cong*=0corresponds to free traffic on the network without congestion, and *Cong*=1 indicates the worst case of congestion or the breakdown of the network.

Another measure of system performance is efficiency proposed by Latora and Marchiori (2001):

$$Effi = \frac{1}{N(N-1)} \sum_{i \neq j \in G} \frac{1}{d_{ij}} \tag{2}$$

where d_{ij} is the shortest path length between two generic nodes i and j, and N is the total number of nodes in the network.

In this paper, we perform the simulation with the scale-free network generated by Molloy-Reed algorithm (Molloy, 1998). At each time, a certain traffic flow (one unit traffic flow) is added into the system and assigned onto the network with user-equilibrium models (Sheffi, 1985). After t steps, the plot of total traffic flow Q_{total} versus congestion factor *Cong* is given in Figure.1(a), in which we can see the relationship between degree exponent λ and the congestion factor, and the congestion factor in the scale-free network is limited. Figure.1(a) also shows that a small λ of scale-free network corresponds to a small congestion factor when the total volume of traffic Q_{total} is larger. Figure.2(b) is the plot of total volume of traffic Q_{total} versus the efficiency *Effi* for the scale-free network with different degree distribution exponents λ=2.2, λ=2.5, λ=3.0.

Figure.1 shows that the efficiency decreases with the volume of traffic flow for all cases considered. Especially, the congestion factor and efficiency are different for the same traffic flow Q_{total}. For example, the congestion factor is about 0.08 when the volume of traffic flow Q_{total}=25 for three cases. But the efficiency is different entirely, as we see from Figure.1(b). The result comes that the level of road traffic conditions is the same from Figure.1, but indeed the efficiency is different. It means that both congestion factor and efficiency should be considered to determine the optimal network topologies.

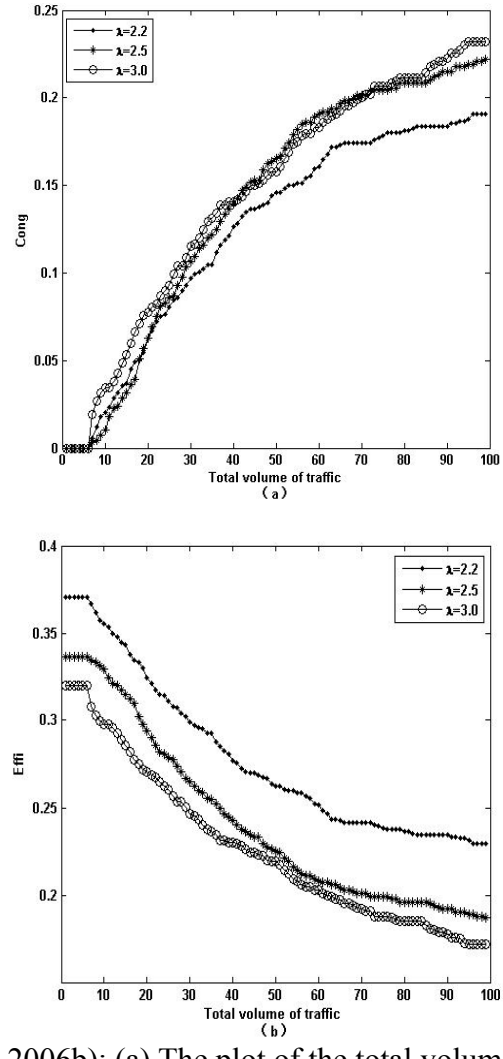


Figure 1. (Wu et al., 2006b): (a) The plot of the total volume of traffic Q_{total} versus the congestion factor *Cong* for scale-free network with N=400 and $\langle k \rangle \approx 4$ for different exponent λ=2.2, λ=2.5 and λ=3.0.
(b) The plot of the total volume of traffic Q_{total} versus the efficiency *Effi* for scale-free network with N=400 and $\langle k \rangle \approx 4$ for different exponent λ=2.2, λ=2.5, λ=3.0.

Performance Evaluation Model and Simulations

Generally, a network will usually be better if its performance is well. As we have seen, however, the way which we measure the performance will vary for different real networks. In a road network, most travelers look for routes that include less congested links. While for network planners, the global efficiency is often considered more importantly. Real-world networks appear to find a remarkably good compromise between them. In order to avoid disagreement and find a remarkably compromise between the congestion and efficiency, we consider a combination form of them. The performance evaluation P is given as:

$$P = h\frac{1}{Cong} + (1-h)Effi \tag{3}$$

where $0 \le h \le 1$ is the weighted parameter which determines the preference for measuring performance in terms of congestion or efficiency. *Cong* expresses the congestion level of networks calculated by the equation (1). *Effi* denotes the network efficiency obtained by equation (2).

Here, we adopt *P* to quantify the capabilities of networks in terms of congestion and efficiency, as described above.

In the network generated by the MR algorithm, the weight on each link is described by the cost (or travel time) on that link. We assume that the cost of one link is a function of the flow on that link, and this function is assumed to be positive and increasing, which is called the congestion effects. Generally, the cost of link $i \to j$ can be described with the following formulation $t_{ij}(y_{ij}) = t_{0ij} \times [1 + \alpha (y_{ij}/U_{ij})^{\beta}]$ offered by U.S. Bureau of Public Road (Sheffi, 1985).

Because of the congestion effects, the network loading mechanism is different from the traditional weighted networks with the constant cost (weights). Here, we adopt the user-equilibrium models to assign the flow to the network (Sheffi, 1985). The node with the largest load in the network can be found with the formulation $f_i = \sum_{j=1}^{N} e_{ij} y_{ij}$, where e_{ij} is an adjacency matrix, y_{ij} is the flow on the link between node *i* and node *j*. If there is an edge between node *i* and node *j* in the network, the entry e_{ij} is 1, otherwise e_{ij}=0. Then, we remove this node and perform UEassignment again. In this process, we calculate the congestion factor and efficiency.

Figure.2 shows the relationship between the performance *P*, congestion factor *Cong* and efficiency *Effi* for different *h* value (h=0, h=0.2, h=0.4, h=0.6, h=0.8, h= 1.0) in the scale-free network with exponents λ=3.0. We see that, *P* decreases with the increase of *Cong* wether the value of *Effi* is, and *P* becomes larger and larger with the increase of the weighted parameter *h*. In Figure.2 (b), it is found that a large *Effi* corresponds to a large *P* value. Especially, for the same congestion factor, *P* increases with the increase of *h* in Figure.2 (a). The similar result also can be derived from Figure.2 (b). We can see the Figure.2 as two projections of Figure.3, and Figure.2 reflect the trend that *P* becomes larger with the increase of the weighted parameter *h*.

Figure.3 represents the relationship among the performance, the congestion factor and the efficiency. Both small *Cong* and large *Effi*, correspond the high performance. For example, when $Cong \approx 0$ and $Effi \approx 0.3$, performance *P* will reach 2 to 2.5, which represents the network has a high performance level. On the contrary, when $Cong \approx 1$ and $Effi \approx 0$, performance *P* approaches to 0, which represents the lower performance with a bad case.

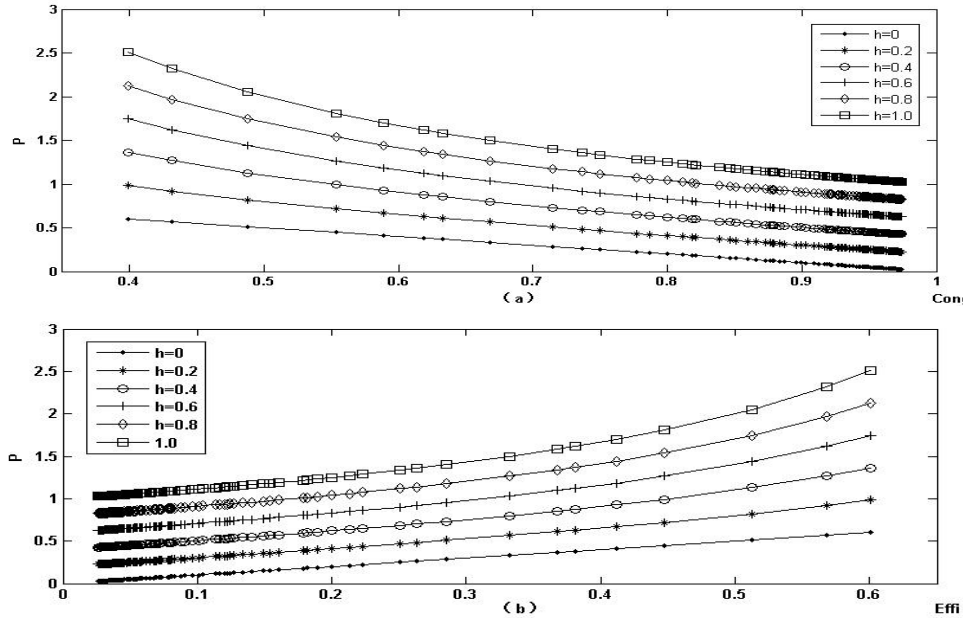


Figure 2. (a) Performance P versus *Cong* of the scale-free network with N=400, λ =3.0 and $< k > \approx 4$ for different parameter h.
(b) Performance P versus *Effi* of the scale-free network with N=400, λ =3.0 and $< k > \approx 4$ for different parameter h.

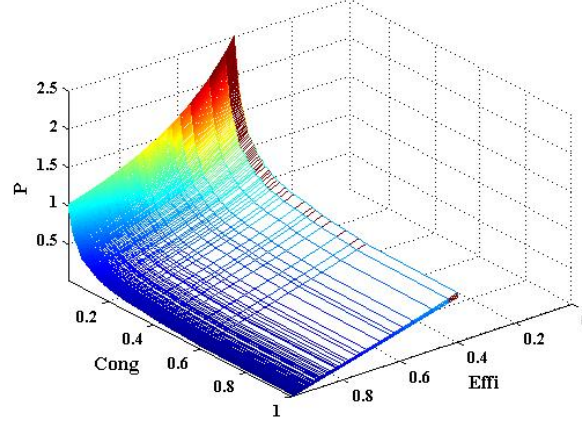


Figure3. Performance P versus congestion factor *Cong* and *Effi* of the scale-free network with N=400, λ =3.0, h=0.2 and $< k > \approx 4$.

For example, in the scale-free network with exponent λ =3.0, when the volume of traffic flow Q_{total} = 25, we can see the congestion factor is about 0.08 from Figure.1(a), and the efficiency is about 0.25 from Figure.1(b). To evaluate the network, the index P can be calculated using formula (3), and the result is as the same as the Figure.3 shows. In the scale-free network with different exponents λ

λ=2.2, λ=2.5and λ=3.0, we can determine the optimal network topologies using formula (3).

Conclusions

In this paper, the performance evaluation method is researched deeply. The simulation shows that only congestion or efficiency is not accurate to evaluate the performance. By introducing the congestion effect and structural efficiency, we propose a combined model with the consideration of congestion factor and efficiency. The simulations are performed in a scale free network. The results show that, in order to design the optimal traffic network, both the congestion and efficiency should be considered in the process of transportation planning and management.

Acknowledgements

This paper is partly supported by National Basic Research Program of China (2006CB705500), FANEDD (200763), NSFC of China (70801005 and 70871009), Beijing Natural Science Foundation (8102029) and the Foundation of State Key Laboratory of Rail Control and Safety (RCS2008ZQ001).

Reference

Albert R, and Barabási A.-L. (2002). "Statistical mechanics of complex networks." *Modern Physics*, 74: 47.

Barabási A.-L., and Albert R. (1999). "Emergence of scaling in random networks." *Science*, 286: 509.

Chandler R.E., Herman R., and Montroll E.W. (1958). "Traffic dynamics: studies in car following." *Operations Research*, vol. 6, 165-184.

Daganzo C. (1994). "Cell transmission model, part II: network traffic." *Transportation Research B*, 28(2), 279-293

Gao Z H, Chen Z J, and Liu Y X, et al. (2007). "Study on the complex network characteristics of urban road system based on GIS." *Proceeding of SPIE*, 6754:67540N.

Jenelius E, Petersen T, and Mattsson L. (2006). "Road network vulnerability: Identifying important links and exposed regions." *Transportation Research Part A*, 20: 537-560.

Latora V, and Marchiori M. (2001). "Efficient behavior of small-world networks." *Physical Review Letters 87*.

Latora V., and Marchiori M. (2003). "Economic small-world behavior in weighted

networks." *The European Physical Journal B*, 32: 249-263.

Latora V., and Marchiori M. (2004). "How the science of complex networks can help developing strategies against terrorism." *Chaos, Solitons and Fractals*, 20: 69-75.

Murray-Tuite P M, and Mahmassani H S. (2004). "Methodology for determining vulnerable links in a transportation network." *Transportation Research Record*, 1882: 88-96.

Molloy, M., and Reed, B. (1995). "Random structure Algorithm." *Comb Probab Comput*, 7, 295.

Prigogine I. (1961). "A Boltzmann-like approach to the statistical theory of traffic flow." *Theory of traffic flow*, 158-164.

Qiang, Q., and Anna, N. (2008). "A unified network performance measure with importance identification and the ranking of network components." *Optimization Letters*, 2(1): 127-142.

Sheffi, Y. (1985). "Urban transportation networks: equilibrium analysis with mathematical programming methods." New Jersey: Prentice-Hall, Englewood Cliffs.

Wu, J. J., Gao, Z. Y., Sun, H. J., and Huang, H. J. (2004). "Urban transit system as a scale-free network." *Modern Physics Letters B*, 18: 1043-1049.

Wu, J. J., Gao, Z. Y., and Sun, H. J. (2006). "Complexity and efficiency of Beijing transit network." *International Journal of Modern Physics B*, 20: 2129.

Wu, J. J., Gao, Z. Y., Sun, and Huang, H. J. (2006b). "Congestion in different topologies of traffic networks." *Euorphysics Letters*, 74: 560-566

Zhu, D., Hu, Y., Li, Y., and Yu, B. (2006). "A new measure for airline networks performance evaluation and critical cities identification." Fudan University.

The Adjustment Factor of Road Section Crosswalk on Urban Road Capacity

Aizeng Li[1], Xinsheng Song[2], Binghua Wu[3] and Lihua Liu[4]

[1] Lecturer, [2] Associate Professor, [3] Lecturer, 4 Lecturer, Department of Traffic Engineering, Henan University of Urban Construction, New District, Pingdingshan City, Henan Province, China 467044; Email: liaizeng@163.com, sxs@hncj.edu.cn, wuiceflower@126.com, liulihuaalicia@126.com

Abstract

This paper quantitatively analyzes the effect of road section crosswalks on urban road capacity and the running characteristics of pedestrians, bicycles, and electric bicycles when they pass through a crosswalk were also studied. The equivalent people group was also defined to describe the common effects of pedestrians, bicycles, and electric bicycles on vehicles' running. Taking sample stability and application convenience into consideration, an analysis period of 15 minutes was adopted. Then, by means of statistical analysis, the values of equivalent people number, average equivalent people groups' time headway, and equivalent people groups' number of each 15 minutes were obtained. By means of a regression analysis, the 11 kinds of regression models between 15 minutes equivalent people number and average equivalent people groups' time headway, and the 11 kinds of regression models between 15 minutes equivalent people number and equivalent people groups' number were obtained. Based on the correlation coefficient values and application convenience, the inverse function was adopted to describe the relationship between equivalent people number and average equivalent people groups' time headway, and the cubic function was adopted to describe the relationship between equivalent people number and equivalent people groups' number. And then, the method of time resource occupancy was adopted to analyze the effect of pedestrians, bicycles, and electric bicycles on urban road capacity, and the adjustment factors under the condition of whether there is a signal for pedestrians to pass through a crosswalk were obtained. Research indicates that the average equivalent people groups' time headway decreases with the increase of equivalent people number, and the equivalent

people groups' number firstly increases and then decreases with the increase of equivalent people number. Case study indicates that when the traffic volume of pedestrians, bicycles, and electric bicycles is relatively high, signal control can effectively reduce the impact of road section crosswalks on urban road capacity.
Key Words: urban road capacity; road section crosswalk; adjustment factor; regression analysis; method of time resource occupancy

Introduction

Urban road capacity is important for urban road designing, rebuilding, and evaluating. When analyzing urban road capacity, the common method is that the designing capacity of a road is equal to its basic capacity multiplied by some adjustment factors including bicycle influence, lane width and number influence and so on (Wang and Guo 2000, TRB 2000). The bicycle influence adjustment factor refers to the bicycle on the sidewalk. In other related literatures, Yang et al. (2008) researched the effects of a bus stop on road capacity, and established models to analyze the relationship between bus arrival rate and road capacity. Gao et al. (2004) analyzed the effects of stochastic factors on actual capacity on road section, established relative capacity calculation modes by means of a BP neural network. Kockelman and Shabih (2000) researched the effects of light-duty vehicles on road capacity and presented a capacity adjustment factor of it. By investigating in Indiana, Perez-Cartagena and Tarko (2005) advised that the Highway Capacity Manual should add an urban population adjustment factor to reflect the effect of urban population on saturation flow rate. By analyzing saturation time headway on four different district types in Florida, Le et al. (2000) considered that the Highway Capacity Manual should add an amusement district type adjustment factor. Due to an increase in older drivers in the United States, Lu and Pernia (2000) provided an older driver adjustment factor to analyze its influence on road capacity. By investigating in Florida, Zhou et al. (2000) pointed out that the drivers coming from other districts have a great influence on saturation flow rate, when the proportion of the drivers is high, saturation flow rate and capacity will decrease 19%. By means of regression analysis, Liu and Fan (2005) provided an adjustment factor of U-Turn vehicle percent to analyze the influence of U-Turn vehicles on road capacity. In addition, Dunlap (2005) analyzed the effect of different weather conditions on saturation flow rate.

Looking from above, there are many literatures that studied the adjustment factors of road capacity considering different aspects. In existing literatures, there are a few that analyzed the influence of road section crosswalks on urban road capacity. In the paper, a convenient method will be provided to analyze the influence of crosswalks on urban road section capacity under the condition of whether there is a pedestrian signal on the crosswalk.

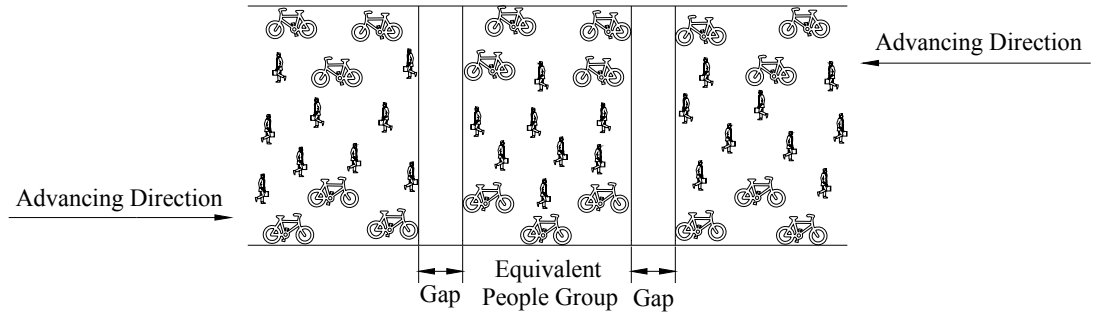


Figure 1. Sketch-map of gap and equivalent people group.

Definition of Equivalent People Group

Running Characteristics of Pedestrians, Bicycles, and Electric Bicycles on Crosswalk. There are pedestrians, bicycles, and electrical bicycles on crosswalks. By observation, though generally speaking, bicycles and electrical bicycles drive on the outside of the pedestrians, they also mutually infiltrate to other parts. In addition, when motor vehicles pass through the non-motorized vehicles and pedestrians, these vehicles generally look for the same gap and pass at the same time.

Definition of Equivalent People Group. On a crosswalk, taking a certain transect as a reference line, if there aren't non-motorized vehicles (including bicycles and electric bicycles) and pedestrians passing the line in a certain time segment, the time segment is defined as a gap, and the non-motorized vehicles and pedestrians between two adjacent gaps are defined as an equivalent people group, as shown in Figure 1.

The Correlation Analysis Models

Related Definition. Taking people as a conversion standard, according to the conversion coefficient values 2.0 and 2.3 (Li et al. 2006) of bicycle and electric bicycle, the number of bicycles and electric bicycles can be transferred to people number, to a certain equivalent people group. The sum of this people number is defined as equivalent people number of the equivalent people group. According to the definition of equivalent people group above, equivalent people groups' time headway is defined as the time interval between two adjacent equivalent people groups, and equivalent people groups' number is defined as the number of equivalent people groups in a determined period of time.

The Model between Equivalent People Number and Average Equivalent People Groups' Time Headway. Taking sample stability and application convenience into consideration, the analysis period of 15 minutes was adopted in the paper. By investigating Shuguang-Guangming, Kaiyuan-Youyue, and Youyue-Laodong

intersections in Pingdingshan city, Henan province, China, 45 samples of each 15 minutes were obtained, as shown in Table 1. Make regression analysis to these samples by Linear, Quadratic, Compound, Growth, Logarithmic, Cubic, S, Exponential, Inverse, Power, and Logistic functions, respectively, the results are shown in Table 2 and Figure 2.

Table 1. Statistical values of equivalent people number and average equivalent people groups' time headway.

Equivalent people number (person)	Average time headway (s)	Equivalent people number (person)	Average time headway (s)	Equivalent people number (person)	Average time headway (s)
330.78	6.38	328.45	6.09	670.13	4.21
349.18	5.98	340.61	6.3	453.34	5.1
321.17	6.54	339.83	6.35	360.92	5.76
348.49	5.77	333.73	5.82	680.99	4.24
351.77	6.17	330.14	7.4	437.11	5.74
266.51	8.89	411.29	5.6	1265.52	3.42
407.43	5.65	288.89	7.36	1035.8	3.75
343.57	6.43	274.8	6.86	507.59	5.36
272.15	7.75	343.74	6.12	309.58	6.61
338.11	6.31	274.13	6.86	1142.65	3.37
343.09	5.94	537.4	4.82	1221.27	3.74
342.81	5.98	424.67	5.45	440.26	5.88
365.18	5.94	327.48	6.18	1334.8	3.51
392.3	6.01	452.21	5.15	778.95	4.09
275.03	7.35	401.29	5.49	492.47	5.5

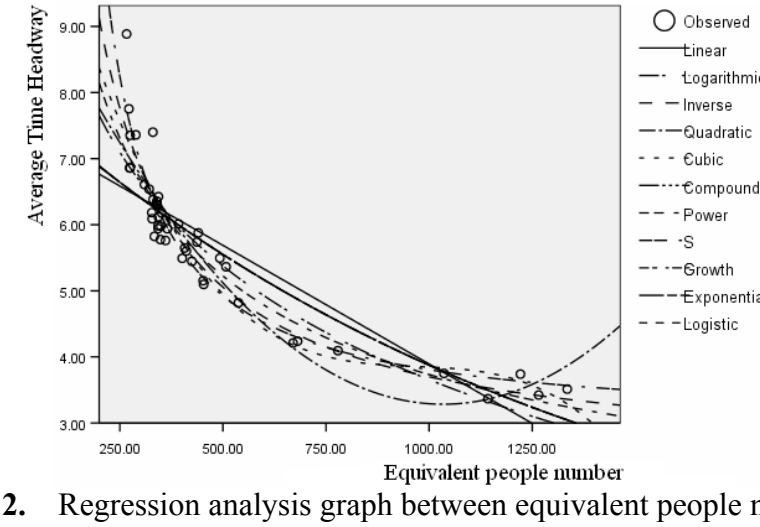


Figure 2. Regression analysis graph between equivalent people number and average time headway.

Table 2. Regression analysis results between equivalent people number and average equivalent people groups' time headway.

Regression Equation	Model Summary					Parameter Estimates			
	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	0.725	113.428	1	43	0.000	7.484	-0.004		
Logarithmic	0.842	228.496	1	43	0.000	20.808	-2.483		
Inverse	0.905	408.498	1	43	0.000	2.342	1356.005		
Quadratic	0.867	137.489	2	42	0.000	10.172	-0.013	6.44E-06	
Cubic	0.891	112.105	3	41	0.000	12.752	-0.027	2.77E-05	-9.55E-09
Compound	0.832	212.755	1	43	0.000	7.950	0.999		
Power	0.919	490.147	1	43	0.000	106.412	-0.485		
S	0.936	627.543	1	43	0.000	1.079	257.724		
Growth	0.832	212.755	1	43	0.000	2.073	-0.001		
Exponential	0.832	212.755	1	43	0.000	7.950	-0.001		
Logistic	0.832	212.755	1	43	0.000	0.126	1.001		

Table 3. Statistical values of equivalent people number and equivalent people groups' number.

Equivalent People Number (person)	Equivalent People Groups' Number (group)	Equivalent People Number (person)	Equivalent People Groups' Number (group)	Equivalent People Number (person)	Equivalent People Groups' Number (group)
266.51	90	330.14	101	670.13	117
407.43	114	411.29	112	453.34	112
343.57	106	288.89	94	360.92	114
272.15	93	274.8	100	680.99	123
330.78	97	328.45	102	437.11	120
349.18	105	340.61	103	1265.52	100
321.17	102	339.83	104	1035.8	108
348.49	112	333.73	113	507.59	122
351.77	108	365.18	111	309.58	105
392.3	107	274.13	102	1142.65	102
275.03	95	537.4	115	1221.27	103
343.74	102	424.67	110	440.26	117
338.11	105	327.48	109	1334.8	98
343.09	107	452.21	115	778.95	120
342.81	112	401.29	115	492.47	114

Taking correlation coefficient R square values and application convenience into consideration comprehensively, the inverse function was adopted to describe the relationship between equivalent people number and average equivalent people groups' time headway, as shown in Eq. 1.

$$y_t = 2.342 + 1356.005/x, \quad \text{for } 200 \le x \le 1400 \qquad (R^2 = 0.905) \qquad (1)$$

where y_t = average equivalent people groups' time headway in 15 minutes(s); and x = equivalent people number in 15 minutes (person).

The Model between Equivalent People Number and Equivalent People Groups' Number. By investigating the same intersections above, the 45 samples of each 15 minutes were obtained, as shown in Table 3. Regression analyses to these samples by Linear, Quadratic, Compound, Growth, Logarithmic, Cubic, S, Exponential, Inverse, Power, and Logistic functions, respectively, are shown in Table 4 and Figure 3.

Taking correlation coefficient R square values and application convenience into consideration comprehensively, the cubic function was adopted to describe the relationship between equivalent people number and equivalent people groups' number, as shown in Eq. 2.

$$y_g = 30.893 + 0.338x - 3.94 \times 10^{-4} x^2 + 1.34 \times 10^{-7} x^3, \quad \text{for } 200 \le x \le 1400 \quad (R^2 = 0.778) \quad (2)$$

where y_g = equivalent people groups' number in 15 minutes (group); and x = equivalent people number in 15 minutes (person).

Table 4. Regression analysis results between equivalent people number and equivalent people groups' number.

Regression Equation	Model Summary					Parameter Estimates			
	R Square	F	df1	df2	Sig.	Constant	b1	b2	b3
Linear	0.002	0.099	1	43	0.755	106.815	0.001		
Logarithmic	0.050	2.272	1	43	0.139	82.714	4.085		
Inverse	0.161	8.255	1	43	0.006	117.182	-3853.973		
Quadratic	0.684	45.466	2	42	0.000	67.187	0.145	-9.50E-05	
Cubic	0.788	50.794	3	41	0.000	30.893	0.338	-3.94E-04	1.34E-07
Compound	0.002	0.100	1	43	0.753	106.515	1.000		
Power	0.050	2.271	1	43	0.139	84.972	0.038		
S	0.162	8.291	1	43	0.006	4.766	-36.220		
Growth	0.002	0.100	1	43	0.753	4.668	1.29E-05		
Exponential	0.002	0.100	1	43	0.753	106.515	1.29E-05		
Logistic	0.002	0.100	1	43	0.753	0.009	1.000		

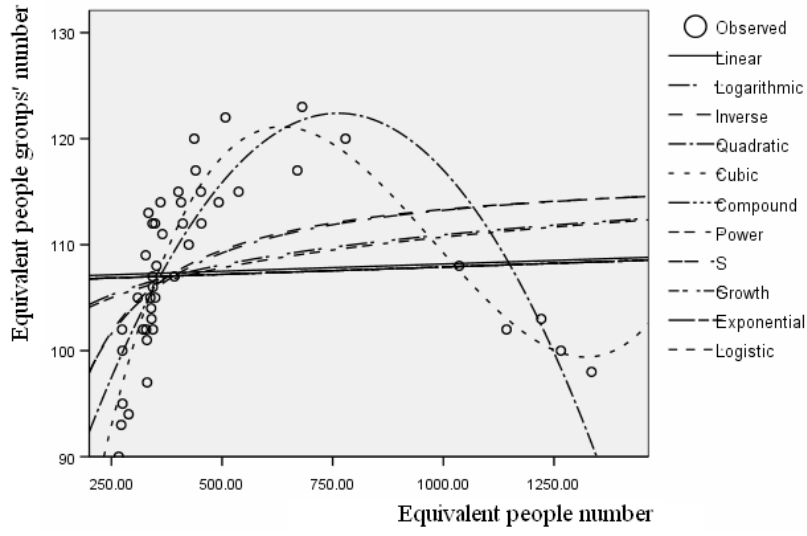


Figure 3. Regression analysis graph between equivalent people number and equivalent people groups' number.

The Adjustment Factor of Road Section Crosswalk

Crosswalk without Pedestrian Signals. For a crosswalk without pedestrian signals, if the traffic demand is known, using Eqs. 1 and 2, the average equivalent people groups' time headway and equivalent people groups' number can be obtained. Based on the method of time resource occupancy, adjustment factor of road section crosswalk can be obtained by Eq. 3:

$$f_c = (t - y_t y_g)/t \tag{3}$$

where f_c = adjustment factor of road section crosswalk on urban road capacity when there isn't a pedestrian signal on a crosswalk; t = length of analyzed time interval(s); y_t = average equivalent people groups' time headway in analyzed time interval(s); and y_g = equivalent people groups' number in analyzed time interval(group).

The Crosswalk with Pedestrian Signal. If there is a pedestrian signal on a crosswalk, based on signal timing design, applying time resource occupancy method, the adjustment factor of road section crosswalk can be obtained by the following Eq. 4：

$$f_c' = (T - t_c)/T = t_v/T \tag{4}$$

where f'_c= adjustment factor of road section crosswalk on urban road capacity when there is a pedestrian signal on the crosswalk; T = signal cycle length(s); and t_c, t_v = pedestrian and vehicle green time in a signal cycle, respectively.

Case Study

Southeast university crosswalk, as shown in Figure 4, is located on the Taiping North Road, Nanjing, China. Traffic demand in the early peak hours (07:40~08:40) is shown in Table 5 (Li et al. 2006).

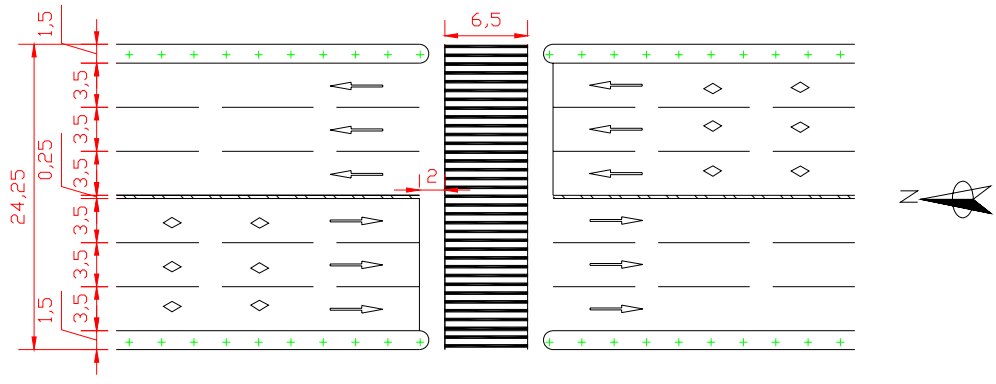


Figure 4. Sketch-map of southeast university crosswalk.

Table 5. Traffic demand in peak hour.

Direction	Pedestrian (Person)	Bicycle (Vehicle)	Electric Bicycle (Vehicle)	Equivalent People Number (Person)
From East to West	125	1174	28	2537
From West to East	67	24	2	120

On the current situation, there is a pedestrian signal lamp on the crosswalk. Pedestrian green time and signal cycle length are 20 and 65 s, respectively. Applying Eq. 4, the adjustment factor of the crosswalk on urban road capacity can be obtained:

$$f_c' = (T - t_c)/T = t_v/T = (65 - 20)/65 = 0.69$$

It is assumed that there isn't a pedestrian signal on the crosswalk and the equivalent people number average distributes every 15 minutes. Then using Eqs. 1, 2, and 3, the adjustment factor of the crosswalk can be obtained:

$$f_c = (t - y_t y_g)/t = (3600 - 4.38 \times 483)/3600 = 0.41$$

By the calculation results, it can be seen that when the traffic volume of pedestrians, bicycles, and electric bicycles increases, signal control can effectively reduce the effect of road section crosswalk on urban road capacity.

Conclusions

Based on the running characteristics of non-motorized vehicles and pedestrians on the crosswalk, the equivalent people group was defined to describe the common effects of this group on motor vehicles' running.

Using 15 minutes as an analysis period, by regression analysis, the correlation model between equivalent people number and average equivalent people groups' time headway was obtained. Research indicates that the average equivalent people groups' time headway decreases with the increase of equivalent people number.

Using 15 minutes as an analysis period, by regression analysis, the correlation model between equivalent people number and equivalent people groups' number was obtained. Research indicates that the equivalent people groups' number first increases and then decreases with the increase of equivalent people number.

Based on the method of time resource occupancy, the calculation formulas of the adjustment factors under the condition of whether there is a pedestrian signal on a crosswalk were obtained. The case study indicates that when the traffic volume of pedestrians, bicycles, and electric bicycles increases, signal control can effectively reduce the effect of road section crosswalk on urban road capacity.

References

Dunlap, B. M. (2005). "Field measurement of ideal saturation flow rate from the highway capacity manual." *M.S.C.E. Dissertation*, West Virginia University, Morgantown, West Virginia.

Gao, A. X., Chen, Y. Y., and Rong, J. (2004). "The road section actual capacity based on the BP neural network." *Journal of Beijing University of Technology*, 12(4), 462-465.

Kockelman1, K. M., and Shabih, R. A. (2000). "Effect of light-duty trucks on the capacity of signalized intersections." *Journal of Transportation Engineering*, 126(6), 506-512.

Le, X., Lu, J. J., and Mierzejewski, E. A. (2000). "Variations in capacity at signalized intersections with different area types." *Transportation Research Record*, 1710, 199-204.

Li, W. Y., Chen, X. W., and Wang, Q. (2006). "Algorithm of mid-block street crossing capacity and cross-street green time." *Journal of Wuhan University of Technology (Transportation Science & Engineering)*, 30(5), 751-754.

Liu, P., Lu, J. J., and Fan, J. J. (2005). "Effects of U-turns on capacities of signalized intersections." *Transportation Research Record*, 1920, 74-80.

Lu, J. J., and Pernia, J. C. (2000). "Traffic capacity reduction at signalized intersections". *Proc., 2000 Annual Conference* - Canadian Society for Civil Engineering, London, Ont., Canada.

Perez-Cartagena, R. I., and Tarko, A. P. (2005). "Calibration of capacity parameters for signalized intersections in Indiana." *Journal of Transportation Engineering*, 131(12), 904-911.

Transportation Research Board (TRB). (2001). *Highway capacity manual 2000*, Transportation Research Board, Washington, D.C.

Wang, W., and Guo, X. C. (2000). Traffic Engineering. Southeast University Press, China.

Yang, X. K., Cao, J., and Gong, J. (2008). "Study of the effect of bus stop on roadway section capacity." *Journal of Beijing University of Technology*, 34(1), 65-71.

Zhou, Y., Lu, J. J., and Mierzejewski, E .A. (2000). "Development of driver population factors for capacity analysis of signalized intersections." *Transportation Research Record*, 1710, 239-245.

Effects of Speed Difference on Capacity of Freeway

Hao Wang[1] and Wei Wang[2]

[1] Lecturer, [2] Prof., Jiangsu Provincial Key Lab of Transportation Planning and Management, Southeast University, 2 SiPaiLou, Nanjing, China 210096,
E-mail: haowang@seu.edu.cn, wangwei@seu.edu.cn

Abstract

This paper studies the effects of speed difference on the real capacity of freeway, based on the speed difference definition of Average Speed Difference. It is found that there is no obvious relationship between speed difference and the real capacity under high level of service, while under medium and low level of service the real capacity is lower than the potential capacity on the effects of speed difference. Based on the analysis of safe car-following behaviors, the relations between vehicle speed, space headway and speed difference are established, from which the modifying factor of real capacity due to effects of speed difference is derived. Compared with existent models, the model presented in this paper is more understood theoretically, and it is able to describe the effects of speed difference on the freeway capacity under various levels of service.
Key Words: traffic engineering; freeway capacity; speed difference; car-following theory

Introduction

The importance of speed difference as one of the most typical characteristics of the unstable traffic flow has been recognized by more and more researchers recently. The empirical observations on the traffic flow on German freeways indicate that the dense traffic flow usually corresponds to extremely broad scattering points on the flow-density plot (Kerner 2004). Cassidy (1998) found that the non-stationary data with large speed difference always fall beneath the near-stationary data on the fundamental diagram. Del Castillo (1995) used data with different speed standard deviation to calibrate the speed-density function. It was found that the data with small speed standard deviation fitted the functional form of the speed-density

relationship nicely. Many researches indicate that speed difference has effects on the flow of traffic stream, and which leads to the reduction of the real capacity of freeways. Zhuge (2000) established an empirical relationship between the speed difference and the flow rate, based on the traffic data from expressways in Beijing. Among all the related researches, the stand deviation of individual speeds is used to describe the speed difference. While there is still no theoretical explanation for the capacity reducing phenomenon resulted from speed difference.

This paper attempts to explore the functional relationship between the speed difference and the real capacity of freeways, based on a new speed difference definition "Average Speed Difference (ASD)" (Wang 2007). In the next section, the new definition of speed difference is introduced. In the third section, the capacity under the effects of speed difference is studied based on the empirical data. In section 4, the relations between vehicle speed, space headway and speed difference are established, based on the analysis of safe car-following behaviors. Moreover, the modifying factor of real capacity related to effects of speed difference is derived. Section 5 is devoted to the conclusion and some further discussion.

Definition of Average Speed Deference

Average Speed Difference (ASD) is defined as the average absolute value of all the differences between speeds of neighbouring vehicles during the observation period. The mathematical expression is given as follows:

$$ASD := \overline{\Delta v} = \frac{\sum_{i=1}^{n-1} |v_i - v_{i+1}|}{n-1} \tag{1}$$

where v_i denotes the speed of the i-th vehicle passing the observation point. In a recent research it was found that ASD is negatively correlated with flow rate in dense traffic flow (Wang 2007). In the following parts, the relation between the speed difference and the real capacity of freeway will be presented based on empirical data and theoretical analysis.

Real Capacity under the Effects of Speed Difference

In order to take deep insight into the effects from the speed difference on the real capacity of freeways, the traffic flow data collected from two segments of Dutch freeway A10 and A20 are studied. Both the two freeway segments have good geometric and physical performance. The proportion of heavy vehicles is less than 2% during the observation period. Such data allows us to neglect the effects on the real capacity that comes from the factors like the width of the lane, the grade, the heavy vehicles, and so on, so that we can concentrate on the effect of the speed difference on the real capacity of freeways.

Considering that HCM 2000 recommends 15 minutes as the analytical period for capacity analysis, all the observed data are integrated into 15-minute data in the first step of data processing. Then the data that involves remarkable fluctuations on volumes or speeds of traffic flow is eliminated, in order to ensure that traffic maintained a relative stable state during the analytical period. After applying the above criterion, 51 data of freeway A10 and 54 data of freeway A20 are obtained for the next studies.

For all these selected data, the flow rate, average speed, density and ASD are calculated. Then the ideal speed-flow curves are estimated for both freeway A10 and freeway A20 according to the methodology of HCM 2000. From figure 1(a) and 1(b), we can see that all the data points fall beneath the ideal speed-flow curves. In order to take deep insight into the relationship between the speed difference and the real capacity of the traffic flows on freeway A10 and A20, the ratio of potential capacity to real capacity is presented on the plot together with the ASD, as shown in figure 1(c) and 1(d). From figure 1(c) and 1(d), we can see that there is a critical flow rate which is around 600 pc/h/ln. The ratio of potential capacity to real capacity is bigger than 1 when the flow rate is higher than the critical flow rate. While for the situation that the flow rate is lower than the critical flow rate, the ratio of potential capacity to real capacity fluctuates around 1. Therefore, the effects on real capacity from the speed difference are obvious when the flow rate is higher than 600 pc/h/ln, namely, at level of service B.

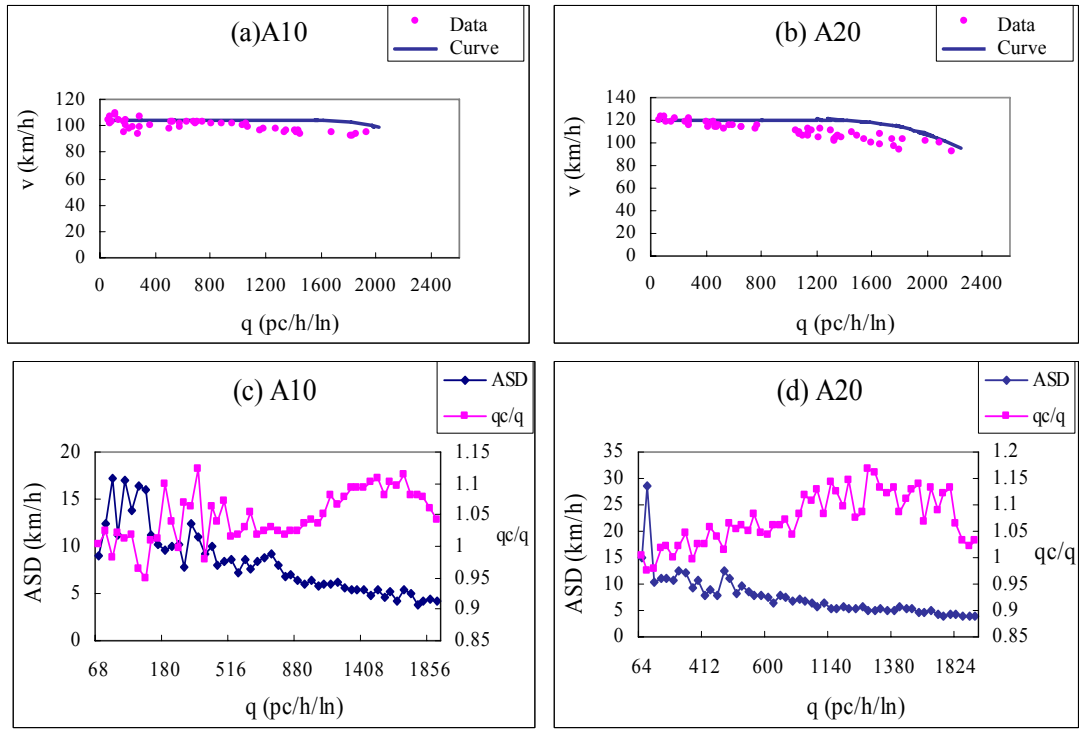


Figure 1. Effects on the real capacity from the speed difference.

Functional Relationship between the Speed Difference and the Capacity

As we all know, drivers always like to keep themselves some distances from previous vehicles, so as to avoid collisions when the previous vehicles perform strong deceleration. If we assume that the following vehicle has the same deceleration performance as the previous vehicle does, then the minimum safe distance headway for the following vehicle should satisfy the following equation:

$$s(v) + s_d = v \cdot \Delta\tau + s_d + s_0 \tag{2}$$

where $s(v)$ is the minimum safe distance headway for the following vehicle when both the previous vehicle and the following one are driving at speed v. s_d is the braking distance. $\Delta\tau$ is the response time of the following driver. s_0 is the minimum safe distance headway of stoppage. Figure 2 presents the trajectories of the vehicles during the deceleration process.

Rewrite the equation (2), we get the minimum safe distance headway function with regard to the speed, as

$$s(v) = v \cdot \Delta\tau + s_0 \tag{3}$$

Equation (3) is consistent with the lower order car following model (Newell 2002). When traffic is in the stable state, all vehicles move with the same distance headways and the same speeds. Therefore, the distance headway can be expressed as the function with regard to the traffic density k, namely, $s(v) = 1/k$ and $s_0 = 1/k_j$. Then the equation (3) can be transformed into

$$v = \frac{k_j - k}{\Delta\tau \cdot kk_j} \tag{4}$$

where k_j denotes the traffic density at jam state.

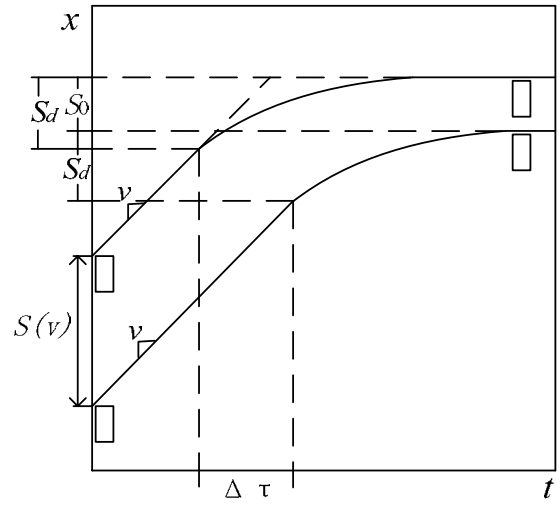


Figure 2. Trajectories of the vehicles during the deceleration process.

On the other hand, when traffic is in unstable state, the speeds of vehicles fluctuate all the time. Therefore, the speed difference and the distance between the two neighboring vehicles also vary with time. At time *t*, the speed difference $\Delta v_i(t)$ and the distance $\Delta x_i(t)$ between the *i*-th vehicle and the (*i*-1) st vehicle can be expressed as follows:

$$\Delta v_i(t) = v_{i-1}(t) - v_i(t) \tag{5}$$

$$\Delta x_i(t) = x_{i-1}(t) - x_i(t) = \Delta x_i(0) + \int_0^t \Delta v_i(t) dt \tag{6}$$

where the (*i*-1) st vehicle is the leading vehicle, the *i*-th vehicle is the following vehicle, $\Delta x_i(0)$ is the initial distance headway.

Considering that the speed difference between every two neighboring vehicles varies and distributes randomly among the traffic stream, for the sake of convenience, we give the following assumptions:

(a) All the speed differences between neighboring vehicles have the same periodic movement as

$$\forall i,\ \exists : \Delta v_i(t) \equiv \Delta v(t) = \Delta v_{\max} \sin(2\pi\omega t) \tag{7}$$

where $\Delta v_{\max}$ is the maximum speed difference, ω is the frequency. Substitute equation (7) into equation (6), we get

$$\forall i,\ \exists : \Delta x_i(t) \equiv \Delta x(t) = \Delta x(0) - \frac{\Delta v_{\max}}{2\pi\omega} \cos(2\pi\omega t) \tag{8}$$

(b) The speed difference ASD can be regarded as the average value of absolute speed difference within one cycle

$$ASD := \overline{\Delta v} = \frac{1}{T}\int_0^T |\Delta v(t)| dt = \omega \Delta v_{\max} \int_0^{\frac{1}{\omega}} |\sin(2\pi\omega t)| dt = \frac{2}{\pi} \Delta v_{\max} \tag{9}$$

(c) The average distance headway can be calculated as

$$s\left(v^{(ASD)}\right) = \frac{1}{T}\int_0^T \Delta x(t) dt = \omega \int_0^{\frac{1}{\omega}} \left[\Delta x(0) - \frac{\Delta v_{\max}}{2\pi\omega} \cos(2\pi\omega t) \right] dt = \Delta x(0) \tag{10}$$

where $s\left(v^{(ASD)}\right)$ is the average distance headway at the average speed $v^{(ASD)}$.

When taking the fluctuation of the distance headway into consideration, the minimum distance headway should satisfy the equation (4)

$$\min(\Delta x(t)) = v^{(ASD)} \cdot \Delta\tau + s_0 \tag{11}$$

Notice the equation (8) and the equation (10), the equation (11) can be transformed into

$$s\left(v^{(ASD)}\right) - \frac{\Delta v_{\max}}{2\pi\omega} = v^{(ASD)} \cdot \Delta\tau + s_0 \tag{12}$$

Substitute equation (9) into equation (12), we get

$$s\left(v^{(ASD)}\right)-\frac{\overline{\Delta v}}{4\omega}=v^{(ASD)}\cdot\Delta\tau+s_0 \tag{13}$$

Moreover, if we transform the equation (13) as the equation (4), we get

$$v^{(ASD)}=\frac{k_j-k}{\Delta\tau\cdot kk_j}-\frac{\overline{\Delta v}}{4\omega\cdot\Delta\tau} \tag{14}$$

Therefore, the ratio of the real capacity to the potential capacity in the same traffic density can be obtained according to equation (4) and equation (14)

$$\frac{q^{(ASD)}}{q_c}=\frac{k\cdot v^{(ASD)}}{k\cdot v}=\frac{v^{(ASD)}}{v}=1-\frac{kk_j}{4\omega\left(k_j-k\right)}\overline{\Delta v} \tag{15}$$

We should mention that the frequency ω of the speed difference is not a constant value. The frequency ω increases with the traffic density increasing, since the shorter the distance headway is, the more sensitive the follower will be. Without loss of generalization, we assume that the frequency ω can be expressed as the function of the traffic density k

$$\omega=\alpha\exp\left(\beta\cdot k\right) \tag{16}$$

where α and β are parameters needed to be calibrated. Using the traffic data of Dutch freeway A10 and A20 to calibrate equation (15) and equation (16), we get the results as shown in table 1.

Table 1. Results of the calibration

	k_j (veh/km)	α	β
Freeway A10	150	26.0	0.14
Freeway A20	150	25.7	0.15

The best-fit parameters for motorway A10 and A20 are very close to each other. After calibrating the model with real traffic flow data, we get the modifying factor $f^{(ASD)}$ of the real capacity under the effects of speed difference

$$f^{(ASD)}=\frac{q^{(ASD)}}{q_c}=1-\frac{kk_j}{104e^{0.14k}\left(k_j-k\right)}\overline{\Delta v} \tag{17}$$

The equation (17) indicates that the speed difference has effects on the real capacity of freeways. Moreover, the real capacity reduces with the speed difference increasing.

The model proposed in this paper is derived from safe driving behaviors in car following. It is not suitable for all kinds of traffic state, especially for the light traffic. When traffic is in free flow state, the distances between vehicles are quite large, which reduces the effects of speed difference on the car following behaviors. Therefore, this model cannot be used for light traffic with high level of service.

Acording to the classification of the level of service (LOS) in HCM 2000, the speed difference has no effects on the real capacity when the traffic is at LOS A. While for LOS B to E, the modifying factor of the real capacity can be obtained based on equation (17) as shown in figure 3.

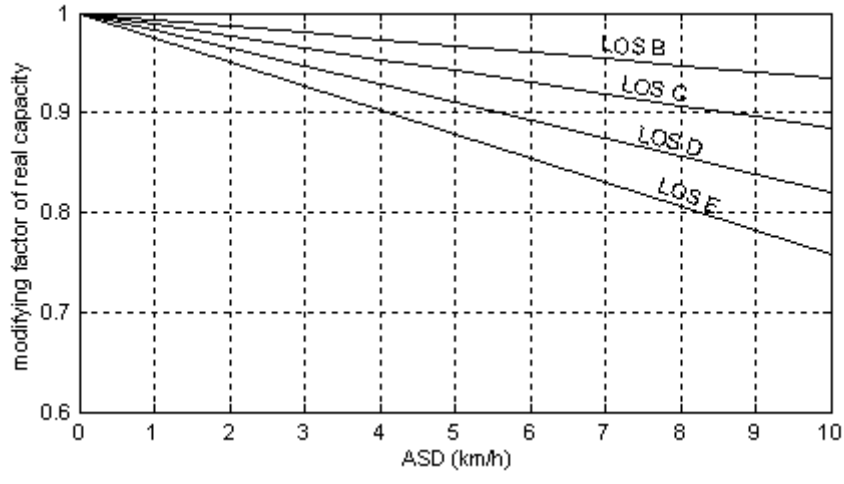


Figure 3. Impact of the speed difference upon the real capacity of freeways

From figure 3 we can see that the speed difference has different effects on the real capacity for LOS B to E. At high LOS such as LOS B, the speed difference has little impact on the real capacity; while at some lower LOS such as LOS D and E, obvious impact of the speed difference can be found upon the real capacity. Furthermore, it is found that modifying factor of the real capacity at LOS D and E comes to 0.8 when ASD reaches 10km/h. Therefore, the impact of the speed difference upon the real capacity cannot be neglected in capacity analysis.

Conclusion

This paper studies the real capacity of Dutch freeways under the impact of speed difference. It is found that when the speed difference is high, the real capacity is lower than potential capacity, even if the other conditions are close to ideal conditions. Such findings verify the viewpoint that speed difference has impact on the real capacity of freeways. A modifying factor model of real capacity under the impact of speed difference is proposed, base on the analysis of safe car following behaviors. This model takes the traffic density into account, which allows it to describe the effects of speed difference at different level of service.

We should mention that the model presented in this paper is derived from car-following theory. Therefore, it might not describe the capacity reducing resulted from lane-changing effects correctly. The effects of lane-changing on the speed difference and the real capacity of freeways will be the research work in the future.

References

Kerner, B. S. (2004). "Three-phase traffic theory and highway capacity." *Physica A*, 379-440.

Cassidy, M. J. (1998). "Bivariate relations in nearly stationary highway traffic." *Transportation Research Part B*, 49-59.

Del Castillo, J. M., and Benitez, F. G. (1995). "On the functional form of the speed-density relationship Ⅱ: empirical investigation." *Transportation Research Part B*, 391-406.

Zhuge, J. M. (2000). "A study of traffic behaviors on urban expressway." *Dissertation of Beijing University of Technology*, China.

Wang, H., Wang, W., and Chen, X. W., *et al.* (2007). "Experimental features and characteristics of speed dispersion in urban freeway traffic." *Transportation Research Record*, 150-160.

TRANSPORTATION RESEARCH BOARD. (2000). "Highway capacity manual 2000." *National Research Council, Washington, D.C*.

Newell, G. F. (2002). "A simplified car-following theory: a lower order model." *Transportation Research Part B*, 195-205.

Spatial Correlation Analysis of Congested Links in Urban Traffic Networks

Dan Ma, Huijun Sun and Linghui Han

MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing, 100044, PRC, Email: Hjsun1@bjtu.edu.cn

Abstract

A large number of facts have shown that emergency of congested links in urban traffic network has certain rule. This rule will help us to identify characteristics of the generation, propagation and dissipation of traffic congestion, which can provide some theoretical foundations for the development of reasonable traffic planning and control methods. In this paper, based on the cell transmission model (CTM), a model is proposed to analyze the propagation of traffic flow. The spatial autocorrelation method (SAM) is introduced to analysis the spatial distribution characteristics of congested links by using the average journey velocity (AJV) of links. Numerical results show that the traffic congestion appearance among adjacent or close links is interrelated, and with the increase of traffic demand, this relationship will become more and more obvious.

Key Words: cell transmission model; congested links; spatial autocorrelation analysis

Introduction

Many metropolises in China are already facing more and more serious traffic problems. Among all the problems, traffic congestion is the most critical one. The growing traffic congestion seriously restricts many cities' economic development and brings inconveniences and damages to people's daily work and life, which has become a major bottleneck for urban sustainable development. Some scholars have proposed methods to alleviate traffic congestion based on different characteristics of the cities, such as， Zhang et al. (2007) analyzed and appraised the status and solving strategy of traffic jam in Kunming city according to social investigation and the method of AHP. But similar studies emphasis more on practical experience in transportation. Other scholars researched traffic congestion characteristics based on

special structure of traffic network and propagation mechanism of traffic flow in the network. Wu et al. (2006) analyzed the degree of congestion and the efficiency in different types of networks. The emergence and evolution of traffic congestion are due to the increase of traffic demand. Long et al. (2008) employed cell transmission model (CTM) to establish a network traffic flow propagation model based on the fundamental work of Daganzo (1994, 1995). They proposed a new method of estimating average journey velocity of both link and network to identify network congestion bottlenecks. Long et al. (2008) investigated jam propagation and dissipation in two-way rectangular grid networks based on CTM. All above mentioned researches provided some help for us to recognize the emergence mechanism and characteristics of traffic congestion. But they do not consider the spatial distribution characteristics of congestion bottleneck.

Spatial correlation analysis as a kind of agglomeration characteristics and similar levels analysis method for spatial elements has made such a wide range of research and application results in population growth, economic growth, urban land prices, and traffic. Xiao et al. (2000) studied the planning information system of highway network by using spatial-temporal technology. Ding et al. (2007) built a system to analyze the spatial-temporal structure of the development about road passenger transportation by using spatial correlation theory. Wen et al. (2008) applied a Moran model to study the mutual relation among spatial locations of road traffic accidents based on the spatial autocorrelation method. Wang et al. (2009) used a spatial analysis approach to study the relationship between traffic congestion and road accidents. The M25 London orbital motorway was chosen to conduct this study. Results suggest that traffic congestion has little or no impact on the frequency of road accidents on the M25 motorway. Quddus (2008) developed a series of relationships between area-wide different traffic casualties and the contributing factors associated with ward characteristics using both non-spatial models and spatial models and to identify the similarities and differences among these relationships. Yuan et al. (2005) investigated spatial-temporal traffic analysis using a cross-correlation method, based on the eigenvector of the largest eigenvalue, motivated by insights about network dynamics at the macroscopic level.

Therefore, in this paper, we will analyze the spatial distribution characteristics of congested links based on the CTM and spatial analysis methods. To show how our proposed method might work in an urban traffic network, the simulation is performed in the given traffic network with different traffic demands. The rest of this paper is structured as four sections. In the following section, we will describe spatial analysis method. In Section 3, we introduce the network traffic flow propagation model based on CTM. Detailed simulation process and discussion of experiment results are described in Section4. Finally, the conclusion is given in Section 5.

Spatial Autocorrelation Method

SAM typically includes two aspects: global and local analysis. The former is to describe overall distribution characteristics of a phenomenon and determine whether

agglomeration characteristics exist in the space, while the latter is to measure similar levels of a phenomenon in region and to identify spatial concentration spots. SAM emphasizes on the "local" factor, Jian et al. (2005) has shown that this method is feasible and effective to analyze traffic performance characteristics in space and its further application will provide some theoretical basis for proposing effective measures to prevent and alleviate traffic congestion. In this paper, we mainly analyze distribution characteristics of congested links in urban traffic network from a global perspective.

In SAM, the matching degree between location and attribute similarity is used to describe the distribution characteristics of certain attributes in space. The local similarity indicates local relationship among each space unit, which is described by the spatial proximity matrix or the spatial weight matrix W. Let x represent certain attribute of space units, attribute similarity will be usually calculated by crossed product $x_i x_j$, or squared differences $(x_i - x_j)^2$ or absolute differences $\left|x_i - x_j\right|$.

Spatial Weight Matrix. Usually, a binary symmetric spatial weight matrix W is defined to indicate the adjacent relationship among multiple space units, as follows:

$$W = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & & \vdots \\ w_{n1} & \cdots & w_{nn} \end{bmatrix},$$

where w_{ij} denotes adjacent relationship between unit i and j, which is measured by adjacency criteria or distance criteria.

Here, links in traffic network as study objects, spatial weight matrix between them is defined as follows:

$$w_{ij} = \begin{cases} 1 & \text{link } i \text{ and link } j \text{ are connected} \\ 0 & \text{others} \end{cases}$$

when $w_{ij} = 1$,that is, link i is the upstream or downstream link of j.

Global spatial autocorrelation analysis. Global spatial autocorrelation analysis is used to describe overall distribution of certain phenomenon and estimate whether this phenomenon has agglomeration characteristics in space. In this paper, we use Moran's I coefficient to analyze the distribution characteristics of congested links in space. Firstly, it is important to select appropriate criterion to judge emergency of congestion. The traffic density, velocity and other parameters are usually adopted to describe travel condition of traffic flow in previous studies, such as Chen et al. (2009). Following Long et al. (2008), we use the average journey velocity (AJV) of links to identify congestion link.

Moran's I coefficient is one of the most representative indicator in global spatial autocorrelation analysis. In the paper, Moran's I coefficient can be reformulated as follows,

$$I = n\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}\left(x_i - \bar{x}\right)\left(x_j - \bar{x}\right) / \sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \quad (1)$$

where, x is the vector of the AJV of each link, $\bar{x}$ is the average AJV of all the links in network, n is the total number of links and W is the spatial weight matrix.

Generally, the value of Moran's I is range of [-1, 1]. If I >0, the correlation is positive, which indicates that distribution characteristics of adjacent areas are similar. The greater of the Moran's I is, the more obvious of this agglomeration phenomenon in space will be. When I <0, the negative correlation is hold. The result indicates that distribution characteristics of adjacent areas are considerably different. If I closes to zero, the random characteristics of the spatial distribution of congested links is derived.

Network traffic flow propagation model

To construct a reasonable model of traffic flow propagation in network is the basic work to study traffic network. However, because of traveler's route choice behavior the dynamics transmission process of traffic flow in the network becomes very complex. So it is important to construct a traffic flow transmission model, which not only considers the traveler's route choice behavior but also reflects the propagation effects of network traffic flow. Sun et al. (2007) has studied the routing mechanisms on complex traffic networks, which is a dynamic selective process. In this section, we employ CTM to establish network traffic flow propagation model based on the fundamental work of Long et al. (2008). In this model, the traveler's route is calculated base on initial impedance of network and fixed in whole process. However, the driver will consider two or more alternative routes instead of the only shortest route in general. So in this paper, the K-th shortest path is introduced in our model, in which the node model and the link model are included.

K-th Shortest Paths. The shortest path can not meet the requirements of travelers in practice, so the idea of K-th shortest paths comes to emerge. Among the algorithms to search K-th shortest paths, "link elimination method" is the most direct algorithm, which is based on the shortest path algorithm. Taking the practical application into consideration, we only use the shortest and the second shortest paths as alternative paths for travelers' random route choice in this paper. The basic step of the algorithm is given as follows:

Step 1. The shortest path generation. Obtain the shortest paths of each OD pair by using the dijkstra algorithm ;

Step 2. The second shortest path generation. Repeat the following for each OD pair:

Step 2.1. Generate a set of sub-networks by removing one link on the shortest path from the original network.

Step 2.2. Obtain a path set that consists of the shortest paths of the OD pair in all sub-networks.

Step 2.3. Find the shortest path in the path set as the second shortest path of the original network.

CTM Link Model. Daganzo (1994, 1995) proposed the CTM which simplified the solution scheme of the LWR model by adopting the following relationship between traffic flow, q, and density, k (Figure 1):

$$q = \min\{vk, q_{\max}, w(k_j - k)\}, \qquad 0 \le k \le k_j \tag{2}$$

In Eq. (2), we let $q_{\max}$, k_j denote the inflow capacity (or maximum allowable inflow) and jam density, v and w denote free-flow speed, the speed of the backward shock wave, respectively.

In the CTM, time is partitioned into intervals δ, each link is divided into a set of homogeneous cells, {1, 2, …,*I*}, where *I* is the number of cells.
Here, we employ the following notations to represent the CTM link model:
$n_i(t)$: The number of vehicles that are presented in cell i in time interval t.
$k_i(t)$: The density of cell i in time interval t.
$y_i(t)$: The number of vehicles that flow into cell i in time interval t.
$Q_i(t)$: The maximum number of vehicles that can flow into cell i in time interval t.
$N_i(t)$: The maximum number of vehicles that can be presented in cell i in time interval t.

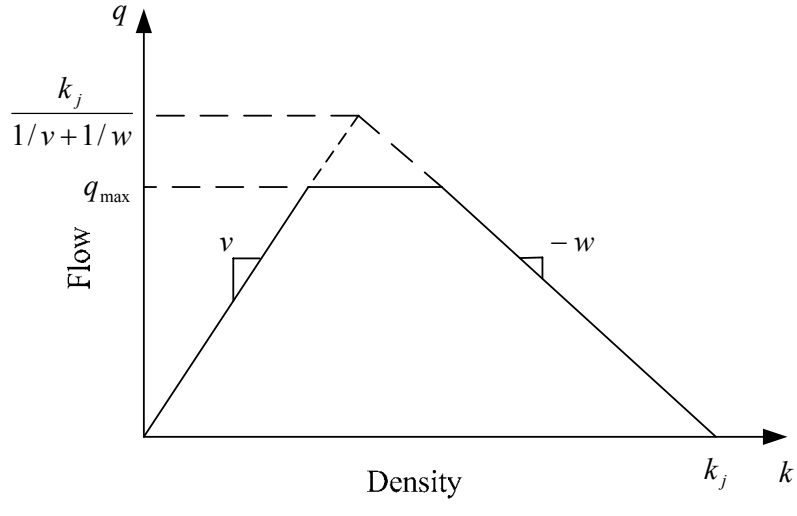


Figure1. The equation of state of CTM.

For any link a, based on the Eq. (2) and the relationship $n_i(t) = k_i(t) \cdot v\delta$ between density and the number of vehicles that are presented in cell i, we can obtain the cell-to-cell flow propagation as follows:

$$y_i(t) = \min\{n_{i-1}(t), Q_i(t), w(N_i(t) - n_i(t))/v\} \tag{3}$$

According to the flow conversation property, the discrete flow conservation equation can be expressed as:

$$n_i(t+1) = n_i(t) + y_i(t) - y_{i+1}(t) \tag{4}$$

CTM Node Model. Here, we employ the following notations to represent the CTM node model:

$y_{ab}(t)$: The number of vehicles that flow from the terminal cell of link a into the first cell of link b in time interval t.

$n_{ab}(t)$: The number of vehicles that will pass by link b and are presented in the terminal cell of link a in time interval t.

$n_{r,b}(t)$: The number of vehicles that are presented at node r (including the waiting vehicles before interval t) and can flow into link b in time interval t.

$y_{r,b}(t)$: The number of vehicles that generate at node r and flow into the first cell of link b in time interval t.

Extending the CTM link model to node condition, we can obtain the propagation equation of traffic flow from link a to link b:

$$y_{ab}(t) = \min\{n_{ab}(t), Q_1^b(t), w(N_1^b(t) - n_1^b(t))/v\} \quad (5)$$

$$y_{r,b}(t) = \min\{n_{r,b}(t), Q_1^b(t) - \sum_{a \in A_r} y_{ab}(t)\} \quad (6)$$

Simulation

Criterion of Congested Links. We use AJV of link to weigh traffic condition whether it is in a good condition or a bad condition. According to a document of the Ministry of Public Security of the People's Republic of China: Traffic Management Evaluation Index System of Urban Freeway 2005 (see web site of www.cein.gov.cn), we adopt the AJV of motor vehicles on main roads in metropolitan areas as the standard for evaluating the degree of traffic congestion. The standard is described as follows:

(1) Freely. The AJV of motor vehicles on main road is not less than 30 km/h in metropolitan areas.

(2) Light congestion. The AJV of motor vehicles on main road is less than 30 km/h but not less than 20 km/h in metropolitan areas.

(3) Congestion. The AJV of motor vehicles on main road is less than 20 km/h but not less than 10 km/h in metropolitan areas.

(4) Serious congestion (jam). The AJV of motor vehicles on main road is less than 10 km/h in metropolitan areas.

According to this standard, we can evaluate the traffic condition of each link by calculating its AJV. We use method of estimating AJV proposed by Long et al. (2008).

Input parameters of models. A large number of empirical studies demonstrate that urban transportation network has small world properties. Therefore, we construct an example network according to SW algorithms (Watts and Strogatz,1998) to illustrate the application of the proposed model. A typical two-way small-world network with 20 nodes and 80 links is generated. The average links of each node is 4 and the rewiring probability is 0.55.

The parameters for the CTM and notations are set as follows:
• The length of each time interval δ : 10 s.
• Jam density: 133 vehicles/km (i.e., 7.5 m for every vehicle).
• Free-flow speed: 54 km/h (15m/s).
• Backward shock-wave speed /Free-flow speed: 0.4.
• Number of lanes: 2.
• Flow capacity: 5 vehicles/time interval/lane.
•The length of each cell is 150 m, and the holding capacity of each cell is 20 vehicles.
• The number of cells of each link: 6.
• The OD demands loading intervals: 350.
• The coefficient of demand ϕ: actual demand/initial demand (the initial demand is one vehicle/s/one OD pair).

Results and Analysis. The coefficient of demand ϕ varies from 0.45 to 0.55 in the numerical test so as to increase the network OD demand. At the beginning, there is no congested link under the small demand level, and the total demand is 350 at each interval. Here, the time window T=[51,350] is chosen. In order to calculate reasonably and conveniently, T is divided into three smaller periods, e.g., T1=[51,150], T2=[151,250] and T3=[251,350]. The Moran's I under different demands can be calculated with Eq.(1) as shown in Table 1.

As shown in Figure 2, the value of I is greater than zero all along with different demand levels. The result implies that the emergency of congested links in network has agglomeration phenomenon. Moreover, with the increase of traffic demand, this phenomenon is more obvious. Therefore, it can be concluded that the network congestion appearance among adjacent or close links is interrelated, and with the increase of traffic demands, the relationship will become more and more obvious.

Table1. Moran's I and AJV of network under different coefficient of demand.

ϕ	AJV of network (km/h)	Moran's I (T1)	Moran's I (T2)	Moran's I (T3)	Moran's I (average)
0.45	49.47	0.0921	0.0659	0.1327	0.0969
0.46	43.11	0.0694	0.1503	0.1695	0.1297
0.47	37.37	0.1003	0.1691	0.2365	0.1686
0.48	28.94	0.1388	0.2058	0.2303	0.1916
0.49	25.4	0.1394	0.2183	0.2797	0.2125
0.5	22.98	0.1567	0.2243	0.2954	0.2255
0.51	22.21	0.1575	0.2478	0.2699	0.2251
0.52	19.1	0.1787	0.2885	0.2314	0.2329
0.53	17.24	0.2125	0.2767	0.2041	0.2311
0.54	16.74	0.2295	0.2808	0.251	0.2538
0.55	15.85	0.2647	0.2954	0.2431	0.2677

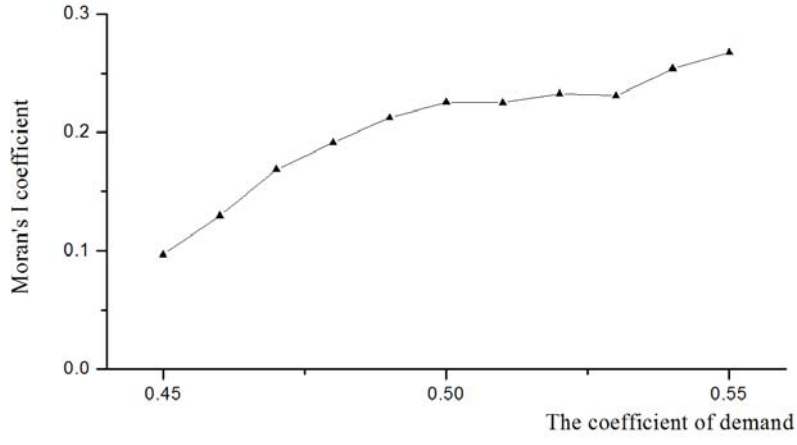


Figure 2. Moran's I under different coefficients of demand.

Conclusions

In this paper, the cell transmission model and the spatial autocorrelation method are used to discuss relationship between traffic demand and spatial distribution characteristics of congested links. The proposed approach is feasible and effective to analyze the complex characteristics of traffic congestion. Numerical results show that emergence of congested links has agglomeration characteristics in network, and with the increase of traffic demand this characteristics becomes more obvious. However, this paper only considers the global case, regional spatial autocorrelation characteristics of congested links will be investigated in the future.

Acknowledgements

This paper is partly supported by National Basic Research Program of China (2006CB705500), FANEDD (200763), Fok Ying Tong Education Foundation (111083), NSFC(70871009) and NSF of Beijing(8102029).

References

Chao, W., Mohammed A. Quddus, and Stephen G. Ison. (2009). "Impact of traffic congestion on road accidents: A spatial analysis of the M25 motorway in England", *Accident Analysis and Prevention,* 41, 798-808.

Chen, L., Qing-pu, Z., and Xue, Z. (2009)."Emergence and disappearance of traffic congestion in weight-evolving networks", *Simulation Modelling Practice and Theory*, 17(10), 1566-1574.

Daganzo, C. F. (1994). "The cell transmission model: a simple dynamic representation of highway traffic", *Transportation Research Part B,* 28(4): 269-287.

Daganzo, C. F. (1995). "The cell transmission model, Part II: Network traffic", *Transportation Research Part B,* 29 (2), 79-93.

Ding, H. L., and Li, L. H. (2007). "Analysis on spatio-temporal character of the development of road passenger transportation", *Journal of Highway and Transportation Research and Development,* 24(8), 149-158.

Jian, Y., and Kevin Mills. (2005). "A cross-correlation-based method for spatial–temporal traffic analysis." *Performance Evaluation,* 61:163-180.

Long, J. C., Gao, Z. Y., Ren, H. L. and Lian, A. P. (2008). "Urban traffic congestion propagation and bottleneck identification." *Science in China Series F: Information Sciences,* 55(7):948-964.

Long, J. C., Gao, Z. Y., Zhao, X. M., Lian, A. P., and Penina Orenstein.（2008）. "Urban traffic jam simulation based on the cell transmission model", *Networks and Spatial Economics,* 1566-113X (Print) 1572-9427 (Online), Springer Netherlands.

Quddus, M. A. (2008). "Modelling area-wide count outcomes with spatial correlation and heterogeneity: an analysis of London crash data", *Accident Analysis and Prevention,*40 (4), 1486-1497.

Sun, H. J.,Wu, J. J., and Gao, Z. Y.(2007). "Dynamics of routing mechanisms on traffic networks", *International Journal of Modern Physics C,* 18(11), 1755-1782.

Sun, H. J., and Wu, J. J. (2007). "Evolution of network from nodes dividing and generation", *Chinese Physics,* 16(6), 1581-1585.

Wen, H. Y. et al. (2008). "Spatial autocorrelation analysis of road traffic accidents based on moran model", *Computer and Communications,* 26(3), 31-33.

Watts, D. J. and Strogatz, S. H. (1998). "Collective dynamics of 'small-world' networks", *Nature (London),*393, 440-442.

Wu, J. J., Gao, Z. Y., and Sun, H. J. (2006). "Congestion in different topologies of traffic networks", *Europhysics Letters,* 74, 560-566.

Xiao, W. Z. et al. (2000). "Research and application of space time technology in highway network planning information system", *Journal of Highway and Transportation Research and Development,* 17(3), 43-46.

Zhang, G. J., and Hua, C. C. (2007). "An appraisal model for the status and solving strategy of trafic jam in Kunming city", *Journal of Yunnan Normal Universit*，27(1), 14-19.

http://www.cein.gov.cn/news/show.asp?rec_no=13698, June, 2006.

Improvement Project Choice Based on Efficiency Analysis

Gu Wang[1], Shengxue Zhu[2] and Xiucheng Guo[3]

[1] Dr., [3] Professor, Transportation College, Southeast University, Nanjing 210096, China, Email: transyscn@gmail.com, xcguo@163.com
[2] Lecturer, Faculty of Transportation Engineering, Huaiyin Institute of Technology, Huai'an 223003, China,; Email:zhushengxue601@126.com

Abstract

Traditionally, the evaluation of the road improvement project is mainly based on the indexes of V/C ratio and travel speed, aiming at improving the traffic performance (Chen, 2002). However, the improvement can guarantee neither the decrease of the spatiotemporal cost of single vehicle, nor the increase of the transportation efficiency. The evaluation is typically a problem of multi-input and multi-output, on which the traditional method fails to create the connection among all the inputs and the traffic performance, and the results selection is apt to be affected by subjective factors. Based on efficiency analysis theory, using data envelopment analysis, this paper builds a CCR model to evaluate an alternative scheme and to analyze and diagnose the ineffective scheme. This paper adopts the popular HCM analysis method and analyzes fluctuation of the index with time and exterior operation conditions (National Research Council, 2000). Results show that travel delay is more appropriate as the traffic operation index in the complex roadway environment. Moreover, approximate calculation methods of this index are discussed.
Key Words: efficiency analysis; DEA model; traffic operation; traffic improvement project

Introduction

There are two traditional assessment methods of improvement of the traffic operation. One looks on the improvement of traffic operation only to aim at improving traffic

operation. In this method, saturation and travel speed are taken as main indexes to evaluate service level with comparison to the foregoing situation and choose the implementation whose indicators are dominant. The other one is a comprehensive evaluation, which integrates broader indicators, such as environmental impacts, capacity, and so on. Analysis hierarchy process (AHP) and gray correlation assessment methods are often used. These methods have the advantages of multi-objective, multi-level, and multi-criteria in decision-making. But the evaluation results are often dominated by subjective factors; what is more to characterize indexes may lead to loss of varying degrees. Therefore, evaluation results need a margin of error.

From the perspective of the road operation and management, the road as a business object also meets the general characteristics of production systems. With the increase of traffic resources input, the level of traffic operation will increase, the operating cost of an individual vehicle will decline, then the formation of economics, the "scale effect" in economics will be produced; However, when the investment of resources exceeds a certain level, although the road can continue to maintain a high level of traffic operation at this time, the input of traffic resources has a big waste viewing on the improvement of individual traffic operating conditions, so that the unit cost is more than that in the original smaller scale, thus it forms the so-called production "scale.". From the perspective of the road operation and management, the road as a business object also meets the general characteristics of production systems. Thus, the objectives for improving programs are based on meeting the largest possible number of traffic demand, with less open road space, a shorter passage time of sections, to maintain traffic operating at a higher speed, and shorter traffic delays. That is to say, considering to improve operational efficiency of the road by reducing the resource consumption level of individual vehicles, the funds available, time and space resources of the road, environmental pollution price, and other resources are the most beneficial improvements for the level of traffic operation.

There are many ways to assess the efficiency. In 1978, based on the concept of relative efficiency, Charnes, Cooper and Rhodes put forward a data envelopment analysis (Data Envelopment Analysis, referred to as DEA) by convex analysis and linear programming. Mathematical programming models are applied to compute and compare relative efficiency of DMUs (Decision Making Units, DMU). DEA can better reflect evaluation objects own information and characteristics, because of full consideration to the optimal input-output program of DMUs. DEA also has the original analysis of multi-input and multiple-output of a complex system (Banker, Das and Datar, 1989).

Information Collection. In the running peak of the highway, in order to enhance the quality of traffic operation, tt is considered to increase transport supply and lower the standard of the speed limit. Because of the many evaluation indicators and the complexity between indicators, and the lack of judgment evidence, it is hard to make sure which control programs are implemented to achieve the optimization of resources. This article is based on operating efficiency of the road and evaluates

improvement programs of traffic with the collection information on road sections. In accordance with the actual operation of the road, this paper considers the assessment of a three-lane freeway after variable lane control and speed control. The original data are in Table 1.

Correlation Inspection. From correlation analysis of the original data in Table 1 leads to Table 2. To consider the relevance of time and average speed as high as 0.99, time index is eliminated, in order to reduce repeat coverage between variables. After the screening, observation variables are: traffic lane number, volume, the average speed, average speed and average delay.

Consistency Test. In the evaluation process of the DEA method, it is often difficult to distinguish the types of factors and what kind of factors input variables (cost variable) are; and what kind of factors are the output variables (efficiency variables). Literature describes the necessary and sufficient conditions of the DEA efficient (CCR) is that (X_0, Y_0) is the Pareto efficient solution of a multi-objective

Table 1. Initial variable table.

Test packet (veh/h)	Program serial number	Input indicator			Output indicator		
		Speed limit (km/h)	Observed traffic volume (veh)	Lane number (lanes)	Average speed (km/h)	Travel time (s)	Average delay (s)
1600	DMU1	60	897	2	53.3	645.4	78.1
	DMU2	80	926	2	73.1	473.4	30.3
	DMU3	100	955	2	82.1	428.7	35.4
3200	DMU4	60	1654	2	48.9	700.5	129.9
	DMU5	80	1805	2	70.3	493.6	50.7
	DMU6	100	1825	2	74.1	464	71.3
3200	DMU7	60	1720	3	52.9	655.5	85.6
	DMU8	80	1806	3	73	477.1	28.1
	DMU9	100	1851	3	82.2	427.6	35.4

Table 2. Relevant analysis of initial variables.

Variables	Volume	Lane	Speed limit	Travel speed	Delay	Travel time
Volume	1.000	–0.743	0.107	–0.008	0.182	–0.017
Lane	–0.743	1.000	0.000	0.009	0.011	–0.004
Speed limit	0.107	0.000	1.000	–0.947	–0.679	–0.924
Travel speed	–0.008	0.009	–0.947	1.000	0.871	0.993
Delay	0.182	0.011	–0.679	0.871	1.000	0.897
Travel time	–0.017	–0.004	–0.924	0.993	0.897	1.000

programming problem (VP): $V--\min\left(x_1,...,x_m\ -\ y_1,...,-\ y_s\right), St.\left(X\ ,\ Y\right)\in \mathrm{T}$, of which (Conceicaowt al., 2006)

$$\mathrm{T}=\left\{(X,Y)\,|\sum_{j\in J}X_j\lambda_j\le X,\sum_{j\in J}Y_j\lambda_j\ge Y,\lambda_j\ge 0,\lambda\in J\right\}$$

According to this theorem, the use of input indicators should be the smaller the better and output indicators should be the bigger the better when using DEA to evaluate. However, in the improvement assessment of road operation, there are input/output indicators that are inconsistent with the request of corresponding indicators, such as delay output where is it preferred that output indicators are smaller, because the output indicators are not expected to be smaller. Therefore, we can not simply use the DEA model to integrate these indicators, we must switch part of variables which impact on evaluate effect, and transform the undesired input and output into the desired input and output by using the linear transformation function: the amendment $x_i^{'}=\nu-x_i$, ν is large enough vector, so that the new variable $x_i^{'}$ is positive which participate in the calculation of operating efficiency in DEA model.

Model Construction

The efficiency measure of traffic operation is expressed as:

$$\theta_i=\frac{v_{Ti}T_i+v_{Di}D_i}{u_{Vi}V_i+u_{Li}L_i+u_{Si}S_i}$$

The unit of this variable does not affect the final results. It is a normalized ratio. The efficiency analysis, shown in formula by looking for optimal decision weight (u , v) of DMUs, based on data points establishes a non-parameter envelope frontier using linear programming method, so all observing point would be on or below the surface of the production frontier.

$$\max_{u,v}\quad \left(u'y_i/v'x_i\right)=\theta_i=\frac{v_{Ti}T_i+v_{Di}D_i}{u_{Vi}V_i+u_{Li}L_i+u_{Si}S_i}$$

$$st\qquad \frac{v_{Tj}T_j+v_{Dj}D_j}{u_{Vj}V_j+u_{Lj}L_j+u_{Sj}S_j}\le 1,\qquad j=1,2,\cdots,n$$

$$u,v\ge 0$$

Using the Charnes-Cooper change, through introducing slack variable s^+ and remaining variables s^- in dual program inequality constraints, the variables will be changed into equality constraints.

$$\min\ \theta_i$$

$$st\quad \sum_{j=1}^{n}\lambda_j\begin{pmatrix}C_j\\L_j\\S_j\end{pmatrix}+\begin{pmatrix}s_{C_i}^-\\s_{L_i}^-\\s_{S_i}^-\end{pmatrix}=\theta_i\begin{pmatrix}C_i\\L_i\\S_i\end{pmatrix}$$

$$\sum_{j=1}^{n}\lambda_j\begin{pmatrix}T_j\\D_j\end{pmatrix}-\begin{pmatrix}s_{T_i}^+\\s_{D_i}^+\end{pmatrix}=\theta_i\begin{pmatrix}T_i\\D_i\end{pmatrix}$$

$$\lambda_j\ge 0,\ j=1,2,\ldots n$$

$$\mathrm{s}^+\ge 0, s^-\ge 0$$

In the above formulations, x_{kj} is the i input of the DMUj, and y_{rj} is the r output of the DMUj. θ_i , the relative efficiency of the DMUi, can be obtained by the above formulation**.** The direction of optimization is determined by using λ_j to determine returns to scale of DMUs and θ^* to determine efficiency value, combined with slack variable value of s-* and s+*. The settlement result is shown in Table 3.

The Analysis of Results

Optimization of Effective Program.In the DEA model, when the decision-making unit optimal solution θ*=1 meets the needs of all s–*= 0, s+*=0 (each slack variable is zero), DEA is valid and vice versa. At this point, in order to keep the output y_j of decision-making unit j unchanged, the various components of x_j can not reduce pro-rata, and even some inputs can not be reduced. From the production theory standpoint, it is shown that the current decision-making unit j is effective in technology and scale.

If θ* <1, it shows that the output unchanged with less input of decision-making unit j, this indicates the current decision-making unit i is not a valid production activity, and its technology is neither effective nor efficient scale. The smaller θ* is, the poorer its effectiveness. According to the ranking of the relative efficiency value, the optimization program for different orders of magnitude of traffic volume is obtained, as shown in Table 3.

Table 3. Solving model results.

DMU	Score	Rank	Excess Volume S–(1)	Excess Lane S–(2)	Excess Speed Limit S–(3)	Shortage Travel Speed S+(1)	Shortage Delay S+(2)
DMU1	0.361	11.00	0.000	0.000	0.034	0.283	0.000
DMU2	0.791	7.00	0.000	0.000	0.000	0.260	0.000
DMU3	**0.901**	**5.00**	**0.000**	**0.000**	**0.000**	**0.227**	**0.000**
DMU4	0.010	15.00	0.000	0.000	0.010	0.000	0.007
DMU5	**1.000**	**1.00**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
DMU6	**1.000**	**1.00**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
DMU7	0.271	12.00	0.000	0.000	0.000	0.232	0.000
DMU8	0.695	10.00	0.000	0.000	0.000	0.309	0.000
DMU9	0.758	9.00	0.000	0.000	0.000	0.203	0.000
DMU10	0.169	13.00	0.000	0.000	0.033	0.110	0.000
DMU11	0.789	8.00	0.000	0.000	0.000	0.169	0.000
DMU12	**1.000**	**1.00**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
DMU13	0.017	14.00	0.000	0.000	0.007	0.014	0.000
DMU14	**1.000**	**1.00**	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
DMU15	0.822	6.00	0.000	0.000	0.000	0.105	0.000

In this article, all of the three implementations for 1600 veh/h can not reach a request of the first measuring value. Therefore, the DMU3, nearest to the first measuring value, is taken as the preferred program. For six kinds of programs for 3200 veh/h, both DMU5 and DMU6 meet the requirements of the first measuring value, whose merits and demerits of the program are difficult to obtain. The frequency for different decision-making units to build a frontier has to be considered.

The MDU who builds frontier more frequently has relative advantages. According to the result above, DMU5 builds a frontier seven times; while DMU6 builds a frontier five times. Therefore DMU5, with a speed limit of 80 km/h and two lanes open, is considered to be more competitive**.**

Diagnosis of Non-effective Program. To analyze and diagnose a non-DEA effective program is to contrast different integrated efficiency indices of DEA and analyze the impact of various indicators on the decision-making units by adjusting the input and output index system (Seiford and Thrall, 1990). Indicators, which affect the program significantly, are made clear, and then they are analyzed specifically. So that, in the implementation process, the effect of the program is optimized (Seiford and Zhu, 2002).

Given D is used as the index system for evaluation, D_i as the index system for evaluation without index i, $\theta(D)$ as the efficiency measure under D and $\theta(D_i)$ as the efficiency measure under D_i, it is obvious that $\theta(D) > \theta(D_i)$. Efficiency rating index $S_j(i)$ is defined below:

$$S_j(i) = \frac{\theta_j(D) - \theta j(D_i)}{\theta_j(D)}, j = 1, 2, ..., n$$

If $S_{j_0}(i) = \max_{1 \le j \le n} S_j(i)$, efficiency rating index most greatly improved after adding index i to DMUj. If i is an input variable, the formula indicates its advantages in use of index i relative to other decision-making units or maybe a poor channel of this resource; while if i is a output variables, it indicates the unit j has obvious advantages on index i.

To analyze and diagnose information in this paper, different combinations of input and output indicators draw a variety of programs. Comprehensive evaluation of the efficiency index is computed by using a C2R model.

From the analysis of Table 4, we can learn:

- The analysis of most influencing elements: For DMUs: $S_3(3) = \max_{1 \le j \le n} S_j(3)$ for the input variables indicates that DMU3 have an advantage in the application of speed limit; while $S_1(1) = \max_{1 \le j \le n} S_j(1)$ for the output variables indicates that DMU1 have an advantage in the application of delay.

Table 4. Efficiency measurement of different combined indicators.

DMU	The value of the efficiency of traffic operation after excluding the following variables				
	Volume	Lane	Speed limit	Travel speed	Delay
DMU1	0.001	0.001	0.361	0.361	0.1
DMU2	0.002	0.003	0.705	0.791	0.529
DMU3	1	0.327	0.79	0.901	0.734
DMU4	0	0	0.01	0	0.01
DMU5	0.002	0.002	1	1	0.836
DMU6	0.761	0.415	1	1	1
DMU7	0	0.001	0.27	0.271	0.076
DMU8	0.002	0.003	0.644	0.695	0.449
DMU9	1	0.6	0.674	0.758	0.63
DMU10	0	0	0.169	0.169	0.049
DMU11	0.002	0.003	0.751	0.789	0.542
DMU12	0.845	1	1	1	1
DMU13	0	0	0.017	0.017	0
DMU14	0.002	0.002	1	1	0.743
DMU15	0.896	1	0.765	0.822	0.736

- The analysis of the programs main influence factor: For the decision-making units, traffic volume and control lane are the main elements to affect road traffic operation. This maintains consistency with the intuitive understanding of tendency to increase traffic supply in large traffic volume. On the output indicators, delay variables vary significantly when they are eliminated; while the speed of change is opposite. This indicates that in the analysis of operating efficiency, the average speed for efficiency analysis remained relatively stable, but there is greater flexibility in terms of delays.

Conclusion

Data envelopment method is used to evaluate improvement program for the traffic operation, make up for the shortcomings of traditional assessment methods, the methods make full use of a variety of data, highlight the objectivity and comprehensiveness of the evaluation, and analyze intrinsic link among elements. The optimum program can effectively reduce the resource consumption of vehicles.

References

Chen, C. (2002). *Freeway Performance Measurement System* (PeMS), University of California, Berkeley, 9-52.

National Research Council (2000). "Highway Capacity Manual—HCM 2000", *Transportation Research Board*, 1780-1802.

Charnes, A.,Cooper, W. W.,and Rhodes, E.(1978). "Measuring the efficiency of decision making units", *European journal of operational research*, 2(4), 429 - 444

Banker, R. D., Das, S., and Datar, S. M. (1989). *Analysis of Cost Variances for Management Control in Hospitals*, Research in Governmental and Nonprofit Accounting, 268-291.

Anersen, P., and Petersen, N. C. A. (1993). "Procedure for ranking efficient unit in data envelopment analysis." *Management Science*, 1261-1264.

Conceicao, A. M., Portela, S., and Thanassoulis, E. (2006). "Comparative efficiency analysis of Portuguese bank branches." *European Journal of Operational Research*, 1-5.

Seiford, L. M, and Thrall, R. M. (1990). "Recent developments in DEA. The mathematical programming approach to frontier analysis." *Journal of Econometrics*, 7-38.

Seiford, L.M., and Zhu, J. (2002). "Modeling undesirable factors in efficiency evaluation European." *Journal of Operational Research*, 16–20.

A Distributed Shortest-Path Routing Algorithm for Transportation Systems

Bin Zou, Jianming Hu[1], Qi Wang and Gangkai Ke

[1] Associate Professor, Department of Automation, Tsinghua University, Beijing, China 100084; E-mail: hujm@mail.tsinghua.edu.cn

Abstract

This paper presents an algorithm based on Dijkstra's algorithm and the distributed Bellman-Ford algorithm (DBF) for path planning in transportation systems, where distributed algorithms are needed to work along with Wireless Sensor Networks. In our algorithm, each node in the network runs Dijkstra's algorithm on the topology it perceived, which we distinguish from the whole topology by giving the name the perceived map, to construct its shortest-path tree, and through exchanging messages with its neighbors it can renew the perceived map. As every node keeps receiving messages for a predefined time and link-costs in a traffic environment often vary moderately, this algorithm can perform well when more than one link changes its cost at the same time.

Key Words: distributed shortest path algorithm; Dijkstra algorithm; bellman-ford algorithm; wireless sensor networks

Introduction

Wireless sensor networks (WSN) can now be deployed into a transportation network, to collect and respond to environmental information, using multiple sensing elements like routing algorithm (Akkaya, 2005), data gathering, data aggregation and data/value fusion and so forth. At the same time, the shortest-path finding problem is one of the most usual and important problems in a transportation system, for instance, in traffic assignment algorithms. With the advent of WSN, distributed routing algorithms are going to perform more efficiently than centralized algorithms, as in

distributed algorithms all nodes are involved in calculating, while in global algorithms calculations are often done at one site, demanding high performance of computing systems. Apart from that, in a distributed algorithm, each node directly uses the information it collects as input, rather than propagates it many hops to some other node, making dynamic shortest paths possible. In (Cicerone, 2010), the dynamic version of the distributed all-pair shortest paths problem has been studied.

Zhan (1998) compared several shortest path algorithms by using real road networks, and found out that the Dijkstra's algorithm ranked among the fastest algorithms in both one-to-all and one-to-one problems. And shortest path algorithms in transportation based on Dijkstra's algorithm can be found in (Wu, 2003), (Pan, 2005) and (Li, 2005). Meanwhile, some parallel shortest path algorithms are mentioned in (Ni, 2006).

We give a routing algorithm here that would fit the traffic environment, where link cost does not make drastic changes in a very short time, thus rendering the shortest paths computed nearly dynamic. As the algorithm runs in a distributed way, it would function well with WSN.

Network Model and Routing Algorithms

The environment for the algorithm is a directed network with constant number of nodes, where link costs can fluctuate over time. Each node is a computing unit that is capable of obtaining the cost of links immediately attached to it, carrying out calculations, communicating with nodes (neighbors) within some range, and also has a local memory. Oblivious to non-neighboring nodes, each node has to communicate with its neighbors to get some information about the whole network.

The Dijkstra's algorithm is one of the most famous shortest-path finding algorithms. But being a global algorithm, it cannot run directly in a distributed way. Many distributed shortest-path algorithms are based on the well-known distributed Bellman-Ford algorithm (DBF), but all of them have to deal with the bouncing effect and counting-to-infinity problems of DBF, such as algorithms in (Cheng, 1989; Humblet, 1991; Murthy 1994; Rajagopalan, 1991).

Combining the Dijkstra's algorithm with DBF, the algorithm we propose, which uses Dijkstra's algorithm as a local shortest path routing algorithm while the whole network works in the manner of DBF, is distributed, yet does not suffer from the problems as DBF does.

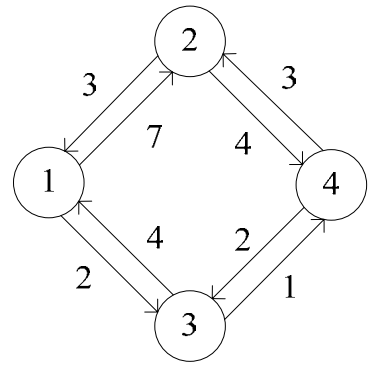


Figure 1. A simple directed network for demonstrating the algorithm.

Algorithm Description

The algorithm in (Cheng, 1989) keeps each node maintaining a distance table, which is comprised of routing tables of its neighbors plus a link-vector, and from which each node obtains its new routing table. And whenever a node is informed of a change in the shortest path from a neighbor to some destination (which means a change in the neighbor's routing table), it renews the corresponding item in its distance table and thus updates its routing table at the same time.

Table 1. The perceived map is denoted by a 4-by-4 weight connection matrix, with rows of matrices corresponding to different loops and columns to different nodes.

	$PMAP_1$				$PMAP_2$				$PMAP_3$				$PMAP_4$			
	-	7	2	-	-	-	-	-	-	-	-	-	-	-	-	-
PMAP(0)	-	-	-	-	3	-	-	4	-	-	-	-	-	-	-	-
	-	-	-	-	-	-	-	-	4	-	-	1	-	-	-	-
	-	-	-	-	-	-	-	-	-	-	-	-	-	3	2	-
	-	7	2	-	-	7	2	-	-	7	2	-	-	-	-	-
PMAP(1)	3	-	-	4	3	-	-	4	-	-	-	-	3	-	-	4
	4	-	-	1	-	-	-	-	4	-	-	1	4	-	-	1
	-	-	-	-	-	3	2	-	-	3	2	-	-	3	2	-
	-	7	2	-	-	7	2	-	-	7	2	-	-	7	2	-
PMAP(2)	3	-	-	4	3	-	-	4	3	-	-	4	3	-	-	4
	4	-	-	1	4	-	-	1	4	-	-	1	4	-	-	1
	-	3	2	-	-	3	2	-	-	3	2	-	-	3	2	-
	-	7	2	-	-	7	2	-	-	7	2	-	-	7	2	-
PMAP(3)	3	-	-	4	3	-	-	4	3	-	-	4	3	-	-	4
	4	-	-	1	4	-	-	1	4	-	-	1	4	-	-	1
	-	3	2	-	-	3	2	-	-	3	2	-	-	3	2	-

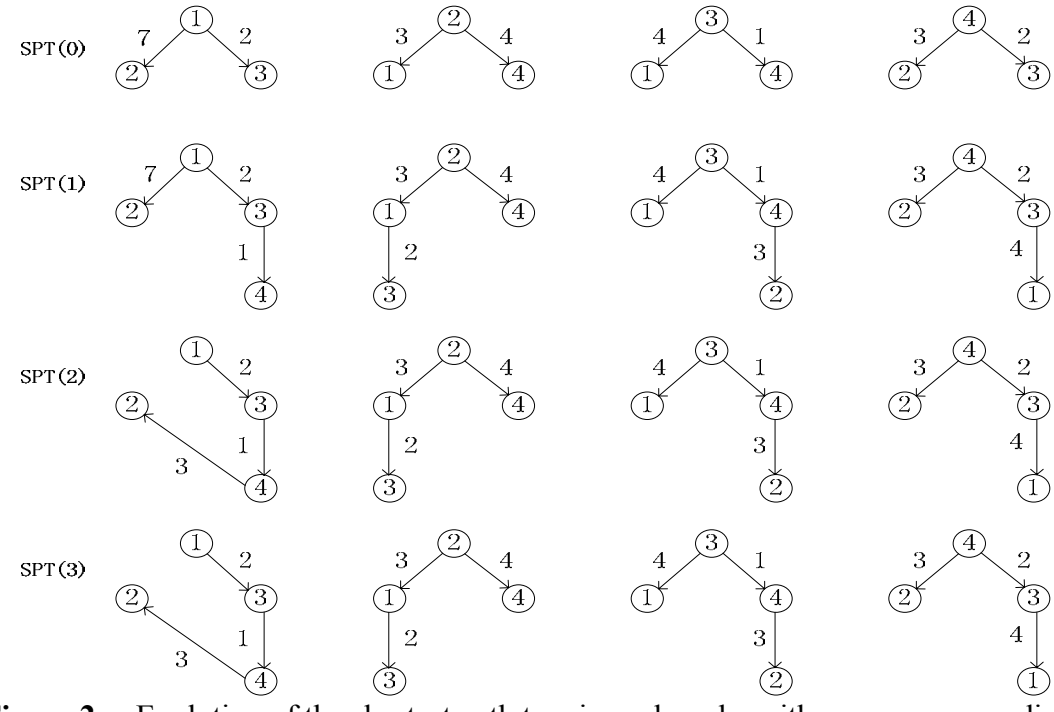


Figure 2. Evolution of the shortest path tree in each node, with rows corresponding to loops and columns to nodes.

Code 1:
begin

- construct the perceived map (PMAP) with links directly attached and send the whole perceived map to neighbors
- keep receiving messages from neighbors for a predefined time *T*
- *if* no links are detected to have two cost-versions
- *then* go back to line 2
- update the current PMAP (PMAP(c)) to a new version PMAP(n)
- run Dijkstra's algorithm on PMAP(n) to get the new SPT (SPT(n))
- *if* SPT(n) is not the same as the previous SPT (SPT(c))
- *then* send to neighbors a message containing links with two cost-versions and those existing in either SPT(n) or SPT(c)
- go back to line 2

end

In (Murthy, 1994), more efficiency is gained by making the node renew additional items corresponding to shortest paths which, sourced at other neighbors, include the known changed shortest path. That trick reminds us that the change in path is aroused by change in some link, so if what a node receives is the information about the link rather than the path, all shortest paths implied in the distance table can be renewed, promisingly resulting much more efficiency.

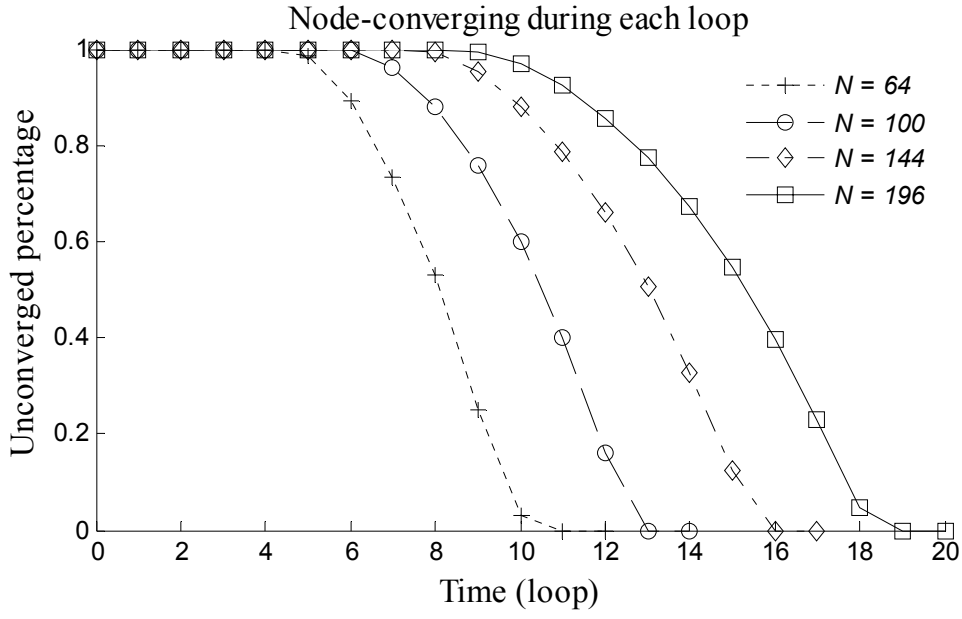


Figure 3. Ordinate indicates the percentage of nodes which have not yet converged.

To avoid the counting-to-infinite problem, each node always sends out only information about links that has newly joined in or removed from the routing table, and those that have introduced a change in the cost of some shortest path originated at it. So, the distance table is replaced by the perceived map which is constructed and renewed with link-information. And as each node will be acquainted with its neighbors' routing tables after communicating with them, the algorithm will converge as well as the one in (Cheng 1989).

Given a network in Figure 1, for each node, we give the involved operations in Code 1 and results for each loop in Table 1 and Figure 2. As for the whole network, if none of the nodes change their SPTs, which means that no messages will be produced, then the algorithm is thought to have converged.

Simulation Results

To gain some insight into the performance of the algorithm over networks of different scales, we have the algorithm run on a network four times, each time including different number of nodes, where nodes within the bound of the network are all 4-indegreed and 4-outdegreed. The results are shown in Figure 3, which are compared with the output of a function in Matlab 2009a running with the same data. The horizontal ordinate is the time axis ticked in loops. *N* is the number of nodes included in the network.

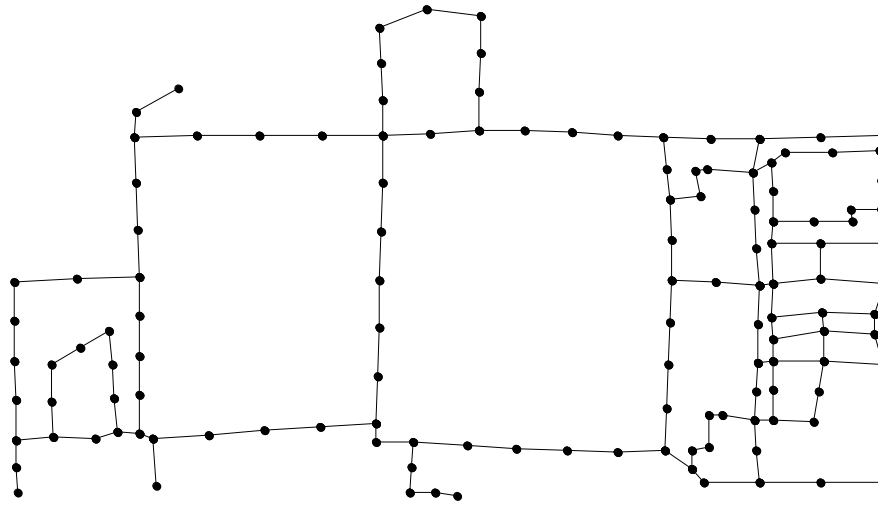


Figure 4. A network extracted from Beijing city map. Nodes are located to make sure that distances to neighbors are less than 500 m.

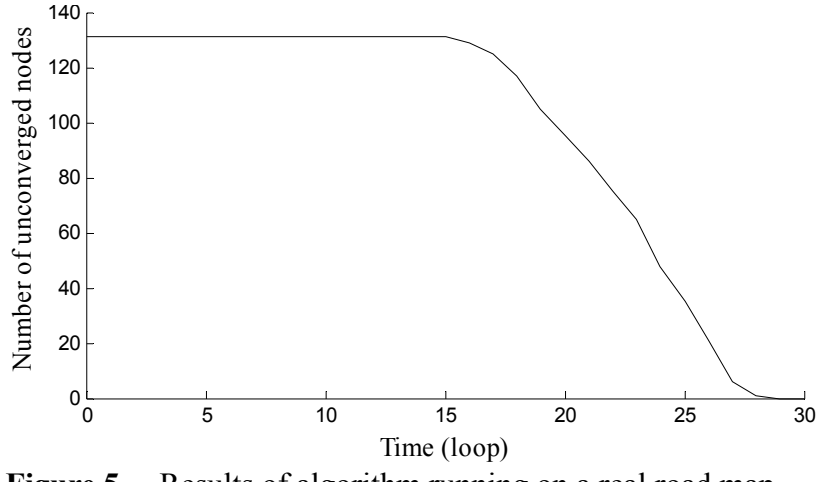


Figure 5. Results of algorithm running on a real road map.

The algorithm is conducted on a 131-node network (Figure 4) extracted from Beijing city map, to show its performance on real network. Results are shown in Figure 5.

Conclusions

In this paper, we present a shortest path routing algorithm for transportation systems where WSNs are deployed (thus a distributed algorithm is needed), and link costs do not change drastically, by taking the Dijkstra's algorithm as a local algorithm for calculating shortest paths under a distributed global framework. The algorithm has been applied to networks with different nodes and a real road network extracted from city map of Beijing, and it has shown nice convergence. Future work will be focused on examining the speeds of this algorithm on real road networks with traffic flow data considered to see if it performs well in a more realistic world.

Acknowledgement

The work described in the paper is partially supported by the National Natural Science Foundation of China (NSFC) (No. 50708054, 60834001), the Hi-Tech Research and Development Program of China (No. 2007AA11Z222), the National Basic Research Program of China (No. 2006CB705506), and the National Key Technology Re-search and Development Program (No. 2006BAJ18B02).

References

Akkaya, K., and Younis, M. (2005). "A survey on routing protocols for wireless sensor networks." *Ad Hoc Networks*, 325-249.

Cheng, C., Riley, R., Kumar, S. P. R., and Garcia-Luna-Aceves, J. J. (1989). "A loop-free extended Bellman-Ford routing protocol without bouncing effect." *Symposium proceedings on Communications architectures & protocols*, 224-236.

Cicerone, S., D'Angelo, G., Stefano, G. D., and Frigioni, D. (2010). "Partially dynamic efficient algorithms for distributed shortest paths." *Theoretical Computer Science*, 1013-1037.

Humblet, P. A. (1991). "Another adaptive shortest-path algorithm." *IEEE Trans. Comm.*, 995-1003.

Li, X., and Wang, J. (2005). "The shortest path search based on city road of database." *Journal of Southwest Jiaotong University*, 14-15, 21.

Murthy, S., and Garcia-Luna-Aceves, J. J. (1994). "A more efficient path-finding algorithm." http://eprints.kfupm.edu.sa/42402/.

Ni, A., Juan, Z., and Gao, L. (2006). "An overview of research on parallel shortest path algorithm in transportation network." *Journal of Highway and Transportation Research and Development*, 128-131.

Pan, F., Wang, F., Zou, X., and Sun, G. (2005). "Shortest path searching based on city road database." *Journal of Highway and Transportation Research and Development*, 105-107.

Rajagopalan, B., and Faiman, M. (1991). "A responsive distributed shortest-path routing algorithm within autonomous systems." *Internetworking: Research and Experience*, 51-69.

Wu, B., Li, L., and Lei, X. (2003). "Shortest path searching based on city road database." *Journal of Southwest Jiaotong University*, 80-83.

Zhan, F. B., and Noon, C. E. (1998). "Shortest path algorithms: an evaluation using real road networks." *Transportation Science*, 65-73.

Multi-Constrained Shortest Path Model and Solution with Improved Ant Colony Algorithm

Yaomin Hu[1], Yuyin Liu[2] and Weiming Liu[3]

[1]Ph.D, [2]Ph.D, [3]Prof., Candidate, School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510006, Guangdong, China, Email: huyaomin@sina.com, yuyin@sina.com, weimingliu@163.com

Abstract

How to provide a route to meet the driver's multiple psychological expectations is the key problem of a navigation system. Essentially, this problem is a resource constrained shortest path problem (RCSP), which belongs to NP hard problems and can not be solved with the traditional shortest path algorithm. A multi-constrained shortest path mathematical model is presented in this paper and solved with an improved ant colony algorithm. The pheromone update rule and heuristic factor are redesigned in the algorithm. Results show that the improved optimization algorithm has a good ability to accurately determine the multi-constrained shortest path in the road network.

Key Words: multi-constrained shortest path; ant colony algorithm; route plan

Introduction

Using the navigation system to plan a route, drivers hope that the route will save time and money, easy to drive, secure and scenic, that is, the route should meet multiple quality constraints. In recent years, many scholars examined the quality of the planned route of the navigation system. Lorraine McGinty et al. (2000) examined the concept of "route quality," but considered that route quality is difficult to define, and proposed a case-based route plan. Yu and Nie(2009) examined the concept of "Reliable route" and argued that the high quality route

should guarantee that a driver can arrive at their destination on time. Choi and et al. (2007) mentioned the concept of "driver preference," and stated that the best route varied with drivers' preferences. Jørgensen et al. (2004) discussed the relationship between the cost and path length, and provided a relationship model of the minimum cost and the best quality of route. However, scholars tried to turn the different dimensional properties into a unified non-dimensional property, and used the value of this non-dimensional property to measure quality of route, and then transformed the multi-constrained shortest path problem to shortest path problem, and finally used the traditional dijstra or A * algorithm to solve it. In essence, this problem is a resource constrained shortest path problem (RCSP). The resource constrained shortest path problem is to find the shortest path in a network such that certain constraints are satisfied. RCSP is NP-hard (Handler and Zang 1980; Wang and Crowcroft 1996) which always can be solved using a simulated annealing algorithm, genetic algorithm, heuristic search algorithm, or Hopfield neural network algorithm (Gao and Shang 2003, Gao and Xu 2003). RCSP can not be solved using dijstra or A* algorithm. Handler and Zang(1980) proposed two methods to solve the CSP problem. One requires solving a k-Shortest Path (kSP) problem, and the other is based upon Lagrangian relaxation, which is the most classical for solving the RCSP but has no satisfying performance. Ant colony algorithm is one of the most recent approximate optimization methods to be developed (Mullen et al. 2009). This algorithm is inspired by the behavior of real ants in the wild (Dorigo et al. 1991), and more specifically, by the indirect communication between ants within the colony via the secretion of chemical pheromones. Ant colony algorithm is a kind of distributed computing, information positive feedback, and heuristic searching optimization tool, and thus it is applied in the optimal path search in transportation with great potential and adaptability.

The rest of this paper is organized as follows: Section 2 will provide a multi-constrained shortest path mathematical model. Section 3 will improve the original ant colony algorithm to adapt to solve the multi-constrained shortest path problem. Section 4 will discuss experiments to test the capability of searching a satisfied path in the road network of the improved algorithm. Section 5 will provide a summary and conclusion.

Mathematical Formulation of the Multi-Constrained Shortest Path

The objective of this paper is to solve the problem to find a shortest path in the road network with multiple path quality constraints (such as the arrival time constraints, security constraints, distance constraints, cost constraints, the road grade constraints and so on) satisfied. Consider a directed graph $N = (V,E)$ where $V = \{1, 2, \ldots, n\}$ represents the set of nodes and $E = \{(i, j): i,j \in V, i \neq j\}$ represents

the set of m directed arcs. Non-negative weights, such as c_{ij} , t_{ij}, and f_{ij} are associated with each arc(***i, j***). These weights may represent, for example, the length (or cost), the travel time, and tolls associated with each arc. Label the origin node, 1 and the destination node, n. There are some other symbols defined as follows:

C: road network directed graph matrix;

c_{ij}: the arc of from node ***i*** to node ***j***, $c_{ij} \in C$;

o: represents an origin;

d: represents a destination;

p: path connecting ***o*** and ***d*** with multi-constraints met;

d_{ij}: the length of road c_{ij};

f_{ij}: road tolls of c_{ij}, different types of vehicles cast different tolls;

t_{ij}: time consumed for each kilometer when driving on the c_{ij} , reference to history average statistics;

s_{ij}: the rate of traffic accidents occurring on the c_{ij} , reference to history average statistics; and

X_{ij}: if c_{ij} on the path connecting o and d , X_{ij} is 1, otherwise 0.

With the travel time, tolls, security, and other aspects of quality constraints imposed, the multi-constrained shortest path can be described as the following mathematical model:

$$\begin{aligned} & Min(G) = \min(\sum_{i\in V}\sum_{j\in V}(X_{ij}\bullet \mathrm{d}_{ij}) \\ & s.t.\begin{cases} \sum_{i\in V}\sum_{j\in V}(X_{ij}\bullet \mathrm{d}_{ij}\bullet \mathrm{t}_{ij}) \le D_p \\ \sum_{i\in V}\sum_{j\in V}(X_{ij}\bullet \mathrm{f}_{ij}) < F_p \\ \prod_{i\in V, j\in V}(1-X_{ij}\bullet \mathrm{s}_{ij}) \ge S_p \\ X_{ij} \in \{1,0\}, i, j \in V \end{cases} \end{aligned} \tag{1}$$

In the model, $\sum_{i\in V}\sum_{j\in V}(X_{ij}\bullet \mathrm{d}_{ij})$ represents length of path, $\sum_{i\in V}\sum_{j\in V}(X_{ij}\bullet \mathrm{d}_{ij}\bullet \mathrm{t}_{ij})$ represents total time consumed on the path, $\sum_{i\in V}\sum_{j\in V}(X_{ij}\bullet \mathrm{f}_{ij})$ represents total tolls on the path, and $\prod_{i\in V, j\in V}(1-X_{ij}\bullet \mathrm{s}_{ij})$ represents the rate that won't cause traffic accidents when driven on the path. D_p is the maximum time allowed to spend on the selected path ***p***, S_p is the minimum safety factor for the selected path ***p***, F_p is the maximum cost allowed for the selected path ***p***.

Improvement of Ant Colony Algorithm for Solving the Model

The original ant system. Ant algorithms are considered under the category of swarm intelligence (Bonabeau et al., 1999). Swarm intelligence encompasses the implementation of intelligent multi-agent systems that are based on the behavior of real-world insect swarms as a problem solving tool. Ant algorithms are considered

under the category of swarm intelligence (Bonabeau et al., 1999). The first ant algorithm, named "Ant System" (AS), was developed in the 1996 by Dorigo et al. As a test-bed for this algorithm, the well-known benchmark Travelling Salesman Problem (TSP) was used for a set of M nodes. The TSP problem involves finding the shortest length closed tour visiting each node only once. In other words finding the shortest-path to visit each node once, ending up back at the starting node. In the case of the Euclidean TSP we consider the path length between any given nodes i and j to be the Euclidean distance between i and j, such that the path length $d_{ij} = [(x_i - x_j)^2 + (y_i - y_j)^2]^{1/2}$. Artificial ants search the solution space, simulating real ants looking for food in the environment. The objective function values correspond to the quality of food sources. The time is discrete in the artificial ants environment. At the beginning of the search process (time $t = 0$), the ants are located in different nodes. The pheromone value on edge (i, j) at time t is denoted $\tau_{ij}(t)$. At time $t = 0$, the value of $\tau_{ij}(0)$ is equal to a constant c. At time t each ant is moving from the current node to the next node, and reaches the next node at time $(t + 1)$. When at node i, ant k choose node j with the transition probability $p_{ij}^k(t)$ defined by the following Eq. 2:

$$P_{ij}^k(t) = \begin{cases} \frac{[\tau_{ij}(t)]^\alpha \bullet [\eta_{ij}(t)]^\beta}{\sum_j [\tau_{ij}(t)]^\alpha \bullet [\eta_{ij}(t)]^\beta}, \text{ j} \in allowed_k \\ 0, otherwise \end{cases} \quad (2)$$

where $allowed_k$ is the set of feasible nodes to be visited by ant k (the set of feasible nodes is updated for each ant after each step), $\eta_{ij} = 1/d_{ij}$ is a heuristic value, which is the "visibility" of node j. Furthermore, α and β are parameters representing relative importance of the pheromone value and the heuristic value, respectively. The greater the value of α, the greater the probability is to select the "important" edge, which was passed by many ants. The greater the value of β, the greater the probability is to select the closest node. Every ant will complete a traveling salesman tour after n iterations. Dorigo et al. (1996, 2006) proposed to update the trail intensity $\tau_{ij}(t)$ after each cycle as follows::

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \rho \bullet \Delta\tau_{ij} \quad (3)$$

Where ρ is the coefficient ($0 < \rho < 1$) called evaporation rate. The total increase of pheromone at edge (i, j) after a complete cycle is equal to:

$$\Delta\tau_{ij}(t) = \sum_{k=1}^{n} \Delta\tau_{ij}^k(t) \quad (4)$$

where $\Delta \tau_{ij}^{k}(t)$ is the pheromone value laid on edge *(i, j)* by the *k*th ant during the cycle. $\Delta \tau_{ij}^{k}(t)$ is calculated as $\Delta \tau_{ij}^{k}(t) = \frac{Q}{L_k(t)}$, if the *k*th ant walks along the edge*(i, j)* in its tour during the cycle. Otherwise, the pheromone quantity equals: $\Delta \tau_{ij}^{k}(t) = 0$, where Q is a constant; $L_k(t)$ is the tour length developed by the *k*th ant within the cycle. As we can see, artificial ants collaborate among themselves in order to discover high-quality solutions. This collaboration is expressed through pheromone deposition.

Algorithm Improvement for the Problem

Ant transition rule. In the new algorithm, being located in node i, ant k chooses the next node j to be visited at time t with the transition probability $p_{ij}^{k}(t)$ defined by the following equation:

If $p_{ij}^{k}(t) < P_0$, transition probability defined as Eq. 5, p_0 is transition probability threshold, Eq. 5 can accelerate the algorithm convergence：

$$P_{ij}^{k}(t) = \begin{cases} \max(\tau_{ij}(t)), \mathrm{j} \in allowed_k \\ 0, otherwise \end{cases} \tag{5}$$

If $p_{ij}^{k}(t) \geq p_0$ transition probability defined as Eq. 6：

$$P_{ij}^{k}(t) = \begin{cases} \frac{[\tau_{ij}(t)]^{\alpha} \bullet [\eta_{ij}(t)]^{\beta}}{\sum_{j} [\tau_{ij}(t)]^{\alpha} \bullet [\eta_{ij}(t)]^{\beta}}, \mathrm{j} \in allowed_k \\ 0, otherwise \end{cases} \tag{6}$$

$\boldsymbol{\eta}_{ij}$ in Eq. 6 is not $1/ d_{ij}$ used in original ant system, As this would cause ants to covet current step faster and deviate from the direction to destination. In order to speed up algorithm convergence, $\boldsymbol{\eta}_{ij}$ should be a heuristic factor for ant direction navigation, and be defined as Eq. 7:

$$\eta_{ij} = \begin{cases} 1/dis_{jd}, dis_{jd} \neq 0 \\ 1, otherwise \end{cases} \tag{7}$$

The dis_{jd} is the Euclidean distance from node j to the destination (node d). In this way ants can move to the destination in a higher probability, rather than covet current step faster.

Pheromone update rule. In order to induce ants in a larger probability to move closer to the path that met the constraints, once a tour is completed, the ant lays a

pheromone trail along each edge visited in the tour. After a cycle, the pheromone values are updated according to Eq. 8:

$$\tau_{ij}(t+1)=\begin{cases}(1-\rho)\tau_{ij}(t), c_{ij}\notin p\\(1-\rho)\tau_{ij}(t)+\rho\bullet F, c_{ij}\in p\end{cases} \tag{8}$$

p represents the current best path, ***ρ*** is the global pheromone evaporation parameter, ***F*** is pheromone update function, which is defined as Eq. 9:

$$F=-A\bullet F_1+B\bullet F_2+C\bullet F_3+D\bullet F_4 \tag{9}$$

F1, F2, F3, and *F4* in Eq. 9, defined respectively as (10), (11), (12), (13) :

$$F_1=\sum_{i\in V}\sum_{j\in V}(X_{ij}\bullet \mathrm{d}_{ij}) \tag{10}$$

In Eq. 10, ***F1*** means the length of path ***p***.

$$F_2=\begin{cases}D_p-\sum_{i\in V}\sum_{j\in V}(X_{ij}\bullet \mathrm{d}_{ij}\bullet \mathrm{t}_{ij}),\ D_p>\sum_{i\in V}\sum_{j\in V}(X_{ij}\bullet \mathrm{d}_{ij}\bullet \mathrm{t}_{ij})\\0,\ D_p<\sum_{i\in V}\sum_{j\in V}(X_{ij}\bullet \mathrm{d}_{ij}\bullet \mathrm{t}_{ij})\end{cases} \tag{11}$$

In Eq. 11, F_2 means path ***p*** should satisfy time constraint.

$$F_3=\begin{cases}S_p-\prod_{i\in V,j\in V}(1-X_{ij}\bullet \mathrm{s}_{ij}),\ S_p>\prod_{i\in V,j\in V}(1-X_{ij}\bullet \mathrm{s}_{ij})\\0,\ S_p<\prod_{i\in V,j\in V}(1-X_{ij}\bullet \mathrm{s}_{ij})\end{cases} \tag{12}$$

In Eq. 12, F_3 means path ***p*** should satisfy security constraint.

$$F_4=\begin{cases}F_p-\sum_{i\in V}\sum_{j\in V}(X_{ij}\bullet \mathrm{f}_{ij}),\ S_p>\sum_{i\in V}\sum_{j\in V}(X_{ij}\bullet \mathrm{f}_{ij})\\0,\ F_p<\sum_{i\in V}\sum_{j\in V}(X_{ij}\bullet \mathrm{f}_{ij})\end{cases} \tag{13}$$

In Eq. 13, F_4 means path ***p*** should satisfy tolls constraint.

A, B, C, and ***D*** are four coefficients that are positive real numbers, represent respectively the relative importance of the four constraints. If the algorithm trend to reinforce some constraint, the coefficient can take a larger number. According to the above definition, the shorter length of the chosen path met all of constraints the more pheromone will deposit on it, which inspires more ants to converge to these edges on the path.

Experiments

In order to verify the ability to search for the optimal path of the algorithm, the authors programmed based on the improved algorithm to solve high-quality multi-constrained path models using Windows XP and Visual C++. Figure 1 is the network topology used in the simulation, which is an un-directed graph representing vehicles that can pass through the edge of bi-directionality. A two-dimensional vector (***x,y***) represents the coordinates of the vertex, which is

used to calculate the distance to destination. 4-tuple (*length, speed, tolls, accident-rate*) indicates the length, speed limitation, tolls, the probability of accident statistics of each edge.

Experiment 1. Let the program plan the optimal path between node 1 and node 14, and add the following quality constraints on the path: allow journey time $\boldsymbol{D}_p$ <4 hours, allow the accident rate (1-$\boldsymbol{S}_p$) <= 10^{-5}, ***tolls*** <100. Set the number of ants $\boldsymbol{m}$ = 10, the pheromone heuristic factor $\boldsymbol{\alpha}$=1, distance heuristic factor $\boldsymbol{\beta}$=5, initial value of pheromone $\boldsymbol{Q}$=100, pheromone evaporation factor $\boldsymbol{\rho}$ = 0.55, coefficient $\boldsymbol{A}$ = 10, $\boldsymbol{B}$ = 15, $\boldsymbol{C}$ = 200, $\boldsymbol{D}$ = 15, the number of iterations set to 100 times. Table 1 shows the results after 300 experiments.

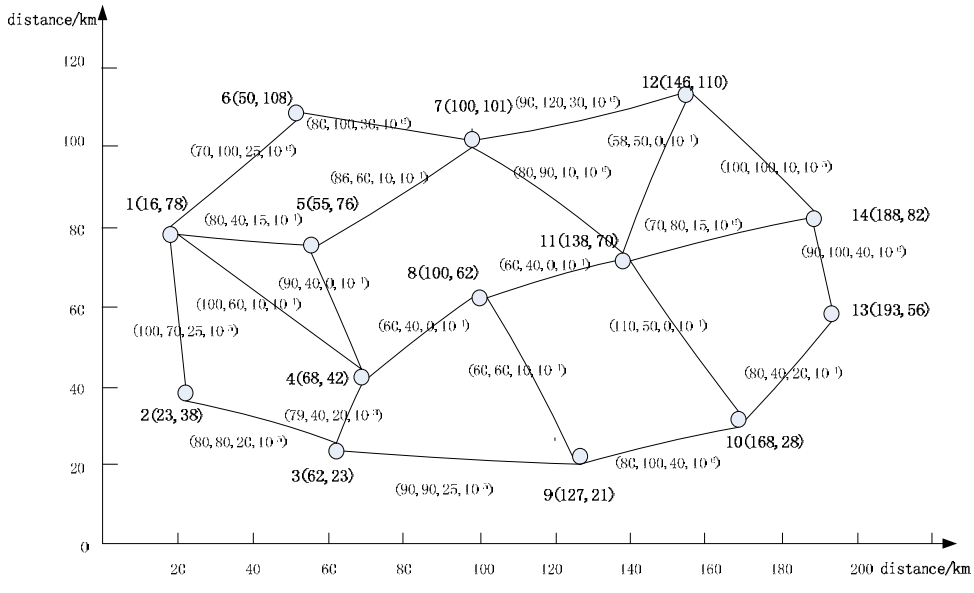


Figure 1. Network topology used in simulation.

Table 1. Table of path planning results 1.

(OD)	Path taken	Tolls	Length	Time estimated	Accident rate	Result times	Rate
1-14	1-6-7-11-14	80	300	3.265	10^{-5}	266	88.67%
	1-6-7-12-14	95	340	3.625	10^{-5}	34	11.33%

Table 2. Table of path planning results 2.

(OD)	Path	Tolls	Length	Time estimated	Accident rate	Result times	Rate
1-9	1-4-8-9	20	220	4.167	10^{-4}	282	94%
	1-2-3-9	70	270	3.43	10^{-5}	18	6%

Experiment 2. In order to examine the behavior when the algorithm can not provide a path to meet expected quality constraints of the driver, let the program plan the optimal path between node 1 and node 9, and add the following quality constraints on the path: allow journey time $\boldsymbol{D}_p$ <3 hours, allow the accident rate (1-$\boldsymbol{S}_p$) <= 10^{-5}, ***tolls*** <5. Set the number of ants $\boldsymbol{m}$ = 10, the pheromone heuristic factor $\boldsymbol{\alpha}$=1, distance heuristic factor $\boldsymbol{\beta}$=5, initial value of pheromone $\boldsymbol{Q}$=100, pheromone evaporation factor $\boldsymbol{\rho}$=0.55, coefficient $\boldsymbol{A}$=10, $\boldsymbol{B}$ = 15, $\boldsymbol{C}$=200, $\boldsymbol{D}$=15, the number of iterations set to 100 times. Table 2 shows the results after 300 experiments.

Experiment 1 indicated that the optimal paths and sub-optimal paths were found respectively in the rate of 88.67% and 11.3%, and that there was no invalid path, which indicated the improved ant colony algorithm has good optimization ability. From Figure 1, it also can be seen that the path 1-4-8-11-14 of routing from node 1 to node 14 is not found by the program. Though the tolls of this path was 25, and the distance was just 290, the time to be consumed was up to 5.55 hours, and the accident rate merely was $10^{-4.}$ The path quality could not meet the requirement, the algorithm did not select it.

Experiment 2 indicated that the algorithm had found the route 1-4-8-9 and 1-2-3-9 when there were not any routes that could meet the quality constraints required by drivers. It was because that algorithm had deteriorated to the shortest distance path planning algorithm. It also indicated that the algorithm was robust and existed output at any time.

Conclusion

A mathematical model of multi-constrained shortest path plan problem of navigation system was created, and the ant colony algorithm was used to solve it, which provided a feasible solution for this NP-C hard problem. The basic ant colony algorithm was advanced for solving the TSP problem. If it was applied to solve the multi-constrained shortest path problem, the algorithm should be modified. The pheromone update rule of the algorithm was re-designed, and the path quality constraints were added to the pheromone update rule. The visibility heuristic factor was improved, and positive feedback was enhanced, so the convergence speed of the algorithm was improved. Simulation results indicated that optimal path or suboptimal path can be obtained by the algorithm quickly and effectively when the quality constraints can be met, and that the shortest or a sub-short path can be found by the algorithm when quality requirements of the path could not be satisfied.

References

McGinty, L., and Smyth, B. (2000). "Personalised route planning: a case-based approach[M]." E. Blanzieri and L. Portinale (eds.), EWCBR, 431-443.

Nie, Y., and Wu, X. (2009). "Shortest path problem considering on-time arrival probability[J]." *Transportation Research Part B: Methodological*, 597-613

Choi, W-K, Kim, S-J, and Kang T-G. (2007). "Study on method of route choice problem based on user preference[M]." B. Apolloni et al. (eds.): KES 2007/ WIRN 2007, Part III, LNAI, 645-652.

Jørgensen, F., and Pedersen, P. A. (2004). "Travel distance and optimal transport policy." *Transportation Research Part B: Methodological*, 415-430.

Handler, G. Y., and Zang, I. (1980). "A dual algorithm for the constrained shortest path problem[J]." *Networks*, 293-310.

Mullen, R. J. Monekosso, D. Barman, S., and Remagnino, P. (2009). "A review of ant algorithms[J]." *Expert Systems with Applications ,* 9608-9617.

Colorm, A. Dorigo, M., and Miniezzo, V. (1991) "Distributed optimization by ant colonies." Proceeding *of the First European Conference on Artificial Life*. Paris France: Elsevier Publishing, 134-142.

Bonabeau, B. Dorigo, M., and Theraulaz, G. (1999). "Swarm intelligence: from natural to artificial systems." *Conference on Artificial Life*, New York, Oxford University Press,156-163.

Dorigo, M., and Gambardella, L. M. (1997). "Ant colony system: a cooperating learning approach to the travelling salesman problem." *IEEE Transactions on Evolutionary Computation*, 53-66.
Dorigo, M., and Di Caro, G. (1999). "The ant colony optimization metaheuristic: new ideas in optimization[J]." *IEEE Transactions on Evolutionary Computation*,11-32.

Wang, Z., and Crowcroft, J. (1996). "Quality of service routing for supporting multimedia applications." *IEEE Journal on Selected Areas in Communications*, 1228-1234.

Dorigo, M., Birattari, M., and Stutzle, T. (2006). "Artificial ants as a computational

intelligence technique[J].” *IEEE Computational Intelligence Magazine,265-276.*

Dorigo, M., and Gambardella, L. M. (1996). “A study of some properties of Ant-Q.” InPPSN IV: *Proceedings of the 4th international conference on parallel problem solving from nature*, 656-665.

Gao, L., and Xu, J. (2003). “A DNA algorithm for graph vertex coloring problem[J].” *ACTA ELECTRONICA SINICA*, 494-497.

Gao, S., and Shang (2003). “Solving TSP with simulated annealing algorithm[J].” *Journal of East China Shipbuilding Institute (Natural Science Edition)*, 13-16.

Congestion Pricing Model Based on Cumulative Prospect Theory

Yuyin Liu, Weiming Liu, Yaomin Hu and Jie Li

Doctotral Candidate, School of Civil Engineering and Transportation, South China University of Technology, Guangzhou, Guangdong Province, China 510640, E-mail: liuyuyin@126.com

Abstract

As an important measure of traffic demand management, congestion pricing can guide and regulate traffic flow to achieve the purpose of easing the traffic congestion. Therefore, it has been paid more and more attention of scholars. Congestion pricing methods are the core and foundation of congestion pricing theory, and the rationality of the pricing model is directly related to the scientific congestion pricing strategies. In the traditional congestion pricing model, it doesn't correspond to reality, if a traveler is assumed absolutely rational, but the decision-making process can be exactly described by the cumulative prospect theory in the uncertain situation, and it is consistent with the bounded rationality of traveler character. In this paper, on the basis of cumulative prospect theory and Wardrop user equilibrium principle, the psychological effect factors of travelers are brought into the congestion pricing model that well overcome the defects of the traditional pricing model, analyses and prove the feasibility of the model at the same time.

Key Words: congestion pricing, pricing model, perceive value, cumulative prospect theory

Introduction

Congestion pricing methods are the core and foundation of congestion pricing theory, and the rationality of the pricing model is directly related to the scientific congestion pricing strategies. The traditional congestion pricing model which based on the expected utility maximization, assume that all the travelers are absolutely rational. However, once some complicated factors are introduced into the decision-making environment, there is obvious departure between the actual traveler choice behavior and the expected utility theory. Such as Allais and Ellsberg paradox, they all illustrate the significant departure between the actual travel behavior and the

"expected utility maximization" hypothesis. With the development of the behavioral sciences, traveler's bounded rationality increasingly attracts more and more attention of domestic and foreign experts. In 1979, Kahneman and Tversky proposed the Prospect theory which can well solve this problem on the basis of the bounded rationality. In 1992, KT formed the cumulative prospect theory after absorbing rank-dependent utility thought proposed by Quiggin.

In the existing literature some experts have done much research and achieved fruitful results on the applications of cumulative prospect theory in transportation. Katsikopoulos etc. have found that, in the path choice experiments, if average travel time below a certain reference, travelers follow "risk aversion"; On the contrary, follow "risk seeking". Bogers have also found such a similar phenomenon in their survey: When a traveler randomly choose a shorter route with a large uncertainty, or a longer route with a steady traffic circumstance, they also show risk aversion behavior, in other words, most of travelers would choose the longer route to travel. Younes Hamdouch etc. extend the toll pricing framework previously developed for vehicular traffic networks to ones with the potential to include many modes of transportation such as walking, driving, and using public conveyance. In this paper, based on these achievements, with cumulative prospect theory established the perceive value for travelers and congestion pricing model combined with UE.

Cumulative prospect theory

In the practical road network systems, road network is a complex dynamic system affected by many external factors. The route travel time is a random variable, under such condition, the travel cost and its probability distribution can be the basic influencing factors. For the calculation convenience, we assume that travelers want to maximize the utility when they choose the travel route.

Each feasible path corresponds to an actual travel disutility variable x and its probability function p. According to the prospect theory, travelers are bounded rationality under uncertainty travel time, and their choice behavior based on the perceive value of their travel route which gains from their subjective utility and probability.

- Subjective utility $v(u)$: According to the actual route utilities, travelers form the subjective utility, representing that the actual level of the utility has an impact on the traveler's psychology factors.
- Subjective probability $w(p)$: According to the actual occurrence probability of route utility, travelers form their subjective occurrence probability, representing that the actual utility probability function p has an impact on the perception utility.

Subjective Utility. The reference point is used for building subjective utility function, in which the carrier of utility is the change of utility, not the terminal state of utility. According to the reference point, subjective utility of traveller can be divided into two parts, gains and losses.The shape of the value function leads to risk aversion for

gains and risk seeking for losses. Loss aversion is accommodated through a kink at zero (the reference point). The corresponding two-part power function is:

$$v(x) = \begin{cases} (u-u_0)^{\alpha} & u \geq u_0 \\ -\lambda(-u+u_0)^{\beta} & u < u_0 \end{cases} \quad (1)$$

Where $v(u)$ stands for the value; u for the outcome; α and β are estimable coefficients determining the convexity/concavity of the function; the level of traveler preference to risk aversion is given by λ, $\lambda \geq 1$ corresponding to the tenet of prospect theory that the value function is steeper for losses than for gains. α and β should fall between 0 and 1 and λ be larger than 1 to accommodate the phenomena violating the expected utility theory axioms and to resemble the shape in Figure.1. Tversky and Kahneman report median values of $\alpha = \beta = 0.88$ and $\lambda = 2.25$.

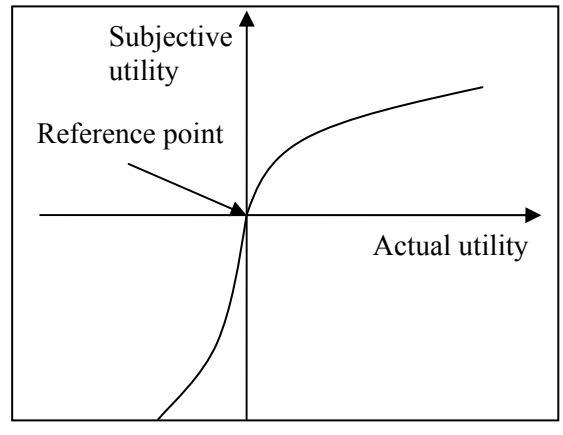


Figure 1. Relationship between perceive value and real value.

Subjective Probability. As making decisions under uncertain conditions, travelers drew their conclusions by actual probability. The subjective probability of travelers would be far away from the real probability because of the uncertain information mixed with subjectivity. Perceive probability is the sense of the objective laws only existed in the travelers' brain, which is based on the analysis of real circumstances, integrating their own experience and prospect. Based on experimental evidence, prospect theory proposes that small probabilities are typically over-weighted and moderate and high probabilities are under-weighted. However, this transformation could also be proposed as a synthesis of experienced and preconceived probabilities in the mind of the decision maker, or an internal representation based on full knowledge of the true probability distribution function of outcomes.

The choice modeling literature provides examples of transformations of the outcome probabilities that have been validated experimentally. Tversky and Kahneman propose two inversely *S*-shaped formulations for the weighting function:

$$w^{+}(p) = \frac{p^{\gamma}}{(p^{\gamma} + (1-p)^{\gamma})^{1/\gamma}} \quad (2)$$

$$w^{-}(p)=\frac{p^{\varphi}}{(p^{\varphi}+(1-p)^{\varphi})^{1/\varphi}} \tag{3}$$

Where w^{+} and w^{-} indicate the decision weights for gains and losses, respectively; p is the probability level; γ and φ estimable parameters indicating the level of distortion in probability judgment. Tversky and Kahneman argued that γ and φ should fall between 0 and 1; their own experiments suggested median values of γ = 0.69 and φ = 0.61.

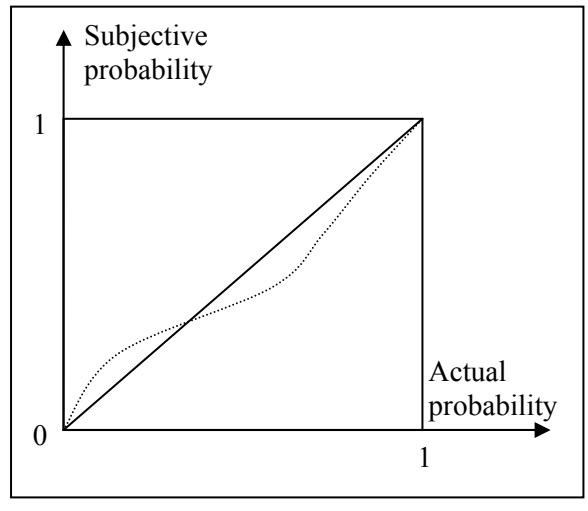


Figure 2. Relationship between perceive probability and real probability.

Perceive Value. This paper assumes that the prospect value of path is finite discrete distribution, shown by u_i, with p_i is the occurrence probability. Based on cumulative prospect theory, perceive value of the route is given by

$$v=v^{+}+v^{-} \tag{4}$$

$$v^{+}=\sum_{i}\pi_i^{+}v(u_i) \tag{5}$$

$$v^{-}=\sum_{i}\pi_i^{-}v(u_i) \tag{6}$$

Where v^{+} is the influence on decision maker made by gains, v^{-} is the influence made by losses, π is the weighting function, a comprehensive value of subjective probability function. The probability weight π_i^{+} for a positive result is the discrepancy between the volume of the events with outcome which is at least as good as event u_i and the events with outcome exactly better than event u_i; the probability weight π_i^{-} for a negative result is the discrepancy between the volume of the events with outcome which is at least as bad as event u_i and the events with outcome exactly worse than event u_i, which can be obtained by the following formula.

$$\pi_n^{+}=w^{+}(p_n);\ \pi_{-m}^{-}=w^{-}(p_{-m})$$

$$\pi_i^{+}=w^{+}(p_i+\cdots+p_n)-w^{+}(p_{i+1}+\cdots+p_n),\quad 0\le i<n$$

$$\pi_i^{-}=w^{-}(p_{-m}+\cdots+p_i)-w^{-}(p_{-m}+\cdots+p_{i-1}),\quad -m<i<0$$

Route Perceive Value Modelling

The object of congest charging is mainly cars; therefore, this paper does not consider the ways of non-motor vehicles' travel, merely taking two ways into account. They are the way of personal travel which is signified by m=1, and the way of public travel signified by m=2, besides, the two ways are independent with each other.
Generalized cost function that relates travel cost with traffic conditions is the basis of making congestion charging model. Several elements aspects must be considered for building route perceive value.
(1) Trip time t_m
Trip time is the time that traveler spend from origin to destination, which is related with road, vehicle and traffic states. Trip time plays an important role in traveler choosing the route.
(2)Trip cost B_m
For private cares, trip cost consists of operation cost and extra congest price, etc.
(3) Individual choice of traveler ε_m.

ε_m is the influence value that road safety, comfort and individual preference impact on traveller to choose m^{th} travel model. We can transform discomfort into extra travel time in the period of jam. Referring to available data, the impedance caused by discomfort could convert into 50% travel time cost.
We represent the road network by a directed graph consisting (G) of N links labeled a=1, 2… N; a demand vector q, with entries q_{rs} representing the travel demand on the r-s origin-destination (OD) movement; and a set of paths k_{rs} connects the r-s OD movement. An assignment of flows to all paths is denoted by the vector f, with f_{rs}^k the flow on the k^{th} path connecting r-s OD. Other parameter and variable is presented as follows:
q_{rs}^m -The traffic volume of m^{th} traffic mode on the r-s OD movement.
q_a^m -The traffic volume of m^{th} traffic mode on the road a.
l_a -The length of road a.
η -The toll rate (Yuan/km) of congestion price, η_a=0 if the road a without toll.
$C_a^m(q_a^m)$ -Road resistance of m^{th} travel model on road a.
μ_{rs}^m -The maximum travel disutility on the r-s OD movement.
N_{ml} -Time value, which related to national income per capita.
δ_{rs}^k -Associative variable between road a and path k. If road a belongs to path k, δ_{rs}^k=1, otherwise, δ_{rs}^k=0.

Perceive Value Modelling for Private Travel Mode. The path-resistance of private travel mode includes travel time, delayed time, operation cost, congestion toll and individual choice of traveler. On the road a, expected generalized travel cost of private travel mode is,

$$C_a^1(q_a^1) = t(q_a^1)N_{ml} + Y_a^1 + y_a^1 l_a + \varepsilon_1 = 1.5t(q_a^1)N_{ml} + Y_a^1 + y_a^1 l_a \quad (7)$$

Where $t(q_a^1)$ is the BPR function that express travel time by the travel flow and road capacity.

$$t(q_a^1) = t_a^0[1 + \varsigma(q_a^1 / Cap_a)^{\psi}]$$

Where t_a^0 is travel time of free flow, Cap_a is actual capacity of road a, ς and ψ are the fixed parameters, ς =0.15, ψ =4. This function takes the path-resistance made by traffic volume and converted time cost of ε_1 into account.

The second part is operation cost,

$$Y_a^1 = l_a \cdot Opr_1$$

Where Opr_1 is operation cost of private travel mode for unite distance (Yuan/km).

The third part is congestion toll on route a.

So, (7) could be expressed as,

$$C_a^1(q_a^1) = 1.5t_a^0[1 + \varsigma(q_a^1 / Cpa_a)^{\psi}]N_{ml} + l_a \cdot Opr_1 + y_a^1 l_a \tag{8}$$

The path-resistance of the *r*-*s* OD movement should be the accumulative total of the impedance on each road of path. Therefore, the path-resistance function of private travel mode on path *k* could be,

$$C_{rs,k}^1(q_k^1) = \sum_a C_a^1(q_a^1)\delta_{rs}^k = \sum_a \left\{ 1.5t_a^0[1 + \varsigma(\frac{q_a^1}{Cap_a})^{\psi}]N_{ml} + l_a \cdot Opr_1 + y_a^1 l_a \right\} \delta_{rs}^k \tag{9}$$

The u_0^1 signify the maximum gain that can be acquired by the traveller at the destination, in that way, the gain acquired by the traveller at the destination when selecting the path *k* is

$$u_i^1(q_k^1) = u_0^1 - C_{rs,k}^1(q_k^1) = u_0^1 - \sum_a C_a^1(q_a^1)\delta_{rs}^k \tag{10}$$

The distribution of $u_i^1(q_k^1)$ relates to the probability distribution of the road section transit time t_a.Suppose the probability of the utility $u_i^1(q_k^1)$ obtained by the travellers is p_k^1, with the substitution of (4), the perceive value of the path *k* is,

$$v_k^1(q_k^1) = \sum_i \pi_{i,1}^+ v(u_i^1(q_k^1)) + \sum_i \pi_{i,1}^- v(u_i^1(q_k^1)) \tag{11}$$

Where $\pi_{i,1}^+, \pi_{i,1}^-$ are the weight factor of the probability p_i^1, and $v(u_i^1(q_k^1))$ is the subjective utility of the expected utility for the path *k*.

Perceive Value Modelling for Public Travel Mode. The road resistance of the public transportation travellers includes travel time, ticket price and value of preference. Suppose the average number of passengers of the public travel mode is N_u, and then the formula for the road resistance function can be

$$C_a^2(q_a^2) = t(q_a^2)N_{ml} + Y_a^2 / N_u + \varepsilon_2$$

So, if the travellers who take the public travel mode choose route *k*, the expected value can be,

$$u_i^2(q_k^2) = u_0^2 - C_{rs,k}^2(q_k^2) = u_0^2 - \sum_a C_a^2(q_a^2)\delta_{rs}^k$$

Let be the probability of the utility u_i^2 obtained by the travellers is p_i^2, and then the perceive value for the public travel mode of the route *k* is,

$$v_k^2(q_k^2) = \sum_i \pi_{i,2}^+ v(u_i^2(q_k^2)) + \sum_i \pi_{i,2}^- v(u_i^2(q_k^2)) \tag{12}$$

Where $\pi_{i,2}^+, \pi_{i,2}^-$ are the weight factors of the probability p_i^2, and $v(u_i^2(q_k^2))$ is the subjective utility of the expected utility on the route *k*.

Road Congestion Pricing Rate Model

Principle of the route choice under the congestion pricing. People always choose the travel route according to the principle of the minimum travel expense, which means the maximize travel disutility, so the travel behaviour generally follows the UE principle. According to the principle of UE, under the equilibrium state, the utilities of all the routes used are equal and are not less than the utility that have not been used. For that reason, to the travelers who choose the different travel routes, this principle can be described as

$$v_k - \mu_{rs} \begin{cases} = 0, & f_{rs}^k > 0 \\ \le 0, & f_{rs}^k = 0 \end{cases} \tag{13}$$

Where v_k is the perceive value of the route *k* between r-s OD under the equilibrium state. The formula above mentioned shows that, when there is travel flow on the route *k* between r-s OD, perceive value of the path *k* is equal to the maximize path perceive value; if there isn't travel flow on the route k, perceive value of the path *k* is less than or equal to the maximize path perceive value.

Model building. The model building in this paper can not only make the allocation proportion of the travel demand among different mode satisfies the Logit model, but also can make route choice behaviour of different travel mode satisfy user equilibrium principle of Wardrop independently. In fact, it means to build a model which combined the travel mode splitting based on the road network equilibrium distribution and distribution of traffic flow under the traffic congestion pricing conditions.

Shao Chunfu (2006) structures a combined mathematic model that including public transportation network, with each travel mode is independent with each other. The model is presented as follow,

$$\min Z(x, \hat{x}, \hat{q}) = \sum_{rs} \int_0^{\hat{q}_{rs}} \left(\frac{1}{\theta} \ln \frac{w}{\overline{q}_{rs} - w} + \phi_{rs}\right) dw + \sum_a \int_0^{x_a} t_a(w) dw + \sum_b \int_0^{\hat{x}_b} \hat{t}_b(w) dw \tag{14.a}$$

$$s.t. \quad \sum_k f_k^{rs} + \hat{q}_{rs} = \overline{q}_{rs}, \ \forall r, s \tag{14.b}$$

$$\sum_l \hat{f}_l^{rs} = \hat{q}_{rs}, \ \forall r, s \tag{14.c}$$

$$\hat{f}_l^{rs}, f_k^{rs} \ge 0, \ \forall k, l, r, s \tag{14.d}$$

$$\hat{q}_{rs} \ge 0, \ \forall r, s \tag{14.e}$$

$$x_a = \sum_r \sum_s \sum_k f_k^{rs} \delta_{a,k}^{rs}, \ \forall a \tag{14.f}$$

$$\hat{x}_b = \sum_r \sum_s \sum_k \hat{f}_l^{rs} \delta_{b,l}^{rs}, \forall b \tag{14.g}$$

Where ϕ_{rs} denote the influence which is brought by other factors to the travel mode chosen besides the travel time. The greater of ϕ_{rs}, the more partial of the user to choose the personal travel mode in the same situation. The objective of this paper is to find out the resistance of congestion pricing, leaving ϕ_{rs} out of consideration. Because of the travel resistance of different travel mode, under the equilibrium state, adopting the balanced constraint of the entire road network is inapposite.

Therefore according to above analysis, a congestion pricing rate model can be concluded eventually in the equilibrium network.

$$\max Z(q_k^1, q_k^2, q_{rs}^1, q_{rs}^2) = \sum_k \int_0^{q_k^2} v_k^2(w)dw + \sum_k \int_0^{q_k^1} v_k^1(w)dw - \sum_{rs} \int_0^{q_{rs}^2} (\frac{1}{\theta} \ln \frac{w}{q_{rs} - w})dw \quad (15.a)$$

$$s.t. \quad q_{rs}^1 + q_{rs}^2 = q_{rs}, \ \forall r,s \quad (15.b)$$

$$\sum_k q_{rs,k}^1 = q_{rs}^1, \ \forall r,s \quad (15.c)$$

$$\sum_k q_{rs,k}^2 = q_{rs}^2, \ \forall r,s \quad (15.d)$$

$$q_{rs,k}^i \geq 0, \forall r,s,k,i \quad (15.e)$$

$$q_{rs}^i \geq 0, \forall r,s,i \quad (15.f)$$

$$q_a^1 = \sum_r \sum_s \sum_k q_{rs}^1 \delta_{rs}^k, \ \forall a \quad (15.g)$$

$$q_a^2 = \sum_r \sum_s \sum_k q_{rs}^2 \delta_{rs}^k, \ \forall a \quad (15.h)$$

Adopting the objective function mentioned above mentioned is to build a maximal model which equivalent to the equilibrium state of mixed-traffic. The constraints (15.b)-(15.d) describe the equation on the traffic flow of all traffic modes and the traffic flow of all paths; the constraints (15.g)-(15.h) describe the equation of the traffic flow on road-segments and the corresponding path.

calculation example analysis

As shown in Figure 3, assume that there are two paths between r-s OD, including 4 road and 4 nodes. Road length and the practical capacity are present in Table1. Presume the traffic demand between r-s OD is 5000.

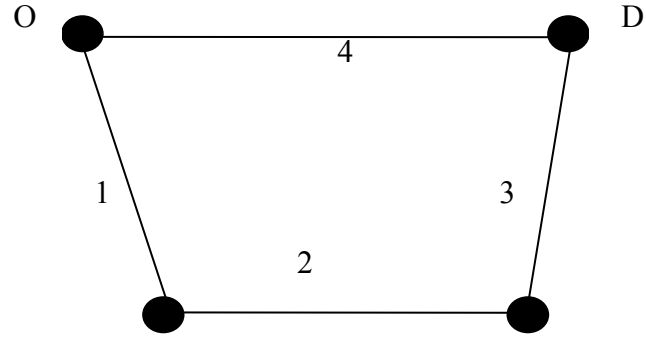


Figure 3. Schematic diagram of the road network to simplify.

Table 1. Road length and the Practical capacity.

Road	1	2	3	4
l_a (km)	0.2	0.6	0.3	0.8
Capacity(veh/h)	2500	2500	2500	2500

Table 2. Model Parameters value.

Parameters	ς	ψ	N_{ml}	Opr_i	θ	N_u	Y_2
Value	0.15	4	6	1.5	1	30	1.0

Table 3. The results of the assignment under Road network free of charge.

Road	1	2	3	4
Trip volume	0	0	0	2992
Impedance	0.25	0.75	0.375	1.0
Saturation	0	0	0	0.87

Table 4. Route choice behaviour changes with the charging rat.

Travel mode	Toll rate changes	Chosen path
Private mode	0-0.6	Path 1
	>0.6	Path 1, 2

The parameters values of model are presented in Table 2. When all roads are free, traveller chose the path and transportation means based on their utility, where traffic modal splitting accord with Logit model. The travel flow of every modal on all roads is present in Table 3.

The result shows that the path-resistance of road 4 is lower than others, travel flow concentrate on this shortest path, and the utilization rate of road (1, 2 and 3) is too low. So, the implementing of congestion pricing on road 4 is necessary.

After implement congestion price on road 4, relevant, traveller changes their travel behaviour.

As shown in Figure 4 and Figure 5, with the increasing of congestion toll rate, the private travel flow on r-s OD movement is decreased, and the public travel flow is increasing. When the toll rate reaches a special point, the distribution of travel flow in road network is invariant. Because of private traveler change their path as toll rate increasing, once the toll rate reaches a special point, all private travelers will choose free way, the change of toll rate is invalidation to the distribution of travel flow. This phenomena is consistent to reality.

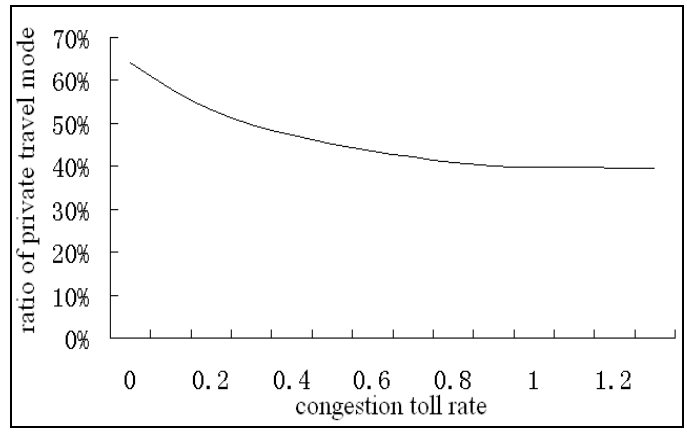


Figure 4. Changes of private travel flow with different congestion toll rates of various travel modes

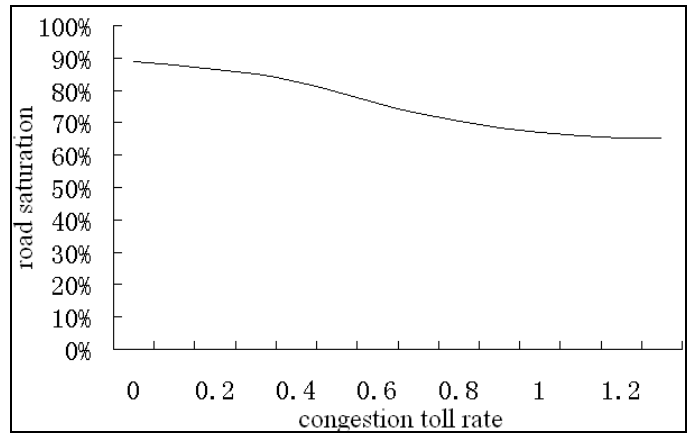


Figure 5. Road saturation with different rates of various transport modes

Conclusion

Congestion pricing is essentially a kind of traffic requirement management measure as well as an important part of city traffic management. The policy can guide people to utilize the existing road traffic facility, thereby alleviate the traffic congestion and lower the traffic pressure. The traditional congestion pricing models are basically established under the assumption of the travelers' absolute rationality, but different kinds of problems arise when applying these models into practice, so it is incapable to well describe the travelers'' response to the congestion pricing.

Cumulative prospect theory appeared in economics domain at the earliest, which was a descriptive model used to simulate the psychological decision process of the investor. This paper goes from the prospective of the traveler's psychological decision process, and utilizes the cumulative prospect theory to optimize the pricing rate model of city traffic congestion pricing, thus fully considers the characteristic of the travelers' limited rationality in the model. Through modeling and cases comparative analysis, we arrives at an conclusion that the cumulative prospect theory

can better describe the routing choice behaviour under the condition of uncertainty, and to some certain extent overcomes the shortcoming of "expectation utility theory".

References

Avineri E. A, "cumulative prospect theory approach to passenger's behavior modeling: waiting time paradox revisited." *Journal of Intelligent Transportation Systems*, vol.8, no.2, pp. 195-204, 2004.

Tim Schwanen, Dick Ettema. "Coping with unreliable transportation when collecting children: Examining parents' behaviour with cumulative prospect theory" *Transportation Reaserch Part A*, vol.43, pp. 511-525, 2009.

Bogers E A I, Zuylen, H J van. "The importance of reliability in route choices in freight transport for various actors on various levels," *Proceeding European Transport Conference 2004.* France: Strasbourg, 2004, 16: 136-154.

Kahneman D, Tversky A. "Prospect theory: an analysis of decisions under risk," *Econometrica*, vol. 47, pp. 313-327, 1979.

Younes Hamdouch, Michael Florian, Donald W. Hearn, Siriphong Lawpongpanich. "Congestion pricing for multi-modal transportation systems" *Transportation Reaserch Part B,* vol.41, pp. 275-291, 2007.

Achim I. Czerny."Airport congestion management under uncertainty"*Transportation Reaserch Part B*, 2009.

Katsikopoulos K V, Fisher D L, Duse Anthony. "Risk attitude reversals in driver's route choice when range of travel time is provided," *Human Factors*, vol. 44, no. 3, pp. 466-467, 2002.

Shao Chunfu, *Traffic planning*, Beijing, China: pp. 227-231, 2006.

Song Gao, Emma Frejinger, Moshe Ben-Akiva. "Adaptive route choices in risky traffic networks: A prospect theory approach" *Transportation Reaserch Part C,* 2009.

Evaluation of Tools for Incorporating Road User Costs in Pavement Rehabilitation Decision Making on Rural Highways in Saskatchewan

Ningning Zheng, Guoxin Liu and Yee-Chuang Jin[1]

[1] Faculty of Engineering and Applied Science University of Regina, 3737 Wascana Parkway Regina, SK, Canada S4S 0A2; Email: andrew.liu@uregina.ca

Abstract

This study is part of an effort to incorporate road user costs and economic impacts into the current asset management system in the network level pavement rehabilitation decision-making process. MicroBENCOST and HDM-4 (Highway Development and Management Model), which are widely used in the road user cost estimation process in North America, are used to calculate user costs, benefits and benefit-cost ratio and results are evaluated. The outputs generated from the two programs are analyzed and compared when the input conditions are consistent for both programs. Several critical input parameters which have a significant influence on the major outputs are identified. In addition, two statistical correlation models based on the outputs of the two programs, which provide a simple and fast way to calculate user benefits, are developed.

Key Words: asset management; pavement rehabilitation; network level decision-making; road user costs; software to estimate road user costs

Introduction

The Canadian province Saskatchewan's economy is largely dependent on exports, and various types of bulk commodity movement are a major part of highway transportation demand. Due to its land-locked location and the widespread of the economic activities, Saskatchewan has a large roadway network and its roadway mileage per capita is among the highest in the world. This means that there are only limited resources available for roadway investment but high requirements for pavement preservation and performance. The transportation policy makers and highway engineers in Saskatchewan have been faced with great challenges to preserve the vast highway network in a most strategic and cost-effective manner to

support the provincial social and economic development. In recent years, as more and more roads have aged and traffic loading on provincial highways has significantly increased due to the abandonment of railway branch lines and the consolidation of the grain handling system, strategic decision making for pavement infrastructure preservation has become a major challenge.

The concept of highway pavement deterioration and preservation can generally be described as Figure 1. Like in many other jurisdictions, it is impossible to provide capital investment to all highways at the same time to meet their rehabilitation needs due to budget constraint. A comprehensive Asset Management System has been implemented in Saskatchewan for more than ten years, which has greatly improved pavement preservation decision-making process. However, due to the absence of social and economic parameters in the Asset Management System, there are questions on whether the best way of investment to preserve the infrastructure is also the best way of investment to support the social and economic development of the province. To address this issue, it is believed that a preservation policy at strategic level to connect the government's overall social and economic development strategy to the Asset Management System tool is necessary. The engineering economic analysis at the pavement network level will also be necessary to help prioritize the pavement preservation projects under the limited budget. The estimation of road user costs and benefits under different road and traffic conditions is an essential step for the network level engineering economic analysis.

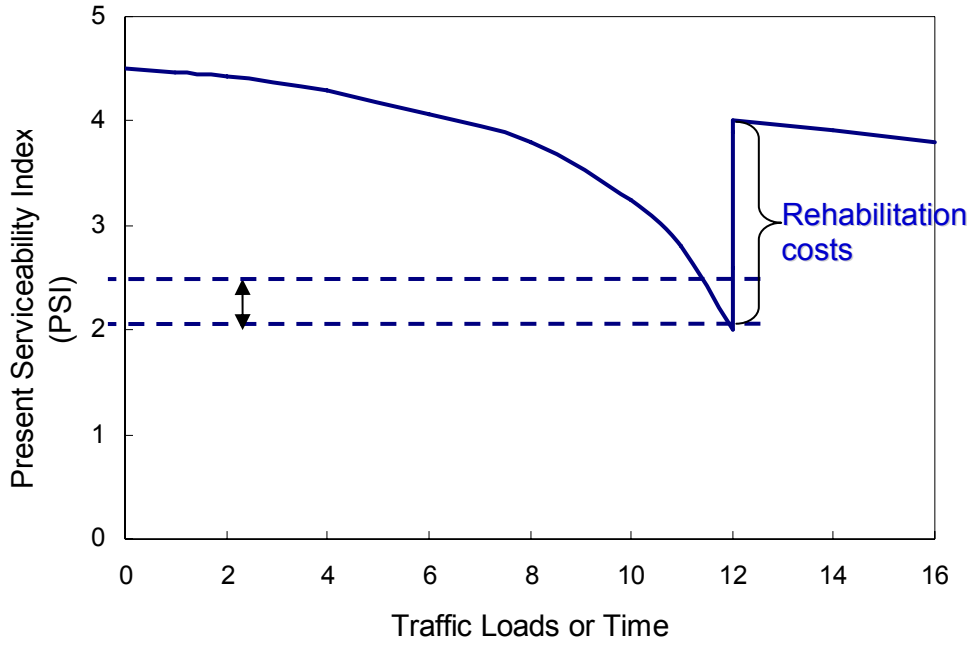


Figure 1. Typical pavement performance curve (Garber and Hoel 2002).

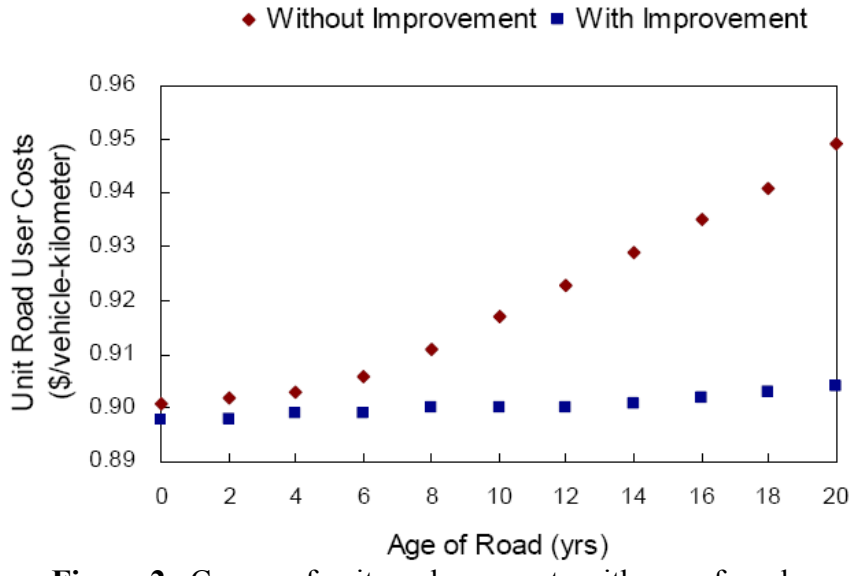


Figure 2. Curves of unit road user costs with age of roads.

In this study, the road user costs include travel time costs, vehicle operating costs and passengers' discomfort costs, as used by the two programs. The user benefits are the savings of road user costs. As shown in Figure 2, the offset between the two scenario curves, without and with road improvement, is the unit user benefits for that year.

Among various software used in estimating road user costs, MicroBENCOST and HDM-4 appear to have wider applications in the North America and the world. However, engineers and other practitioners often just use one of the software to calculate the benefits-cost ratio at the project level and use the result as a reference for their decision making. It is still not clear if the results of the two programs are consistent. Therefore, it is necessary to conduct the comparison between the outputs generated by MicroBENCOST and HDM-4 when inputs for the two programs are consistent.

Literature Review

There have been researchers who used and evaluated the two programs. Roper and Keltner (1999) used MicroBENCOST to estimate road user costs and benefits resulting from a railroad grade separation project. The authors identified some weaknesses associated with MicroBENCOST. Lamptey et al. (2004) gave a comprehensive introduction to MicroBENCOST and HDM-4. The merits and drawbacks of them were identified. The authors also analyzed the existing methods for estimating user costs by the two programs. Reilly et al. (2005) presented a review of existing analytical tools for asset management system. Both MicroBENCOST and HDM-4 were described in details. The strengths and limitations of these two programs were listed in their study.

In Jung et al.'s research (2002), MicroBENCOST and HDM-4 were used to estimate the agency costs and user costs on asphalt rubber pavement and conventional asphalt

pavement, respectively. The author also discussed the drawbacks of two software. Salem and Genaidy (2007) conducted a review of tools including MicroBENCOST and HDM-4, and raised the concern with regard to the lack of consideration of user costs in Life Cycle Cost Analysis (LCCA). The study also identified the reasons for not considering user costs in highway pavement improvement projects. In addition, the author presented the recommendations for the inclusion of user costs into the Ohio Department of Transportation pavement selection process.

According to these literatures, there have been comprehensive evaluation of MicroBENCOST and HDM-4, in terms of methodology and drawbacks, however, there is no literature comparing the results generated by the two programs and it is not clear if the results of the two programs are consistent.

Inputs for MicroBENCOST and HDM-4

The process of determination of inputs was quite time-consuming. Various sources including reports, books, guides, reviews, and websites were searched and evaluated to determine the appropriate inputs. The input data used in this study are based on the most recent data for the low volume rural highway conditions in Saskatchewan and Canada. All of the input parameters required by MicroBENCOST and HDM-4 are presented in Table 1. Some of them have been converted to the appropriate format to better fit the situation of Saskatchewan. Details of all the inputs data are discussed in the study (Zheng 2009).

Among all inputs listed in Table 1, some input parameters have far more significant influence on outputs than others. Some are crucial to the outputs while others just have a slight impact on outputs. Therefore, the critical input parameters which have large impact on outputs were identified. The result of a sensitivity analysis (Callao, 2007) was used to determine the extent of impact of inputs on the unit road user benefits. According to the impact on road user benefits, six input parameters were considered most important to unit road user benefits, these include:

- New vehicle price
- Kilometers driven per year
- Service life (years)
- Maintenance labor costs
- Gross Vehicle Weight, and
- Fuel prices

Based on the above analysis result, and considering the fact that this study is interested in the total road user benefit after a highway section's improvement on rural Saskatchewan highways, it is reasonable to assume that many of the identified critical factors influencing the unit road user benefits are relatively steady over the analysis period. Such factors include new vehicle price, kilometers driven per year, service life (years), maintenance labor costs, and gross vehicle weight. It is also reasonable to assume that the vehicle fleet composition in terms of age of vehicles, vehicle brand, and vehicle types is also relatively steady. It is therefore only those

Table 1. List of Inputs parameters for MicroBENCOST and HDM-4.

MicroBENCOST Input parameters	HDM-4 Input parameters
Analysis period (yrs)	Annual interest (%)
Discount rate (%)	Rise and fall (m/km)
Value of time for passenger cars ($/hr)	Number of rise and fall (per km)
Value of time for trucks ($/hr)	Horizontal curvature (deg/km)
Depreciation costs ($/veh)	Superelevation (%)
Maintenance and repair cost ($/1000km)	Altitude (m)
Lubricating oil costs ($/l)	Speed limit enforcement factor
Tire costs ($/veh)	Roadside friction factor
Gross vehicle weight for trucks (ton)	NMT friction factor
Accident costs ($)	Number of passengers on trucks
Value of time multipliers	Percent time driven on snow (%)
Unit discomfort costs ($/veh-km)	Percent time driven on water (%)
Vehicle occupancy rate for cars	Paved roads texture depth
Hourly traffic distribution (%)	Private use percentage (%)
Fatal, injury and PDO accident rates (per 100 million vehicle-km)	Work-related passenger trips (%)
Lane width and shoulder width (m)	New vehicle prices ($)
Design speed and speed limit (km/h)	Service life (years)
Average running speed (km/h)	Annual working hours (hours)
Emission data	Annual km driven (km)
Annual AADT growth rate (%)	Crew wages ($/hr)
Annual PSI deterioration	Annual overhead ($)
Road section length (km)	Cargo time ($/hr)
Total rehabilitation costs ($)	Passenger non-working-time value ($/hr)
Annual maintenance costs ($)	
Vehicle composition for each class (%)	

input parameters that vary from one highway section to another will be critical to the total road user benefits. These critical input parameters are identified and their impacts on road user benefits are analyzed.

The vehicle fuel prices fluctuated dramatically in the recent years in North America while other inputs do not seem to change significantly in a short period. Therefore, diesel and gasoline prices can be considered as two critical input parameters which have major impact on road user costs. According to the change of diesel and gasoline prices in 2008, the minimum values were set at 1.1 and 1.14 dollars per liter and the upper limits are 2.0 and 2.04 dollars per liter. The analysis interval of 0.3 dollars per liter within the price range was set for this study.

Although the changes of AADT do not affect the unit road user costs when AADT is below 6,500 (Callao, 2007), AADT is definitely a crucial factor affecting total user costs and benefits incurred on the proposed road section. It is obvious that the higher AADT represents more road users who can benefit from the road improvement, and

the total user benefits are increased with the increasing traffic demand. Based on the Saskatchewan Ministry of Transportation and Infrastructure's Traffic Information System, AADTs for rural highways are typically in the range from 400 to 2,400, and an analysis interval of 400 within this range is adopted in this study.

IRI is the International Roughness Index, and a higher IRI value indicates a rougher pavement surface that will have higher road user costs. A rougher pavement surface may also require more expensive treatment for pavement improvements. Normally, highway agencies adopt a threshold value of IRI based on road users' response to the roughness of the pavement surface, and any rougher pavement surface than the threshold will be considered for surface rehabilitation. The improvement of the pavements with worse pavement conditions will achieve higher user benefits than roads that had been kept in good condition. Therefore, IRI value of the pavement surface before improvement should be considered as a critical input parameter. Based on normal highway agency operation, IRI value of 2.5 is mostly used as the threshold, the analysis interval of 0.5 is selected between the minimum IRI value of 2.0 and the maximum IRI value of 4.0 in this study.

Trucking is an important part of transportation on rural highways in Saskatchewan. Higher truck traffic will accelerate the deterioration of pavements, and a rougher pavement will also cause higher road user costs for trucks due to higher vehicle operating cost for trucks and higher value of time for truckers. Therefore, the truck percentage in the traffic stream should be a critical input parameter according to the impact on road user costs. According to Traffic Information System in the Saskatchewan Ministry of Transportation and Infrastructure, the truck percentage for rural highways normally ranges from 10% to 30%. The analysis interval of 10% within this range is used in this study.

The analysis was conducted by using MicroBENCOST and HDM-4 and by varying one critical input parameter while holding other input parameters constant to calculate user benefits and benefit-cost ratios. The value of each critical parameter was changed alternately to guarantee the outputs can be generated under every condition. Therefore, a large number of runs of two softwares were performed.

Output Conversion

It is notable that HDM-4 is able to generate user costs rather than user benefits and benefit-cost ratio. However, for the reason of comparison of the outputs generated by MicroBENCOST and HDM-4, the conversion of user costs to user benefits and benefit-cost ratio is necessary. The equations of derivation of benefit-cost ratio are expressed below:

$$B/C = (B + S) / (CCOST + MCOST) \quad (1)$$

$$B = \sum_{N=1}^{20} \frac{B_N}{(1+i)^N} \quad (2)$$

$$B_N = (U_{1N} - U_{2N}) \times VKT_N \times 365 \quad (3)$$

$$S = \frac{s}{(1+i)^N} \tag{4}$$

$$MCOST = \sum_{N=1}^{20} \frac{MCOST_N}{(1+i)^N} \tag{5}$$

where, B is the total discounted total user benefits; S is the discounted salvage value at the end of the analysis period; CCOST is the construction costs; MCOST is the total discounted maintenance costs; B_N is the user benefits for the N_{th} year; i is the discount rate; U_{1N} and U_{2N} are the unit road user costs for the N_{th} year without and with improvement, respectively; VKT_N is the vehicle-kilometres travelled in the N_{th} year.

In the following assessment, the benefits refer to the road user benefits, and the costs refer to the construction costs, and the IRI value refers to the original IRI value for the road section before improvement.

Comparison of Outputs Generated by MicroBENCOST and HDM-4

Figures 3 - 7 show the relationships between each critical parameter and benefit-cost ratio generated by MicroBENCOST and HDM-4.

In Figure 3 and 4, both programs show benefit-cost ratio increase with the increase of AADT and truck percentage as expected if other inputs are constant. However, according to Figure 5 and 6, the change of fuel price does not seem to significantly affect the benefit-cost ratio if other inputs are constant.

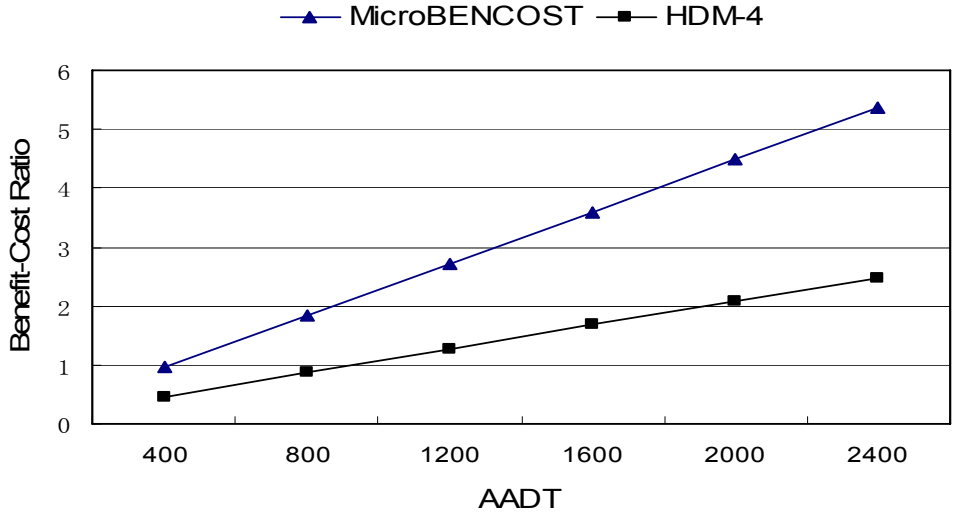


Figure 3. Marginal means of benefit-cost ratio by AADT.

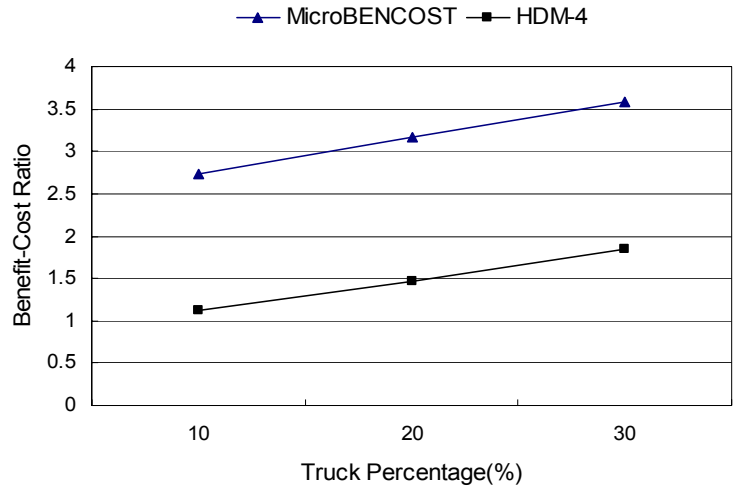


Figure 4. Marginal means of benefit-cost ratio by truck percentage.

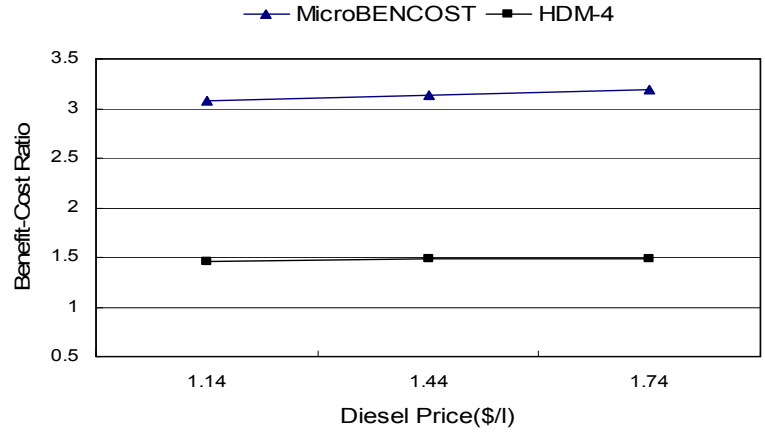


Figure 5. Marginal means of benefit-cost ratio by diesel price.

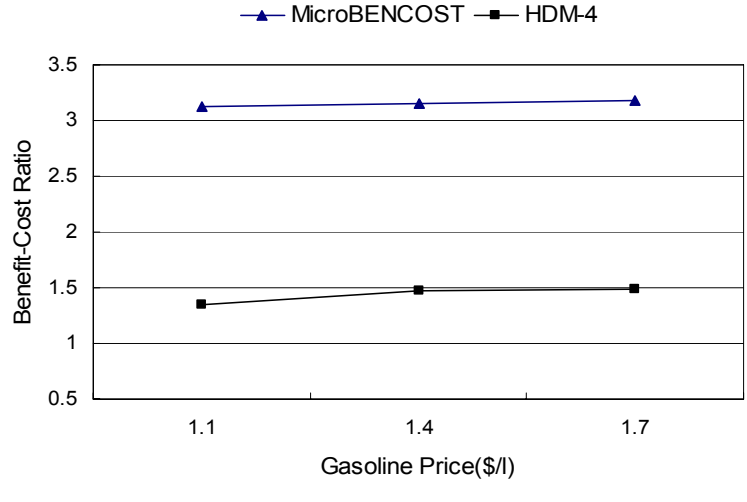


Figure 6. Marginal means of benefit-cost ratio by gasoline price.

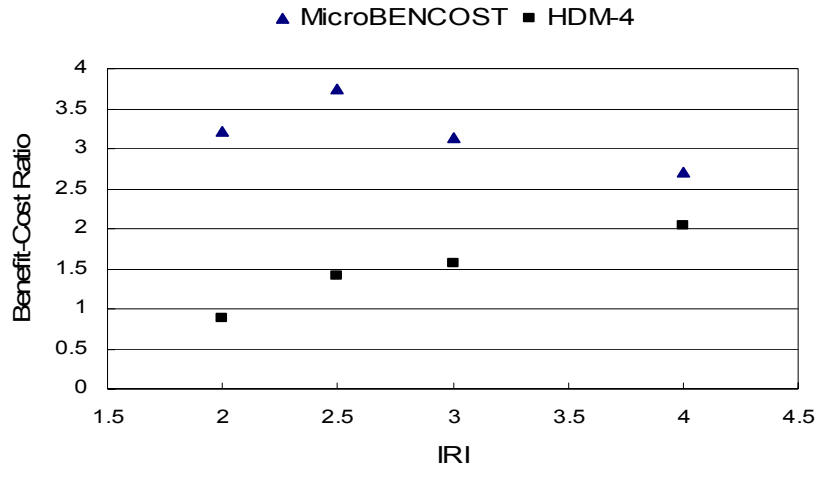


Figure 7. Marginal means of benefit-cost ratio by IRI.

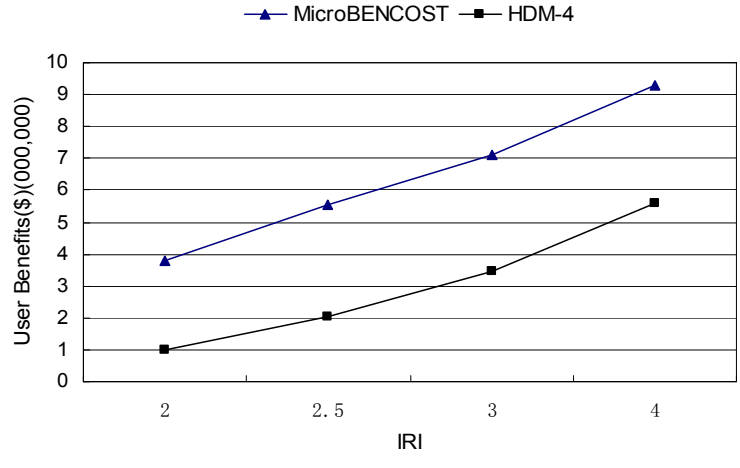


Figure 8. Marginal means of user benefits by IRI.

In Figure 7, the unsteady changes of benefit-cost ratios are caused by the changes of construction costs associated with different pavement conditions represented by IRI values. Thus, the decline of benefit-cost ratios generated by MicroBENCOST with the increase of IRI indicates the improvement of worse pavements requires higher construction costs than the road user benefits which the pavement improvement brings. Nevertheless, in HDM-4, the lasting increase of benefit-cost ratio with the increasing IRI shows the improvement project on the worse pavement can generate more benefits than the increase in construction costs.

Since the exact construction costs related to IRI are difficult to accurately quantify and the construction costs vary with various factors in addition to road surface condition, the road user benefits as another important output were generated by the two programs. Therefore, the effect of construction cost variations is removed in order to obtain relationships between road user benefits and critical input parameters. It has been found that the relationships between road user benefits and critical input

parameters are similar to the relationship between the benefit-cost ratios and critical input parameters except IRI. Figure 8 shows that the road user benefits increase with the increase of the original IRI values of the improved pavements for both programs.

Although there are different relationships between road user benefits and each of the critical input parameters, there seem to be similar trends from the outputs of the two programs. It has also been found the outputs from MicroBENCOST are normally higher than the outputs from HDM-4 under the same input conditions due to different methodologies used in developing the two programs. Therefore, if engineers use MicroBENCOST to calculate benefit-cost ratio, a positive engineering economic case may be obtained (the benefit-cost ratio is greater than one), while if HDM-4 is used to analyze the same project, a negative case may be obtained. This result shows that engineers and decision-makers should not consider the benefit-cost ratio obtained from either of two software as an absolute value, especially when AADT level is low. It is interesting to note that under the studied ranges of IRI, truck percentage, and fuel prices, both softwares have all output benefit-cost ratios greater than 1.0 when AADT is greater than 2400.

Correlation Models based on Outputs of MicroBENCOST and HDM-4

As described above, it is quite time-consuming to determine the exact values of all input parameters when using MicroBENCOST and HDM-4 for application. It is noted that the two programs are more appropriate for project level engineering economic analysis, and it is difficult to use the software for road user benefit calculation of many highway sections at the highway network level. Since many of the software input parameters are not sensitive enough to the road user benefit, it is possible to simplify the calculation process for road user benefit estimation. Therefore, two correlation models, which are helpful in estimating user benefits based on the outputs of the two software, were developed using statistical analysis. The estimation of the road user benefits or benefit-cost ratio on Saskatchewan rural highways for many pavement improvement projects will be more efficient and convenient by using these models rather than running the software themselves. Five critical input parameters, IRI, AADT, diesel price, gasoline price, and truck percentage are used as independent variables in the models to calculate road user benefits.

Table 2 shows the coefficients of the independent variables as well as the constant for the regression analysis based on the outputs from MicroBENCOST and the five selected independent variables. For a given level of significance, $\alpha = 0.05$, it is proven all coefficients, for each independent variable, are statistically significant since the values of P are all smaller than 0.05. The column of non-standardized coefficients indicates the expected benefit-cost ratio can be expressed as Eq.6 below:

$$\text{User benefits} = -1.103E7 + 2.745E6(\text{IRI}) + 4849(\text{AADT}) + 383618(\text{Diesel Price}) + 221823(\text{Gasoline Price}) + 93065(\text{Truck Percentage}) \quad (6)$$

Table 2. Coefficients of independent variables for model of MicroBENCOST.

Variables	Non-standardized Coefficients		Standardized Coefficients	t	P
	B	Std. Error	Beta		
(Constant)	-1.103E7	243108		-45.386	.000
IRI	2.745E6	40232	.478	68.221	.000
AADT	4849	41.644	.816	116.451	.000
Diesel price($/l)	383618	84816	.032	4.523	.000
Gasoline price($/l)	221823	84816	.018	2.615	.009
Truck%	93065	3484	.187	26.711	.000

Table 3. Coefficients of independent variables for model of HDM-4.

Variables	Non-standardized Coefficients		Standardized Coefficients	t	P
	B	Std. Error	Beta		
(Constant)	-8.180E6	196423		-41.643	.000
IRI	2.322E6	51778	.677	44.836	.000
AADT	2150	56.052	.579	38.364	.000
Truck%	76161	4689	.245	16.240	.000

The coefficient of determination (R2) is 93% for the regression model (Eq.6), which means about 93% of variation in road user benefit estimation can be explained by this model.

Table 3 shows the coefficients of the independent variables as well as the constant for the regression analysis based on the outputs from HDM-4 and the selected independent variables. For a given level of significance, $\alpha = 0.05$, it can be seen three independent variables are statistically significant in the finalized model and diesel and gasoline prices are excluded. Based on the results shown in Table 3, the correlation model for HDM-4 is presented below:

User benefits = -8.180E6 + 2150(AADT) + 2.322E6(IRI) + 76161(Truck Percentage) (Eq.7)

The R2 value is 85.3% for the regression model (Eq.7). This may be due to the exclusion of the diesel and gasoline prices in the mode, which shows fuel prices may not be the most important factor on road user benefits but they do have influence on results. This value is lower than the R2 value of model of MicroBENCOST because more independent variables usually result in a greater R2.

Conclusions

Several conclusions can be drawn from this study:

(1)Literature review shows that road user costs and economic impact of the pavement preservation investment have not been specifically considered in the network level Asset Management System. It is important to incorporate economic parameters into the network level pavement rehabilitation decision-making process.

(2)Determination of input parameters for running MicroBENCOST and HDM-4 are very time-consuming. However, not all input parameters have the obvious influence on road user benefits output. Several critical input parameters including IRI, AADT, diesel price, gasoline price, and truck percentage were identified.

(3)The comparison between outputs generated by MicroBENCOST and HDM-4 shows that the two programs generally have a similar changing trend, and the outputs from MicroBENCOST are normally higher than the outputs from HDM-4 under the same input conditions. Engineers and decision-makers should not consider the benefit-cost ratio obtained from the software as an absolute value, especially when AADT level is low.

(4)Two correlation models based on the outputs of MicroBENCOST and HDM-4 and the selected critical input parameters were developed. The models will help the road user benefit estimation in a simple and fast manner at the highway network level. The high values of coefficient of determination indicate both models can provide estimates of road user benefits as accurate as MicroBENCOST and HDM-4 do under the Saskatchewan rural highway conditions.

References

Callao, R. A. (2007). “HDM-4 road use costs model documentation.” The World Bank, USA.

Garber, N. J., and Hoel, L. A. (2002). “Traffic & Highway Engineering.” Wadsworth Group, USA.

Jung, J., Kaloush, K. E., and Way, G. B. (2002). “Life Cycle Costs Analysis: Conventional Versus Asphalt-Rubber Pavements.” Dept. of Civil & Environmental Engineering, Arizona State University, and Arizona Dept. of Transporation, Arizona, USA.

Lamptey, G., Ahmad, M., Labi, S., and Sinha, K. (2005). “Life cycle costs analysis for indot pavement design procedures.” Purdue e-Pubs: docs.lib.purdue.edu/jtrp/138, Indiana, USA.

Lyall, P. (2007). “2007 Updates for Microbencost Default Values.” *BC MoTH, Planning and Major Projects*, British Columbia, Canada.

Memmott, J. L., Richter, M., Pardo, A. C., and Wildenthal, M. (1999). MicroBENCOST User's Manual, Texas Transportation Institute and The Texas A&M University System College Station, USA.

Reilly, R. J., Jencks, C. F., Hedges, C. J., Delaney, E. P., Freer, H., and Barnes, N. (2005). "Analytical Tools for Asset Management." Transportation Research Board, Washington D.C, USA.

Roper, B. A., and Keltner, D. M. (1999). "User Benefits of Railroad Grade Separation in a Small Community: Practical Techniques for Applying MicroBENCOST." St. Cloud Area Planning Organization, and BRW, Inc., USA

Salem, O., and Genaidy, A. (2007). "Improved Models for User Costs Analysis." Dept. of Civil & Environmental Engineering, Dept. of Industrial Engineering, University of Cincinnati, USA.

Sinha, K. C., and Labi, S. (2007). "Transportation Decision Making: Principles of Project Evaluation and Programming." John Wiley & Sons, Inc., USA.

Zheng, N. N. (2009). "Evaluation of tools for calculating road user costs in decision making process for pavement capital improvement on low volume Saskatchewan rural highway network." Master Thesis, University of Regina

Comparison of Cordon and Optimal Toll Points Road Pricing Using Genetic Algorithm

Zhi Zuo[1], Ryo Kanamori[2], Tomio Miwa[3] and Takayuki Morikawa[4]

[1] Ph.D. Candidate, [2] Research fellow, [4] Professor, Graduate School of Civil Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Japan, 464-8603; Email: zuozhi@trans.civil.nagoya-u.ac.jp, kanamori@bin.t.u-tokyo.ac.jp, morikawa@nagoya-u.jp
[3] Associate Professor, EcoTopia science institute, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Japan, 464-8603; Email: miwa@civil.nagoya-u.ac.jp

Abstract

Taking advantage of a genetic algorithm based bi-level programming solution to find the optimal scheme for second-best road pricing. The highlight of the work is a case study using the actual Nagoya metropolitan area road network instead of a numerical network. The effectiveness of the method is evaluated for this network by comparing 3 cordon pricing schemes with optimal tolled schemes having the same number of links. The results show that under a situation with the same number of links, the scheme using optimal toll points works better than the cordon pricing scheme.
Key Words: road pricing; second best; GA; BLPP

Introduction

As economies develop and technology improves, the availability of vehicles increases and car ownership in urban areas rises. As a result, many cities have been suffering from heavy traffic congestion in recent years. The sharp increase in car ownership has led to a series of problems in metropolitan areas, mostly relating to congestion, safety, and environmental issues. Researchers have looked into many technology and policy alternatives to deal with this situation. As one method of traffic control, road pricing had attracted much attention. London, Singapore, and Stockholm have all implemented various forms of congestion pricing and obtained effective results. Through road pricing, changes have been achieved in traffic demand, traffic destination, departure times, and vehicle routing, while unnecessary trips have been reduced and the overall operating conditions of the road network have improved greatly.

The idea of road pricing derives from the viewpoint that road users are actually paying less than the cost (Marginal Cost) that should be paid. By implementing a charging method, traffic conditions move closer to the optimal state; this kind of pricing is called first-best pricing. Under the assumption that all links can be charged, first-best road pricing can be solved by system optimization assignment. In fact, however, first-best pricing is not practically appealing, because of high operational costs and poor public acceptance. As a result, few first-best pricing schemes have been implemented.

In recent years, more practical methods based on second-best road pricing have received considerable attention. Yang and Huang (1997) investigated the principle of marginal cost pricing and presented a general model of link congestion tolls with elastic demand. This model can be used not only in first-best pricing but also in alternative second-best pricing schemes. Verhoef (2002) proposed a method for determining second-best toll levels in a study in which only a subset of all links can be charged. Shepherd et al. (2004) showed that a genetic algorithm (GA) approach can be used for large-scale networks, but the paper presented only a mini network case with few links. Sumalee et al. (2005) developed a GA-based method and compared judgmental and optimal road pricing cordons. Zhang et al. (2004) and Sumalee (2004) presented an investigation of cordon-based second-best congestion pricing, with both papers focusing on the design of the cordon using GA. Zhang et al. (2004) looked into variable demand equilibrium under second-best pricing with a parameterized pricing scheme. Santos (2004) simulated cordons in eight towns and estimated the impacts of schemes using SATURN (Simulation and Assignment of Traffic to Urban Road Networks); the paper compared the benefits of single cordon pricing and double cordon pricing. Maruyama et al. (2007) compared cordon- and area-based road pricing using an elastic demand model. Kanamori et al. (2008) used a semi-dynamic model of stochastic user equilibrium (UE) to evaluate the effect of cordon road pricing. In most cases, this research work focused mainly on numerical case studies using small-scale networks.

In fact, for a given toll level on a certain network, it is unknown where charges should be imposed to obtain the optimal effect. Cordon pricing for a central business district is a practical approach but it might not be an optimal one. The objective of this study is to compare cordon pricing at a certain toll level with a system of optimal toll points obtained using GA-based bi-level programming. The study is based on the full Nagoya metropolitan area road network. In order to compare the effect of cordon pricing with optimal toll points, a cordon pricing scheme with a uniform toll level and two optimized schemes with various link sets are examined.

The paper is organized as follows. In the next section, we will discuss the bi-level programming model for second-best road pricing and explain how a GA is used to solve the problem. Details of the GA coding method will be given. In section 3, the Nagoya metropolitan area case study will be described, including a comparison between the cordon pricing scheme and the optimal toll scheme. Section 4 presents

the results of the case study and offers an analysis and our final conclusions are given in Section 5.

Model Formulation

Tolls charged at optimal points can be regarded as link congestion tolls in the network, so the problem can be depicted as a bi-level second-best pricing model that is subject to standard traffic equilibrium. The problem can be solved using a GA, an algorithm that has proven very useful in solving large-scale network questions. In this section the first part introduced the bi-level programming of road pricing problem and the latter parts give out details of GA process.

Upper Level. The upper level model can be evaluated to maximize the social welfare gained from the tolled network.

$$Max \quad SW = UB - SC \tag{1}$$

Here user benefit (UB) is evaluated by the following formulation:

$$UB = \sum_{w \in W} \int_0^{d_w} D_w^{-1}(x)dx \tag{2}$$

where w is the OD pair, W is the set of OD pairs, d_w is the demand between OD pair w, and $D_w^{-1}(x)$ is the Inverse Demand function, whose value depends on the travel cost, and the social cost (SC) is formulated as:

$$SC = \sum_{a \in A} v_a t_a(v_a) \tag{3}$$

where a is the link, A is the set of links, v_a is the flow on link a belonging to links set A, and $t_a(v_a)$ is the travel time on link a, which is a function of link flow. Then the social welfare can be expressed as following formulation:

$$SW = UB - SC = \sum_{w \in W} \int_0^{d_w} D_w^{-1}(x)dx - \sum_{a \in A} v_a t_a(v_a) \tag{4}$$

The demand function is set as:

$$D_{rs}(C_{rs}) = D_{rs}^0 \exp\left[\sigma(1.0 - \frac{c_{rs}}{c_{rs}^0})\right] \tag{5}$$

where D_{rs}^0 is the demand for minimum generalized travel cost during assignment, C_{rs}^0 is the minimum generalized travel cost, and σ is the price elasticity of demand.

Lower Level. The lower level model is the user equilibrium with variable demand:

$$Min \qquad Z(x,d) = \sum_a \int_0^{x_a} t_a(\omega) d\omega - \sum_{rs} \int_0^{d_{rs}} D_{rs}^{-1}(\omega) d\omega \tag{6a}$$

Subject to

$$f_k^{rs} \geq 0 \tag{6b}$$

$$d_{rs} \leq \bar{d}_{rs} \tag{6c}$$

$$d_{rs} = \sum_k f_k^{rs} \tag{6d}$$

In this formulation, f_k^{rs} is the flow on the *kth* path between OD pair *rs*, d_{rs} is the demand between the OD pair, $\bar{d}_{rs}$ is the upper bound of demand between origin r and destination s, $D_{rs}^{-1}(x)$ is the Inverse Demand function, whose value depends on the travel cost. (Because the algorithm for solving the variable demand user equilibrium problem is very complex, it cannot be depicted here in details. Full details of the algorithm for user equilibrium with variable demand can be found in Urban Transportation Network, Chapter 6, Sheffi, 1985.)

The form of travel time is evaluated as:

$$t_a(x_a) = t_0 \left[1 + \alpha (v/\mathrm{c})^{\beta} \right] + toll / VOT \tag{7}$$

In this formulation, tolls are determined by toll level; once the toll level is determined, the monetary cost will be a certain amount.

In order to make the problem practicable, we choose several subsets of the whole network links as candidate toll links; only links in these subsets can be selected for tolls. These subsets are determined empirically and we place links into various combinations as road pricing schemes. When a link is selected for one combination, it means the link is charged. For an actual network, the total number of combinations is huge, even if they are selected only from one subset. Going through all combinations using an optimal result finding procedure would be very difficult. A GA is able to determine the optimal result because of its favorable procedure of natural selection. The GA procedure used is outlined below.

Step 1. (Initialize population). Randomly select links from the subset and randomly generate toll levels.

Step 2. (Fitness evaluation). Employ the model to obtain the optimum scheme from each combination of toll links.

Step 3. (Selection). Select schemes with higher values of social welfare for reproduction.

Step 4. (Crossover and mutation). Pair the populations and randomly exchange the toll links between pairs. Randomly change some toll links and toll levels.

Step 5. (Verification of halt criterion). If the convergence criterion has not been satisfied, go to Step 2 and continue with the next generation; otherwise stop the program.

The detailed GA process used in this study is shown in the Figure 1. During the initialization procedure, information about the network and its subsets, OD data, and model parameters are all read by the program. The fitness of all schemes is evaluated in the next procedure, where lower-level programming plays the core role. With respect to a certain scheme, there will be an assignment of variable demand user equilibrium and after this assignment the social welfare, user benefit, and social cost will be output.

Figure 2 shows the detailed structure of one chromosome, where one chromosome represents a toll scheme, a combination of toll links. For example, when there are 20 links need to be tolled, the length of each chromosome is set as 20. The upper rows are gene indexes of binary format. The bottom row represents the toll link index; in Figure 2 the first entry in this row, 231, means that the link with ID 231 in the subset is selected as the toll link.

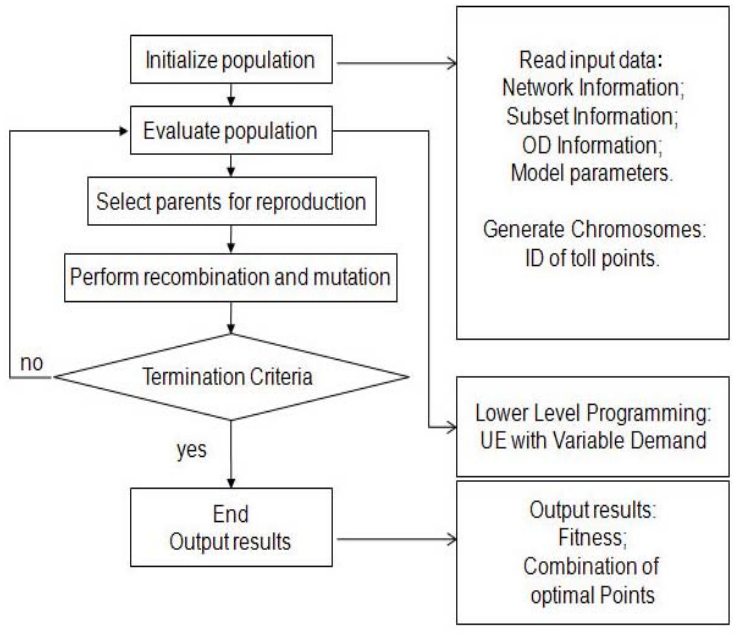


Figure 1. GA process.

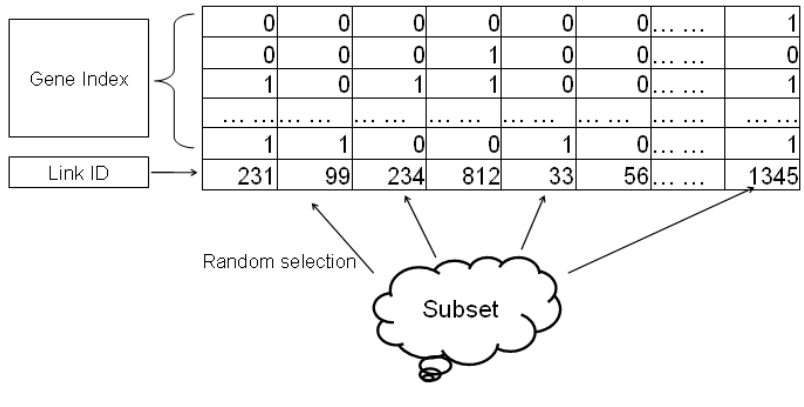


Figure 2. Structure of chromosome.

Nagoya Metropolitan Area Network Case Study

As a case study, the model was applied to the Nagoya metropolitan area network. This network comprises 4209 one-direction links and there are at total of 279 zones in the city area. Personal trip (PT) data collected in 1996 is used to simulate traffic flows on the network. The unit of traffic flow placed on the network is the standard Passenger Car Unit (PCU) and we assume that the value of vehicle driver time is 83.4 Japanese Yen per minute. In this case study, we set only a uniform toll level of 300 Japanese Yen, it means once a link is selected as toll link, it is tolled 300 Japanese Yen.

In order to compare optimized toll points with cordon pricing, we introduce three cases, as detailed in the paragraphs that follow.

Case A. **Cordon pricing.** In this case we choose the 3 cordon areas of Nagoya central urban area. The small cordon is central business district of Nagoya city, the large cordon is the central urban area of Nagoya and the one between these 2 cordons is the middle cordon, because when vehicles enter these 3 areas there are respectively 20, 50 and 42 possible toll links, we then choose schemes with the same number of links for comparison in the next 2 cases.

Case B. **Optimal toll points for artery road subset pricing**. The subset in this case is selected from among the artery roads in the central urban area of the Nagoya road network. There are roads with at least 4 lanes in the Nagoya city area and that are not expressway links or existing toll links. The total link number of this subset is 1198, this means there are 1198 potential links can be chosen as the toll links.

Case C. **Optimal toll points for central urban area road subset pricing.** The subset in this case is selected from the central urban area of the Nagoya road network. It is easy to know that the subset of Case B is a subset of Case C and in Case C also the links in the subset exclude expressway links and existing toll links. The total link number of this subset is 2096.

Figure 3 shows the location and range of the three subsets taken from the road network of Nagoya city. The top-left image shows the location and range of cases in the Nagoya road network, all cases are carried out in the Nagoya central metropolitan area. The Case A subset is the set of links within the 3 cordon lines of the top-right image, the red line is inner cordon, the pink one is the middle cordon and the brown one is outer cordon. All links in the combination are one-direction links and vehicles are charged only upon entering the cordon. The Case B subset consists of the bold blue links in the left-lower image. These links are artery roads in the central city area with at least 4 lanes, Expressway links and existing toll links are excluded. The Case C subset consists of the bold grown links in the right-lower image. All links in the central city area are included, while as with Case B the expressway links and existing toll links are excluded too.

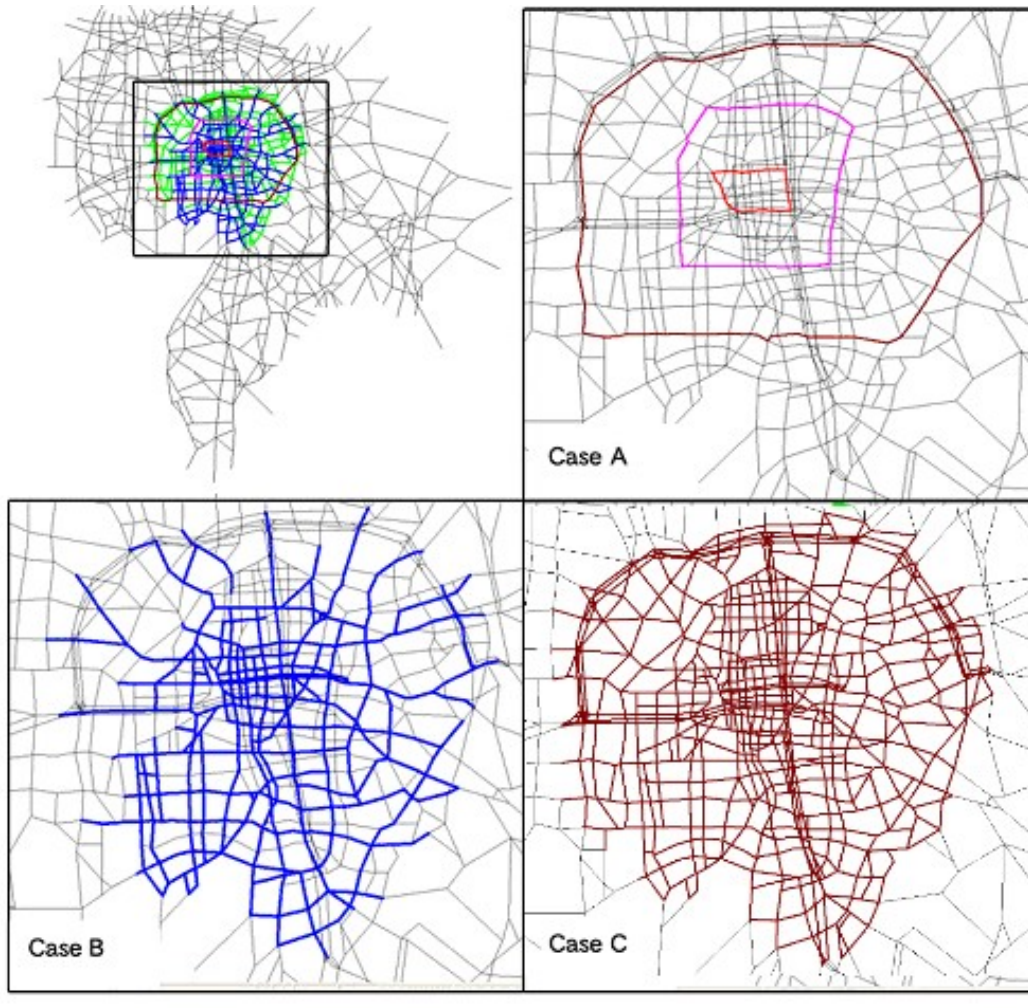


Figure 3. Cases of the study.

Result of Case Study

With the introduction of road pricing schemes on the network, traffic flows are affected in various ways: most drivers try to change their routes, some adjust their departure time or move to another mode of transport, such as subway. But from the point of traffic conditions on the network as a whole, it moves toward the system optimum. In other words, the network becomes less congested and users of the network benefit.

Results were obtained for the three cases by applying the GA-based bi-level programming model. As anticipated, the results showed that the cases with optimal toll points and levels achieved better results than the cordon pricing case. In figure 4, we can see that for all of the cordons Case A results in the lowest value of social welfare; both Cases B and C yield better values of social welfare than cordon pricing. This comparison demonstrates that, although cordon pricing schemes are more feasible and easier to realize, schemes based on optimal toll points are more attractive as regards social welfare.

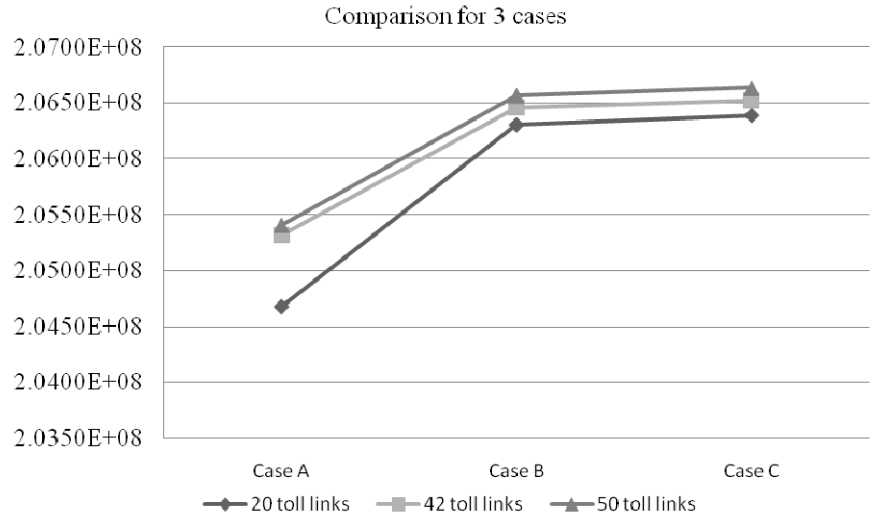


Figure 4. Comparison of the three cases.

Table 1. Result of 3 Cases (Unit: min).

Toll link numbers	Case A	Case B	Case C
20	2.0468112E+08	2.0630126E+08	2.0638543E+08
42	2.0531886E+08	2.0646006E+08	2.0652708E+08
50	2.0540568E+08	2.0656883E+08	2.0663627E+08

Table 1 shows the detailed result in terms of number. The result of different toll link numbers share the same trend in 3 cases, all results of Case B and Case C are better than cordon pricing cases with the same toll link numbers. As for the same toll link numbers, in this case study it seems that with more toll links the social welfare tend to be higher.



Figure 5. Results for Case C.

Figure 5 shows the example of optimal result for Case C of 20 links with the same number of inner cordon. The figure shows the combination of optimal links in the network, with the toll links in the scheme marked in bold style. It is clear that most of the toll links tend to be located in the central part of the urban network or on arterial corridors, non-artery links can hardly be chosen as toll links. The table at the right part of image is table of link ID of these optimal links.

Conclusions

In this study, we develop a bi-level model based on a genetic algorithm to determine optimal combination of toll points. As a case study, the model was applied to the actual large-scale road network of the Nagoya metropolitan area. In order to compare the effectiveness of the method with other policy options, we introduced three cases for the case study: the cordon pricing case, the optimal combination of toll points case with tolls on central city area artery roads only, and the optimal combination of toll points pricing case with tolls on any road in the central urban area road network. This comparison between the cordon pricing case and the two optimal toll points cases demonstrated that the latter cases performed better than the former one, although cordon pricing schemes are more feasible and easier to realize, schemes based on optimal toll points are more attractive as regards social welfare.

References

Kanamori, R., Miwa, T., and Morikawa, T. (2008). "Evaluation of road pricing policy with semi-dynamic combined stochastic user equilibrium model", *International Journal of ITS Research*, 6(2).

Maruyama, T. and Sumalee, A. (2007). "Efficiency and equity comparison of cordon- and area-based road pricing schemes using a trip-train equilibrium model", *Transportation Research Part A 41*, 655-671.

Santos, G. (2004). "Urban Congestion Charging A Second-Best Alternative", *Journal of Transport Economics and Policy,* 38(3), 345-369.

Sheffi, Y. (1985) *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall, Englewood Cliffs, N.J.134-159

Shepherd, S. and Sumalee, A. (2004) "A Genetic Algorithm Based Approach to Optimal Toll Level and Location Problems", *Networks and Spatial Economics,* 4, 161-179.

Sumalee, A. (2004) "Optimal Road User Charging Cordon Design: A Heuristic Optimization Approach", *Computer-Aided Civil and Infrastructure Engineering* 19, 377-392.

Sumalee, A., May, T., and Shepherd, S. (2005) "Comparison of judgmental and optimal road pricing cordons", *Transport Policy* 12, 384-390.

Verhoef, E. T. (2002) "Second-best congestion pricing in general networks", Heuristic algorithms for finding second-best optimal toll levels and toll points." *Transportation Research Part B* 36, 707-729.

Yang, H. and Huang and H. J. (1998) "Principle of marginal-cost pricing: How does it work in a general road network?" *Transportation Research Part A,* 32(1), 45-54.

Yang, H. and Huang, H. J. (2005) *Mathematical and Economic Theory of Road Pricing.* ELSEVIER.

Zhang, H. M. and Ge, Y. E. (2004) "Modeling variable demand equilibrium under second-best road pricing", *Transportation Research Part B,* 38,733-749.

Zhang, X. N. and Yang, H. (2004) "The optimal cordon-based network congestion pricing problem", *Transportation Research Part B* 38, 517-537.

Identifying Key Factors Affecting Students' Travel Modes Using the Multi-perspectives Diagnosis Approach

Huaguo Zhou[1], Peter Hsu[2] and Shaoqiang Chen[3]

[1] Assistant Professor, Department of Civil Engineering, Southern Illinois University Edwardsville, Edwardsville, IL 62026-1800, Email: hzhou@siue.edu
[2] District Safety Program Engineer, Florida Department of Transportation, 11201 N McKinley Dr. Tampa, Florida 33612, Email: ping.hsu@dot.state.fl.us
[3] Graduate Research Assistant, Center for Urban Transportation Research, University of South Florida, 4202 E. Fowler Ave., Tampa, FL 33620

Abstract

In this paper, survey data of students' travel modes from both parents and students were collected at 18 schools (16 elementary and 2 middle schools) in Pinellas County, Florida. A new diagnosis approach was used to pin down the factors significantly affecting students' travel modes. Multi-perspective diagnosis results identified different factors that significantly affect the walking/biking rate for different groups of students. Generally, students living in different distance intervals are subject to different barriers. Security and safety are the primary factors of concern for short distance interval people, while distance is the primary factor of concern for those who live farther away from the school. The multi-perspectives diagnosis proved to be an effective method to mathematically search for the necessary-sufficient condition for increasing students' walking/biking rate.

Introduction

Due to obesity and worsening health conditions of children, researchers have put more efforts to develop efficient treatments in recent decades. Among different therapies, physical treatments play an important role in reducing obesity. That is to encourage children to do more exercise everyday, in which walking/biking to school is designed as an integral part before and after school. However, only a very small portion of students choose to walk or bike to school. The Safe Routes to School (SRTS) program was thus developed, which originated from Denmark in 1976 and aims to encourage elementary and middle school students to walk or bike to school by providing safer and more comfortable routes to students.

In the United States, studies that explore the potential conditions on which students' travel behavior depends upon were conducted before the SRTS program was implemented. Research in the literature has shown that students' travel behavior to school is a complicated socio-economic activity, which is affected by many potential factors such as household demographics and the household location (*1, 2*). Economic factors (travel cost, income, expense, etc.) and demographic factors (gender and ethnicity) were demonstrated as important factors (*2*). Travel time is also a significant factor affecting students' travel modes (*2, 3*). Travel time is an indirect reflection of distance. As distance increases, more time is needed to walk/bike to school. And there is a greater concern with safety issues such as traffic safety and crime that children are exposed to on the way to school/home. The studies have shown that distance, traffic, and crime are three key factors preventing students from walking/biking to school (*4*). In the real world, there is often more than one single significant factor that influences students' travel mode (*5*). Previous research concluded that under different conditions (locations, units, and time, etc.), impact factors on students' travel modes are different, especially for walking/biking. Therefore, an efficient approach is necessary to facilitate the efforts to identify the significant factors and quantify their impacts.

In the literature, a statistical test such as chi-square test was employed to explore relationships between the level of bicycling and demographics, objective environmental factors, perceptions of the environment and attitudes (*6*). A multivariate logistic model has been developed to quantify the relative safety of a bicycle route based on bicycle crash data in the previous four years (*8*). Moreover, a multinomial logit model (*1*) was fitted to estimate students' travel mode choice to school. Logistic regression method (*7*) has shown to be an effective method to quantify the factors' impacts. However, few statistical test methods were used in testing the relationship between students' choice on

walking/biking and potential influential factors. No modeling approaches have been well developed exclusively for predicting the impacts on walking/biking rate, which is very important to the success of the SRTS program.

In this paper, a new statistical diagnosis approach was applied to conduct a sequential search of the necessary-sufficient significant factors in multiple perspectives. The data was collected from a survey on student's walking/biking to school for 18 schools (16 elementary and 2 middle schools) in Pinellas County, Florida. The survey forms (*9, 10*) and online input tools designed by the National Center for Safe Routes to School (SRTS Clearinghouse) were used for both students and parents. A total of 644 classrooms were surveyed, with 347 tally sheets returned by teachers and a student participation rate of 53.9%. A total of 5,885 parent surveys were distributed, and 2,551 were returned, a response rate of 44%.

Descriptive Statistics of Survey Data from Different Perspectives

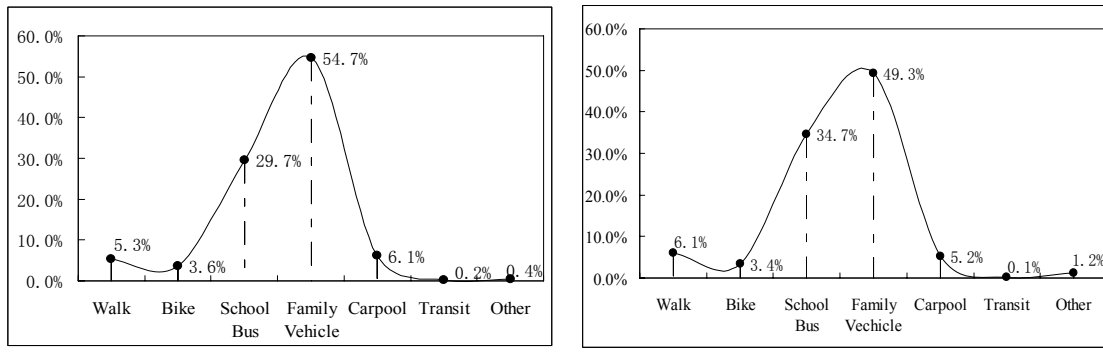
To perform multiple perspectives analysis, questions in the survey were selected and treated as impact factors. Different choices of each question were treated as different factor levels. For example, the student's grade level is a factor. The different choices from K through 8th grade are defined as nine different levels. Under this setting, data can be analyzed in three perspectives, namely, overview perspective, forward direction perspective, and backward direction perspective.

Overview perspective*:* The basic descriptive statistics of the parents' survey are summarized in Table 1. It lists the distribution of distance, travel time, grade level, school attitude, and parent's attitude toward the walking/biking. Approximately 28.5% people live in the distance that is practical for walking or biking to school (less than 1 mile). Over 50% students' travel time is less than 10 minutes. 75.5% of all surveyed students are elementary school students (K-5th grades). It is obvious that a considerable amount of the parents' attitudes towards walking and biking are positive. Approximately 32.6% of parents think that it is fun and very fun to walk/bike to school, and 71.7% of parents think that it is healthy and very healthy to walk or bike to school. The survey results also showed that only a small portion (16.7%) of schools hold positive attitudes towards walking/biking. About 58.0% parents will never allow children to walk/bike alone to school.

Table 1. Basic Descriptive Statistics (Percentage)

Distribution of Distance between Home and School					
<1/4 mile	1/4~1/2 mile	1/2~1 mile	1~2 miles	>2 miles	DK
9.4	7.2	11.9	17.7	46.9	4.6
Distribution of Normal Travel Time to School					
<5 min	5~10 min	11~20 min	>20 min	DK	
19.6	31.6	27.9	18.1	2.5	
Distribution of Student's Grade Levels					
K	1	2	3	4	5	6	7	8
15.1	13.1	13.7	12.3	11.3	10.0	11.1	8.4	5.0
Distribution of Grade Levels That Students are Allowed to Walking/Biking Along										
K	1	2	3	4	5	6	7	8	9plus	Never
0.2	0.5	1.2	4.0	5.3	6.4	10.6	3.2	6.3	0.7	58.0
Distribution of School Attitudes towards Walking/Biking				
Strongly Encourage	Encourage	Neutral	Discourage	Strongly Discourage
2.8	13.9	72.8	2.5	2.8
Distribution of Walking/Biking Enioyment				
Very Fun	Fun	Neutral	Boring	Very Boring
10.1	22.5	48.1	4.1	4.1
Distribution of Walking/Biking Health				
Very Healthy	Healthy	Neutral	Unhealthy	Very Unhealthy
38.1	33.6	18.1	1.1	1.6

Figure 1 presents the distribution of travel modes from both parent and student surveys. The student survey showed that the percentage of walking and biking to school is approximately 8.9% and parent survey showed that percentage of walking and biking to school is approximately 9.5%. The answers are also very consistent from parents and their children for other travel modes. The majority of the students choose school buses or family vehicles. The low walking/biking rate may be attributed to distance, travel time, school attitudes, traffic speed, safety, and students' grade.



A. Distribution of travel modes (Parent) B. Distribution of travel modes (Student)

Figure 1. Distributions of travel modes

Figure 2 shows the ranks of issues affecting parents' decision toward students' walking/biking. The distance was ranked as the No.1 issue with 67.7% of parents selecting it and 22% of parents willing to change their minds if the distance changes. Besides distance, travel modes choices are also affected by other barriers, including traffic volume, traffic speed, intersection safety, sidewalk violence or crime, weather, crossing guards, time, driving convenience, and adults to walk with.

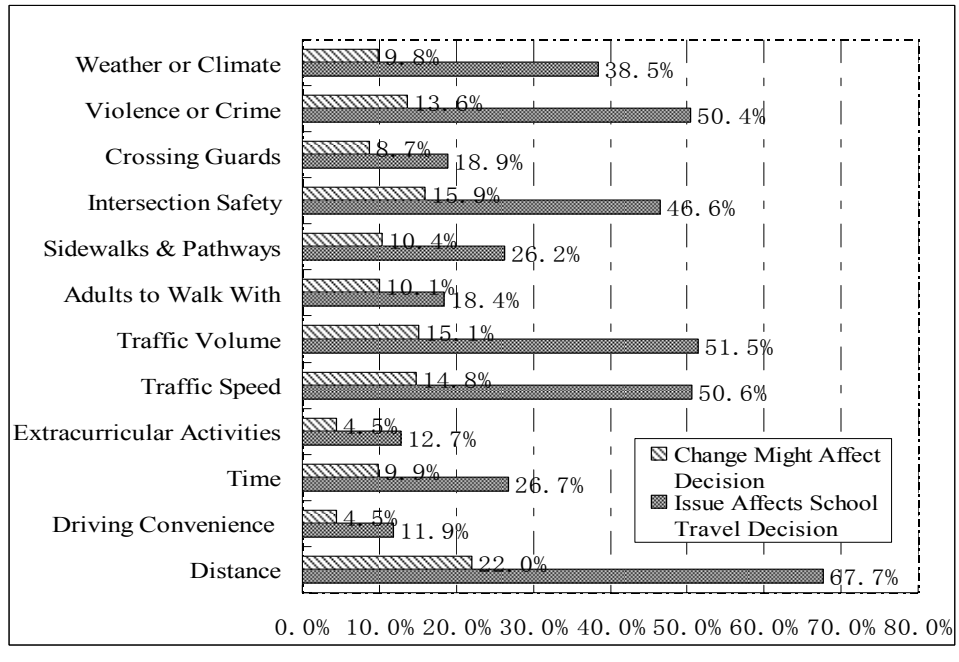


Figure 2. Ranks of the issues affecting parents' decision towards children's walking/biking with opportunities to improve

Further descriptive statistical analysis was conducted for survey data on the parents (sample size = 728) who reported their home is located less than one mile from the school. The results showed that the percentage of students walking/biking to school significantly increases by about 16%. The percentage of travel time less than 20 minutes also significantly increases by 18.5%. A higher percentage of parents believed that their children's schools have encouraging attitudes toward walking/biking (up by 8.5%), and a higher percentage of parents believed that walking/biking is fun (up by 15%) and healthy (up by 11.7%), for a total of 728 families living at less than one mile to school, 25.8% of children choose walking/biking. The percentage of students allowed to walk to school alone increases as the grade level increases. The non-positive attitudes from schools and parents are still considerably high with 72.4% of schools holding neutral and discouragement attitudes. Also, 41.6% of parents never allow their children to walk/bike alone to school.

Figure 3 shows the ranks of issues affecting parents' decision on walking/biking for those living less than one mile. The top 4 barriers are violence or crime (about 50.4%), intersection safety (40.2%), traffic speed (39.6%), and traffic volume (39.3%), respectively. This indicated that safety becomes the major concern when the distance is not a problem. The SRTS engineering program designed to improve intersection safety and to reduce traffic speed by installing speed flashers works better for those schools with a high percentage of students living within a walkable distance.

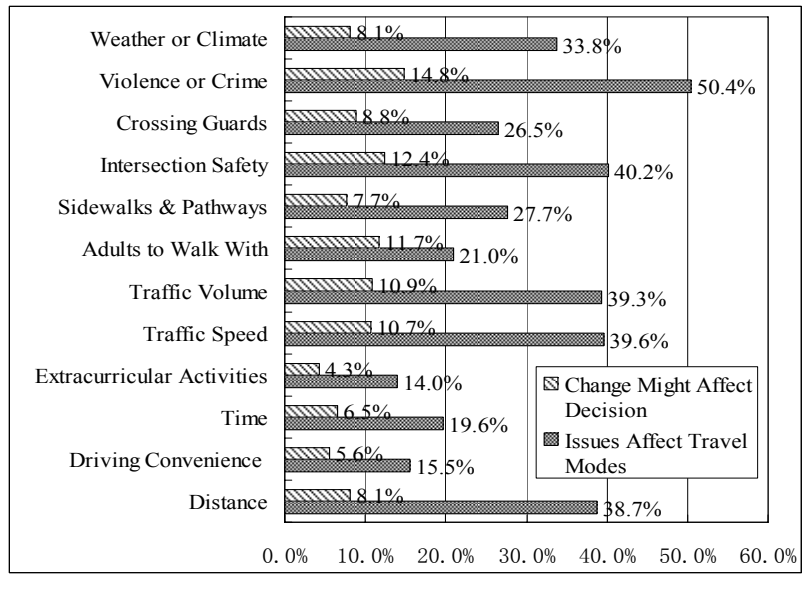
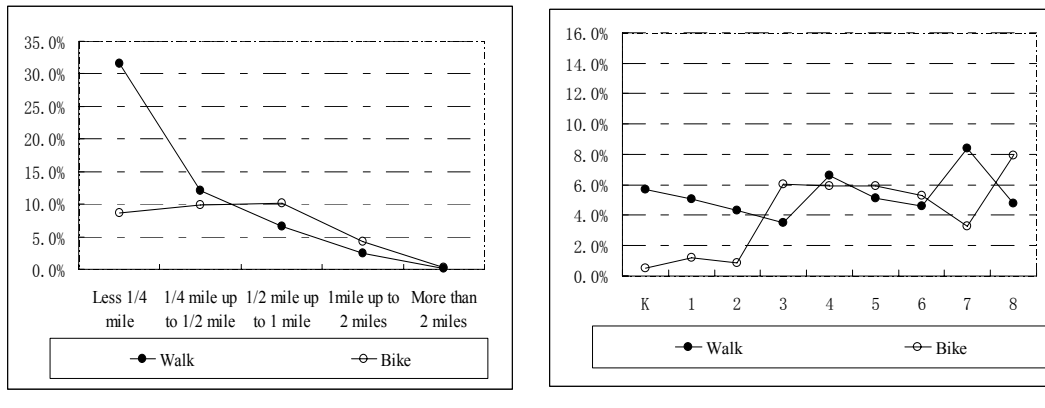


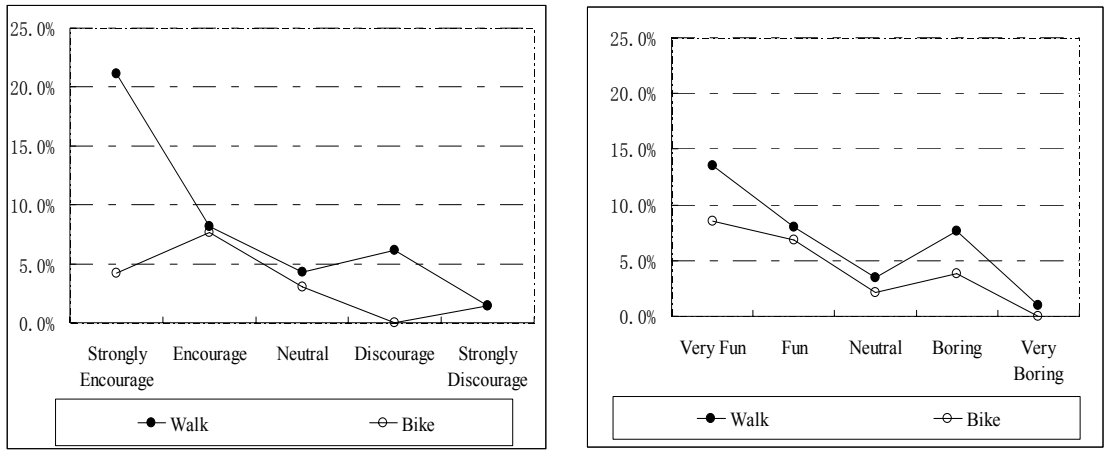
Figure 3. Ranking of the Issues affecting Parents Decision towards Children's Walking/Biking with Opportunities to Improve (within 1 mile)

Forward Direction: Cause-Effect Study of How Walking/Biking Rate Change by Different Factor Levels: In the cause-effect study, a chi-square test and a cause-effect plot were used to search for the significant factors. The chi-square test was used to assess whether paired observations on two variables/factors are independent. For instance, one variable is students' travel modes (value 1 for walking/biking, 0 otherwise), other variables are demographic variables (e.g., grade, gender, and distance) and subjective variables (e.g., school attitudes towards walking/biking, enjoyment, and health). The cause-effect plot visualizes how the travel modes change as values of other variables change. In the plot, the vertical axis denotes the walking/biking rate, and the horizontal axis represents factor levels. Figure 4 illustrates the cause-effect relationship between the walking/biking rate and demographic and subjective variables. The cause-effect plots reveal that the biking rate within 1 mile almost stays at a constant level (about 10%), while the walking rate drops significantly as distance increases from the 0.25-mile interval to the 0.5-mile interval. This matches a common sense assumption that within a very short distance region, students' are more likely to go to school via walking as they don't need to go across too many intersections. As the distance increases, biking is more convenient than walking. Generally, the walking/biking rate decreases as distance increases. On average, the walking/biking rate increases as the grade level increases. The biking rate remains low for the low grade students (K - 2nd). The biking rate between male and female is more significantly different than the walking rate. The biking rate deceased as the family size increased while the walking rate increased as the family size increased. For example, the sample size for family size of 7 is very small (only 2). This is why the walk rate for this family size is close to zero.



A. Walking/Biking Rate by Distance B. Walking/Biking Rate by Grade Levels

Figure 4. Cause-Effect plot of walking/biking rate by demographic factors



A. Walking/Biking Rate by School Attitudes B. Walking/Biking Rate by Enjoyment

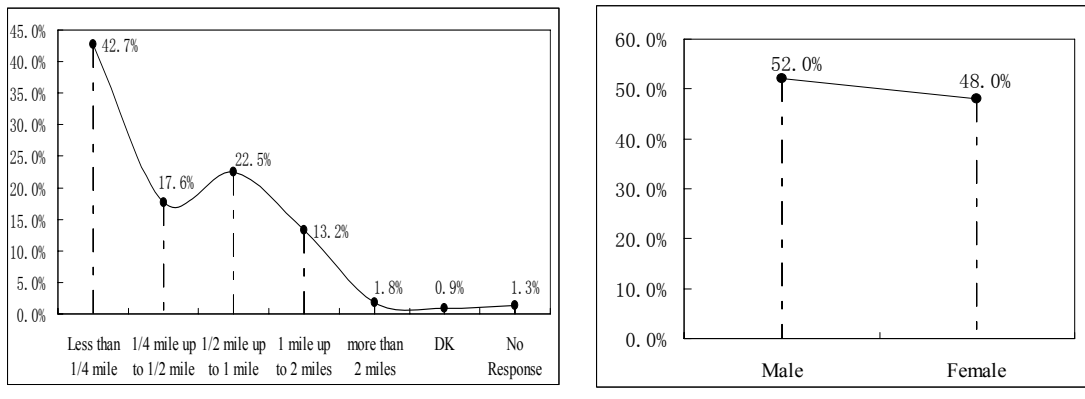
Figure 5. Cause-Effect plot of walking/biking rate by subjective factors

Figure 5 illustrates the cause-effect relationship between the walking/biking rate and subjective factors. It indicates that school, parents', and students' attitudes had significant impact on the students' walking/biking rates. The SRTS education program is aimed to improve students' walking/biking rate by education and encouragement. The survey results proved that this will be an effective improvement method. Table 2 gives the chi-square test results for all the demographic and subjective factors. From the test results, most of demographic and subjective factors are statistically significant (α level = 0.05), except for gender and family size. Significant variables include distance, grade levels, school attitudes, enjoyment, healthy, allowable grade level (defined as what grade level will parents allow students to walk/bike to school alone), and student's attitude.

Table 2. Chi-Square Test for the Factors in the entire Parents' Survey

Factor	Significant value
Distanc	0.000
Grade Level	0.002
Gender	0.064
Family Size	0.406
School Attitude	0.000
Enjoyment	0.000
Health	0.000
Allowable Grade Levels	0.000
Student's Attitudes	0.000

Backward Direction: Similarity Study of What Is in Common Within the Walking/Biking Group: The potential factors to be studied here are all the factors included in overview perspective section and forward direction perspective section. The sample size of the walking/biking group is 227. Figures 6 and 7 summarize the demographic as well as subjective similarities within this group.

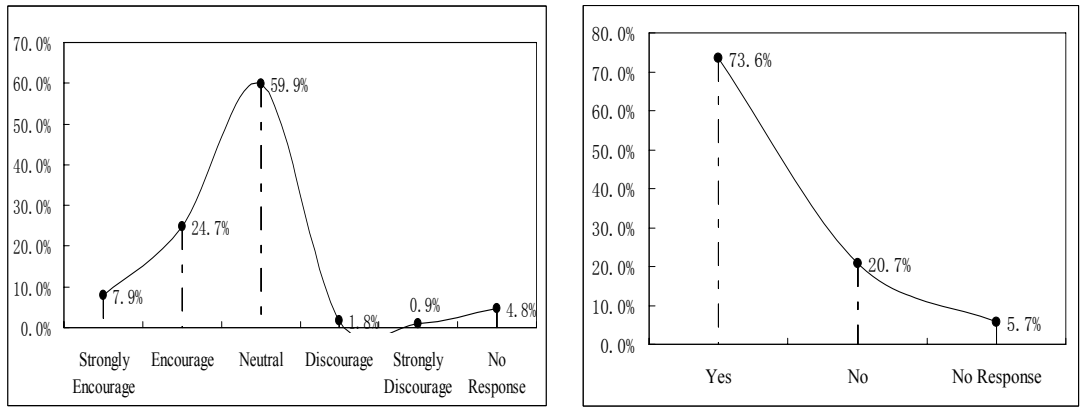


A. Distance similarities B. Gender similarities

Figure 6. Demographic similarities within walking/biking group

Within the walking/biking group, the majority of students live less than one mile from school, and most of the families have one or two children. Interestingly, the sample size of the families with more than three children is very small even among all the survey data. This explains the large percentage difference between small size families and large size families. There is no significant difference between males and females. It also shows that percentage of students at middle level grades (3^{rd}- 6^{th}) is slightly higher than other grade level.

Figure 7.A shows that more schools hold a positive attitude toward walking/biking within the walking/biking group. The student's attitude was measured by the question on if the students have ever asked for parents' permission for walking/biking to school. Figure 7.B shows majority of parents' answer to this question is yes for the walking/biking group. In addition, except for those who did not answer the question, all other people hold non-negative views on health factors; and most of them hold non-negative attitudes on enjoyment factors.



A. School attitudes similarities B. Student's attitudes similarities

Figure 7. Subjective Similarities within Walking/Biking Group

Figure 8 shows the rankings of the issues affecting parents' decision towards children's walking/biking (within the walking/biking group). It shows that distance is still the primary concern among parents within this group, followed closely by crossing guards and sidewalks (safety related issues). This indicates that the current SRTS engineering program (installing more crossing guards, building more sidewalks, and reducing traffic speed) would have a great effect on increasing the number of students walking or biking to school.

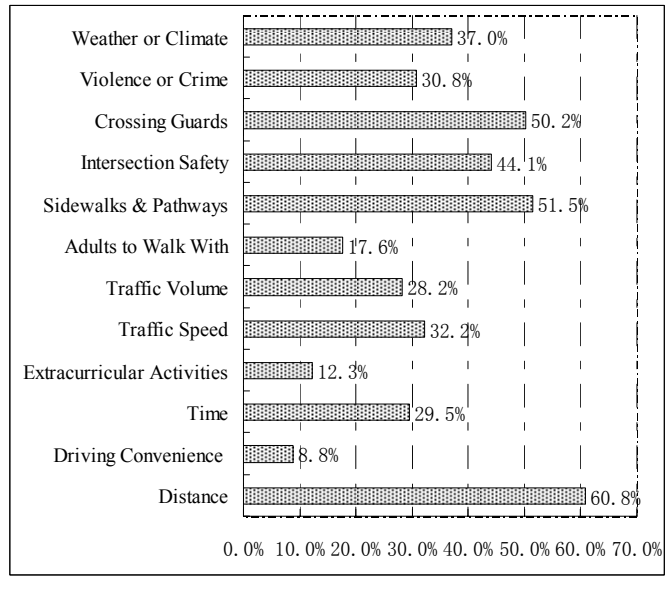


Figure 8. Rankings of the Issues Affecting Parents Decision towards Children's Walking/Biking (within the Walking/Biking Group)

Conclusions

This paper presents a novel diagnosis approach for efficiently identifying the significant factors that affect students' walking/biking rate from different perspectives. Multiple perspectives analysis proved to be effective to mathematically search the necessary and sufficient significant factors on walking/biking rate at different levels. Specifically, distance is the most significant factor and the largest barrier for the long distance interval group. For short distance interval, safety issues such as crime and intersection safety become the top concerns that prevent children from walking/biking to school. Subjective factors (school attitudes, allowable grade, and student's attitudes) are also significant factors. Improvement of sidewalk conditions also results in increasing the walking/biking rate. The analysis results provide deeper understanding of the barriers that result in the low walking/biking rates.

These findings are very helpful into developing guidelines of implementation of the SRTS program. It proves that the current SRTS program will significantly increase walking/biking rates by engineering, education and enforcement. During the last few years, many speed flashers, speed feedback signs, and sidewalk have been installed around the school areas to increase the traffic safety. The additional crossing guards and enhanced law enforcement have been used to strengthen the safety and security. The education and encouragement programs have been implemented to change some subjective factors such as student, parent and school attitudes to encourage students to walk/bike to school. All these SRTS improvements have been proven to be effective based on the analysis results of this SRTS survey.

References

Yarlagadda A. K, and Sivaramakrishnan, S. (2007). Modeling Children's School Travel Mode and Parental Escort Decisions. In *the 86th Annual Meeting of the Transportation Research Board*. CD_ROM. Transportation Research Board of the National Academies, Washington DC.

Mahalawat M., and Selvaraj Rayan, et al. (2007). Examination of Student Travel Mode Choice. In *the 86th Annual Meeting of the Transportation Research Board*. CD_ROM. Transportation Research Board of the National Academies, Washington DC.

Ewing Reid, Schroeer William, and Greene William. (2004). School Location and Student Travel-Analysis of Factors Affecting Model Choice. In *Transportation*

Research Record: Journal of the Transportation Research Board, No. 1895, TRB, Transportation Research Board of the National Academies, Washington D. C., pp. 55-63.

Parisi David, Hondorp Brett. (2005). Transportation Professionals Get Involved with Safe Routes to School, *ITE journal*, pp. 41-46.

McMillan Tracy E. (2007). The Relative Influence of Urban Form on a Child's Travel Mode to School. *Transportation Research*, part A, Vol. 41, pp. 69-79.

Dill Jennifer, Voros Kim. (2007). Factors Affecting Bicycling Demand: Initial Survey Findings from the Portland Region. In *the 86th Annual Meeting of the Transportation Research Board*. CD_ROM. Transportation Research Board of the National Academies, Washington DC.

David W. Hosmer, Stanely Lemeshow. (1989). *Applied Logistic Regression*. New York & Sons, Inc. (US).

Cheryl Allen-munley, Daniel Janice. Urban Bicycle Route Rating Model Application in Jersey City, New Jersey. (2006). ASCE *Journal of Transportation Engineering*, vol. 132, No. 6,, pp. 499-507.

Survey about Walking and Biking to School – for Parents. http://www.saferoutesinfo.org/resources/collateral/SRTS_Parent%20Survey_scan2007.pdf.

Safe Routes to School Student Arrival and Departure Tally Sheet. http://www.saferoutesinfo.org/resources/collateral/TravelBehaviorIn-ClassTallySheet_scan2007.pdf

Modeling Destination Choice for the Intermediate Stop Trips Included in the Commute Activity Patterns

Min Yang[1], Wei Wang, Xuewu Chen and Yingxiang Yang

[1] Ph.D, lecturer, Jiangsu Provincial Key Laboratory of Transportation Planning and Management, Southeast University, China; Email: transtar2002@163.com

Abstract

Using the travel-activity data collected from Shangyu, China, this paper develops multinomial Logit model with aggregate alternatives, including multiple size variables in representation for intermediate stop destination choices in the commute activity patterns. We use the amounts of facilities with different land use in the surveyed destinations as the size variables. Apart from the zonal attractions (size variables), locational factors and OD matching characteristics, Hanson accessibility are introduced into the models. The empirical results show that the commercial establishments are most likely to be chosen by intermediate stopping. And CBD area offers more opportunity to the intermediate stops. Moreover, the results show that the destination containing maintenance/leisure places compete with those nearby to win more intermediate stopping opportunity, which makes it more attractive in Shangyu, China. The findings of this paper provide some useful insights into the spatial activity-travel behaviors in the context of China.

Key words: Destination choice; multinomial Logit model; intermediate stop; size variable

Introduction

Destination choice model could answer the question that how people choose the trip destination under the drive of some specific activities, so it can be used to analyze the trip distribution. McFadden initiated the idea that turned the economic theory of

choice behavior into the practical model to analyze the suitable residential zones (1978). Daly applied improved Newton-Raphson algorithm to calibrate choice models containing attraction variables (1982). Ben-Akiva listed the empirical studies to estimate destination and mode choice models simultaneously by using nested Logit model (1985). Different from the gravity model that takes the traffic zone as the unit, destination choice model belongs to discrete category, which fully considers the individual's choice behaviors. And it introduces multiple size variables to display the destination's attraction degree while gravity model has single attraction variable. Also it brings in some dummy variables to show us how the land use pattern influences the trip distribution. As a result, the destination choice model could better explain the choice behavior mechanism.

With the further study of activity based travel demand forecasting techniques, greater concerns are given on the trip characteristics of different activity patterns, with these activities constrained within the temporal and spatial framework. To be specific, the activity destination model with spatial correlation and constraints could illustrate the influence from different activity patterns to trip distribution. With regard to the commute activity, the primary activity is work, while other flexible activities, i.e., intermediate stop activities, are restrained by the work destination, schedule and the occupations of workers. Bhat estimated the destination models of home based work trip and shopping/personal business trip (1998). Kitamura studied how the activity period of time, lasting time and time cost between home and destination affected the destination choice (1998). Bowman estimated destination and mode choice model at the same time based on the data collected from 1991 Boston travel survey (2000). Pozsgay studied the influence on choosing entertainment destination from effects of LOS, zonal attributes, trip attributes and social-demographic characteristics (2001). Chow and Li built the destination choice models respectively for three income groups (high, middle and low) of residents in Broward, Palm Beach, Volusia towns of Florida (2005). Yao and Guan gave a method for calculating the trip efficiency and the effect of traffic zones combined with a destination selection model based on disaggregate theory for trip generation in Beijing (2008). However, the destination choice behaviors of the intermediate stop trips included in the commute activity patterns are still not clear in China.

The object of this paper is to formulate and estimate destination choice model for the intermediate stop trips included in the commute activity patterns that will facilitate the replacement of aggregate trip distribution model in the context of a Chinese city. The model will comprehensively consider zonal attractions (size variables), locational factors, OD matching characteristics and Hanson accessibility.

Destination Choice Model for Mulation

The individual's destination choice could be considered as a specific traffic zone, nonetheless, this choice is the very establishment within the zone that can satisfy his activity demand. We call the establishment that the decision maker directly chooses as elemental alternative, and the aggregate alternative is the grouping of elemental alternatives in each zone. If the decision maker chooses any alternative, the aggregate alternative that contains the elemental alternative is also chosen accordingly (Ben-Akiva, 1985; Guan, 2004).

Denote the random utility of an aggregate alternative i to decision maker n by U_{in} and assume that:

$$U_{in} = V_{in} + \varepsilon_{in} \tag{1}$$

where V_{in} and ε_{in} are respectively the systematic and random components of the utility for aggregate alternative i and decision maker n. Define the aggregate alternative's systematic utility by:

$$V_{in} = \sum_{k=1}^{K'} \theta_k x_{ink} + \mu' \ln M_i \tag{2}$$

where K' is the number of unknown coefficients of the non-size attribute variables. θ_k is the coefficient of the k-th attribute and $\ln M_i$ is a measure of the "size" of aggregate alternative i. μ' is a scale parameter.

Actually we hardly know how many elemental alternatives are in any aggregate alternative. As for destination choice, we cannot clearly find out how many specific establishments in any traffic zone. Definite measure of variable $\ln M_i$ in the utility Eq. (1) is often unobservable. Fortunately, the introduction of size variables is able to solve the estimation problem of $\ln M_i$, shown as follow:

$$V_{in} = \sum_{k=1}^{K'} \theta_k x_{ink} + \mu' \ln M_i = \sum_{k=1}^{K'} \theta_k x_{ink} + \mu' \ln \left(\sum_{k=K'+1}^{K} \theta_k x_{ink} \right) \tag{3}$$

where K is the number of size variables, $\theta_{K'+k}$ is the parameter of the k-th size variable, x_{ink} is the value of the k-th attribute variable for destination alternative i and decision maker n.

The probability that n-th decision maker chooses destination alternative i can be expressed as:

$$P_{in} = \frac{\exp(V_{in})}{\sum_{j \in A_n} \exp(V_{jn})} = \frac{\exp\left\{ \sum_{k=1}^{K'} \theta_k x_{ink} + \mu' \ln \left(\sum_{k=K'+1}^{K} \theta_k x_{ink} \right) \right\}}{\sum_{j \in A_n} \exp\left\{ \sum_{k=1}^{K'} \theta_k x_{ink} + \mu' \ln \left(\sum_{k=K'+1}^{K} \theta_k x_{ink} \right) \right\}} \tag{4}$$

From Eq. (3), it is clearly that size variables yield a Logit model that is nonlinear with a series of attribute variables. However, for the nonlinear form, the traditional Newton Raphson algorithm which is applied to estimate the general linear Logit utility function cannot always work effectively. Fortunately, Daly proposed an improved NR algorithm to guarantee the astringency of the new model (1982). In this paper, models are calibrated according to Daly's method.

Data Survey and Processing

The city selected for this study is Shangyu, a small city located in the middle of the Yangtse River delta region in China, with a population of 204,900, and covers an area of 111 km^2. The travel diary survey had three distinct sections (Yang, 2007):

(1) Household characteristics: including residential location, household size, number of household members with jobs, number of children under education age, identification of household vehicles (including number of cars, motorcycles and bicycles) and annual household income.

(2) Individual socio-demographics: including questions designed to classify the household members according to the following aspects: gender, occupation, age, educational level and possession of a driving license.

(3) Travel-activity attributes: including trip sequence, start and end times, origin and destination, mode, trip purpose (containing activity type) and facility type.

We select commute activity patterns, in which work is the primary activity among the sets of all patterns. Meanwhile, extract data of intermediate stops from the commute activity patterns, including trips derived from maintenance activities and leisure activities. The size variables measuring the attraction of traffic zones are the employments. However, the data is difficult to get in China and a complete survey costs a lot. To solve this problem, in the travel survey of Shangyu, we added some questions about the activity facility types according to "the Classification Standard of the City Land-use" instituted by Chinese Construction Ministry. Based on the travel-activity survey results, various facilities in destination zones are added as proxy size variables according to activity types. The obtained data set to model destination choice for the intermediate stops contains 648 cases. The destination alternatives are selected from 22 traffic zones in Shangyu.

Destination Choice Model for the Intermediate Stop Trips

Table 1 gives the specific variables of the intermediate stop destination choice model, including zonal attractions (size variables), locational factors, OD matching

characteristics, Hanson accessibility and the impedance. The size variables of the zonal attractions contain four kinds: administration, commercial, education and other facilities.

Meanwhile, the OD matching characteristics not only include the matching between the intermediate stop and the living, origin places, but also contain the intermediate stop location matching with the working place, so that the restriction from work activity could be reflected.

In addition, The Hanson accessibility is used to analyze the "competition" and "aggregation" of the facilities for the intermediate stop activity like the shopping malls. Fotheringham and Bhat have discussed the impact of the "competition" and "aggregation" of the shopping facilities to the shopping destination choice (Bhat, 1998). The shopping destination probably combines with others nearby to form an aggregated area in order to enhance its attraction. On the other side, it may compete with its neighbor shopping malls to increase its chosen probability. The "competition" and "aggregation" not only affect the shopping destination choice, but also work on other maintenance/leisure destination choices.

Table 1. Explanatory variables of intermediate stop destination choice model.

Variable	Name	Description	Data source
Zonal attractions (size variables)	EMB1–EMB22 EMJ1–EMJ22 EMX1–EMX22 EMO1–EMO22	Administration Commercial Education Other	The accumulative amounts of facilities of each land use within the specific zone
Locational factors	CBD	Central business district	Including CBD zones 3 and 5
	Center	Administration center	Including zone 10
OD matching Characteristics of the intermediate stop location and the residential, working and origin places	RS	Whether the residential and the intermediate stop locations are in the same zone	If same, 1; otherwise, 0
	WS	Whether the working and the intermediate stop places are in the same zone	If same, 1; otherwise, 0
	OS	Whether the origin and the intermediate stop locations are in the same zone	If same, 1; otherwise, 0
Hanson accessibility	M	Measure the competition and aggregation of the destination	Computation from Eq. (5)
Impedance	IMP1–IMP22	Auto travel time (s)	Calculated using traffic assignment by TranStar

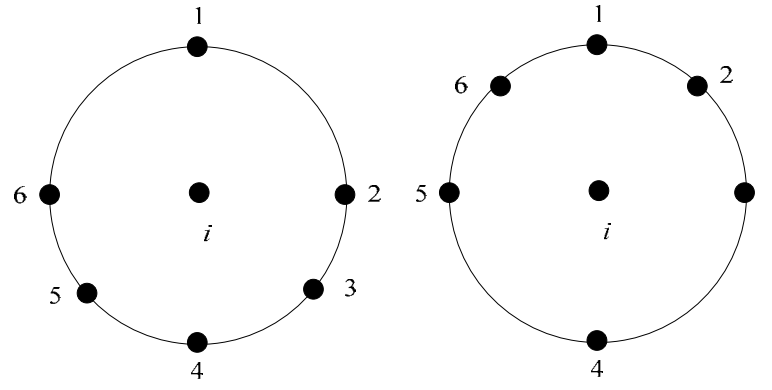


Figure 1. Competition and aggregation of maintenance/leisure facilities.

It can be seen from Figure 1, zone 1 might compete with zones 2 to 6 nearby to increase its own chosen probability and also might aggregate with zones 2,3,5,6 to strengthen its attraction. The "competition" and "aggregation" are in positive correlation with the number of maintenance/leisure facilities, while in negative relation with the sum of impedance from the specific zone to the ones nearby. The Hanson accessibility function can be expressed as follow:

$$M_j = \left[\frac{1}{L} \sum_{l=1}^{L} \left(\frac{\log R_l}{\log H_{ij}} \right) \right] \tag{5}$$

where M_j is the measurement of Hanson accessibility. R_l is the number of maintenance/leisure facilities of zone l close to zone j. H_{ij} is the impedance from zone j to zone l. L is the total number of zones near to zone j.

Table 2 shows the parameters and their corresponding T values of the intermediate stop destination choice model.

(1) Locational Factors (Size variables): The parameters of EMB (administration), EMJ (commercial), EMX (education), EMO (other) are 1, 4.29, 2.96, 2.76 respectively, among which the value of EMJ is the biggest but that of EMB is the smallest. This result accords with the fact that commercial facilities attract a great number of maintenance/leisure activities, which have a great tendency to generate intermediate stops. This is in accordance with previous studies on home based shopping destination choice behaviors by Bhat and others. However, the administration facilities have less attraction to the intermediate stops, possibly because they mainly attract work activities. Notably, the education facilities have a great attraction to the intermediate stops. It is the same with the expectation that the parents usually send or pick up their children to school during commuting, which also reflect the tendency of commute trip chaining.

Table 2. intermediate stop destination choice model estimation results.

No.	Variable	Description	E	T
1	EMB	The number of administration establishments	1	-
2	EMJ	The number of commercial establishments	4.29	5.41
3	EMX	The number of education establishments	2.96	5.38
4	EMO	The number of other establishments	2.76	5.18
5	Mu	Scale parameter	0.95	9.95
6	CBD	Dummy variable: if the zone is CBD, 1; otherwise 0	0.20	1.72
7	Center	Dummy variable: if the zone is the administration center, 1; otherwise 0	0.13	1.68
8	RS	Dummy variable: if the residential and intermediate stop places are in the same zone, 1; otherwise 0	1.50	10.31
9	WS	Dummy variable: if the working and intermediate stop places are in the same zone, 1; otherwise 0	0.58	4.47
10	OS	Dummy variable: if the origin and intermediate stop places are in the same zone, 1; otherwise 0	–0.72	–3.70
11	M	Hanson accessibility	–2.80	–1.86
12	IMP	Auto travel time (s)	–6.88e-3	–8.50
Log-likelihood at zero: -1825.58				
Log-likelihood at end: -1594.87				
Rho-square (ρ^2): 0.204				
Adjusted rho-square ($\overline{\rho}^2$): 0.198				

(2) The Mu value is 0.95, close to 1, which means the model is independent with the division of the traffic zones.

(3) Locational factors: The parameter of variable "CBD" is positive, so CBD area has strong attraction to the intermediate stop activities. This is because plenty of commercial infrastructures located in CBD area offer more stopping opportunities. The parameter of "Center" variable is also positive, which means the administration center are also attractive to the intermediate stops because some matching commercial infrastructures are located there.

(4) OD matching characteristics of the intermediate stop destination with the residential, working and origin places: The parameters of RS, WS are positive, while that of OS is negative. This reveals that commuters prefer to choose the intermediate stop destination which is close to residential or working place but do not so care where the origin is. Furthermore, the choice of intermediate stop locations included

in the commute activity patterns, is not only influenced by the residential places but also influenced by the completion of work, which is the primary activity in the commute activity pattern. In addition, the result reflects that the working place has some restrictions to the choice of the intermediate stop destination, but "RS" is bigger than "WS", so the individual shows a greater tendency to make an intermediate stop near his residential location than working place.

(5) Hanson accessibility: The parameter of M is negative, which suggests "competition" is more significant than "aggregation" to affect the intermediate stop destination choice in Shangyu city. As a result, the destination increases its attraction by competing with the places nearby. Results show that people in Shangyu city are more likely to choose dispersed intermediate stops, which may because that these dispersed locations are closer to their residential or working places.

(6) The impedance: IMP is negative, which means the bigger the IMP becomes, the smaller the attraction of the destination and the chosen probability will be.

Conclusions

In this paper, we explore the intermediate stop destination choice model with aggregate alternatives by introducing multiple size variables, using the travel-activity data collected from a Chinese city. Since the amounts of employment opportunities in urban China could not be obtained conveniently under current situations, we use the amounts of facilities of each land use pattern in the surveyed destinations as the size variables. Zonal attractions (size variables), locational factors, OD matching characteristics and Hanson accessibility are found to be the significant factors influencing destination choice for the intermediate stops. Major findings are as follows: Fist, commercial facilities attract a great number of maintenance/leisure activities, notably, the education facilities have a great attraction to the intermediate stops. Second, CBD area has strong attraction to the intermediate stop activities. Third, commuters prefer to choose the intermediate stop destination which is close to residential or working place but do not so care where the origin is. Four, people in Shangyu city are more likely to choose dispersed intermediate stops according to Hanson accessibility analysis. The utility function estimated by the model can be applied in a straightforward way to obtain discretional trip interchanges by aggregating individual's choice probability.

Acknowledgement

This research was founded by the National Natural Science Foundation of China (50738001, 50908052) and the National Basic Research Program (2006CB705501).

References

McFadden, D. (1978). "Modeling the Choice of Residential Location." In *Transportation Research Record: Journal of the Transportation Board*, No. 673, Transportation Research Board of the National Academies, Washington, D.C., 531-552.

Daly, A. (1982). "Estimating choice models containing attraction variables." *Transportation Research Part B*, Vol.16, pp. 5-15.

Ben-Akiva, M., and Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, Massachusetts: MIT Press, 253-275.

Bhat, C. R., Govindarajan, A., and Pulugurta, V. (1998). "Disaggregate attraction-end choice modeling: formulation and empirical analysis." *Journal of the Transportation Research Board*, 1645, 60-68.

Kitamura, R., Chen, C., and Narayanan, R. (1998). "The effects of time of day, activity duration and home location on travelers' destination choice behavior." *Journal of the Transportation Research Board*, 1645, 76-81.

Bowman, J. L., and Ben-Akiva, M. (2000). "Activity-based disaggregate travel demand model system with activity schedules." *Transportation Research Part A*, 35, 1-28.

Pozsgay, M. A., and Bhat, C. R. (2001). "Destination choice modeling for home-based recreational trips; analysis and implication for land-use, transportation, and air quality planning." *Journal of the Transportation Research Board* 1777, 47-54.

Chow L. F., Zhao F., and Li M. T. (2005). "Development and evaluation of aggregate destination choice models for trip distribution in Florida." *Journal of the Transportation Research Board*, 1931, 18-27.

Li M. T., Chow L. F., and Zhao F. (2005). Application of Geographically Stratified Importance Sampling in the Calibration of Aggregated Destination Choice Models for Trip Distribution. *Journal of the Transportation Research Board*, No. 1935, pp. 85-92.

Yao L. Y., Guan H. Z., and Yan H. (2008). "Trip generation model based on destination attractiveness." *Tsinghua Science and Technology*, 13(5), 632-635.
Yang M., Wang W., Chen X.W., et al. (2007). "Empirical analysis of commute trip chaining: case study of Shangyu, China." *Journal of the Transportation Research Board*, 2038, 139–147.

Guan, H. Z. (2004). *Discrete Choice Model- Tool for Travel Behavior Analysis*. China Communication Press, Beijing, 26-60.

Analysis on Effective Information and Travel Activity Returning Home by Car under Downpour

Jun Sakamoto[1] and Motohiro Fujita[2]

[1] Assistant Professor, Gifu National College of Technology, 2236-2 Kamimakuwa, Motosu, Gifu, 501-0495, JAPAN, Email: sakamoto@gifu-nct.ac.jp
[2] Prof., Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, Aichi, 466-8555 Japan, Email: fujita.motohiro@nitech.ac.jp

Abstract

The way of providing information for reducing disaster damage has been more important in Japan. By providing appropriate information to citizens under downpour, citizens can understand disaster situation and behave appropriately. Therefore, this paper focuses on the use of privately owned cars under downpour which is often brought up the deteriorating factor of traffic jam under downpour, and examines the way of providing traffic & weather information for reducing car use under downpour through the development of behavior model of returning home and other analyses. As a result, it found that the effect factor to driver consciousness of returning home under downpour.
Key Words: downpour; effective information; behavior models of returning home; paired comparison method

Introduction

The way of providing information for reducing disaster damage has been more important in Japan. For example, the local government lessons of the downpours on Sep.11 and 12, 2000 at Tokai region, Japan (Tokai downpour) is as follows. (a) They couldn't get downpour information immediately. (b) They couldn't utilize the downpour information from citizens. (c) They couldn't provide sufficient downpour information. (d) Many citizens didn't know disaster prevention information such as hazard point due to lack of daily disaster education.

With these points as background, this paper focuses on "How to drive car under downpour" which is one of traffic jam cause. Because of that many citizens drive car carelessly under downpour, there is a possibility that emergency vehicle such as

ambulance is involved heavy traffic jam. For this reason, this study analyzes the changes in driver behavior and traffic & weather information at returning home by car, and examines the provision of information method for that recognition of current situation and encouraging appropriate behavior for drivers under downpour.

In addition, this paper focuses on not evacuation behavior but returning home under downpour. In other words, it analyzes not the situation of that citizens have to evacuate due to severe downpour disaster, but the situation of that drivers can choose returning home or not due to immediately after downpour disaster. Therefore, the feature of this study is to analyze the relationship between traffic & weather information and the driver behavior at returning home under downpour.

The rest of this paper is organized into six sections. The next section introduces survey outline for this study analysis. In section 3 ~ section 5 analyze the provision of information impact based on information content and the timing of providing information under downpour. In section 6 develops the behavior model of returning home, and analyzes that driver's attribute which affects the behavior of returning home and the effect of information. In section 7 analyzes the relative importance between traffic & weather information by paired comparison analysis.

Survey Outline

The questionnaire for this study was conducted on February - April 2006. The object person of the survey is the Nagoya city resident who commutes by car. The number of distribution is 1,628, and the number of collection is 259 (The collection rate is 15.9 %). The survey outline is shown in Table 1. The main questionnaire contents are follows.

- Usual situation of returning home.
- The consciousness of returning home based on information content and the timing of providing information under downpour.
- The relative importance between traffic information and weather information.
- Respondents attribute (Gender, Age, and Job etc.).

Table 1. Survey outline.

Gender			Age			Job		
Male	197	81%	10s	0	0%	Independent business	31	13%
Female	46	19%	20s	6	2%	Company employee	142	58%
			30s	29	12%	Government employ	29	12%
			40s	37	15%	Full-time homemaker	2	1%
			50s	88	36%	Student	1	0%
			60s	61	25%	Part-time worker	3	1%
			70s ~	22	9%	Disemployment	5	2%
						Others	30	12%
VICS			**Living with child or elderly**			**Disaster experience**		
Mount	140	58%	Yes	88	37%	Yes	134	56%
Unmount	100	42%	No	148	63%	No	105	44%

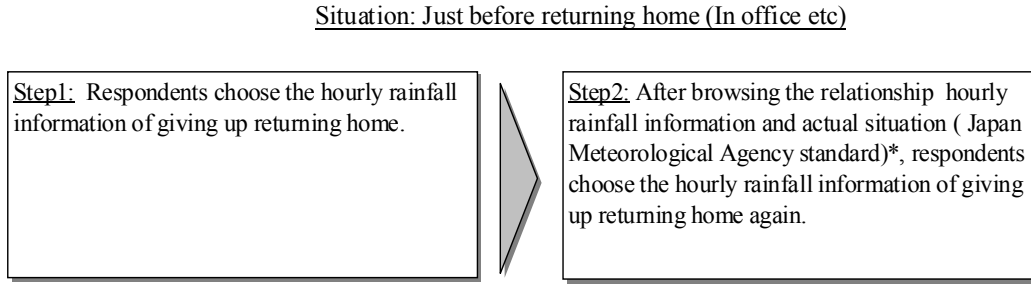


Figure 1. Question about rainfall information.

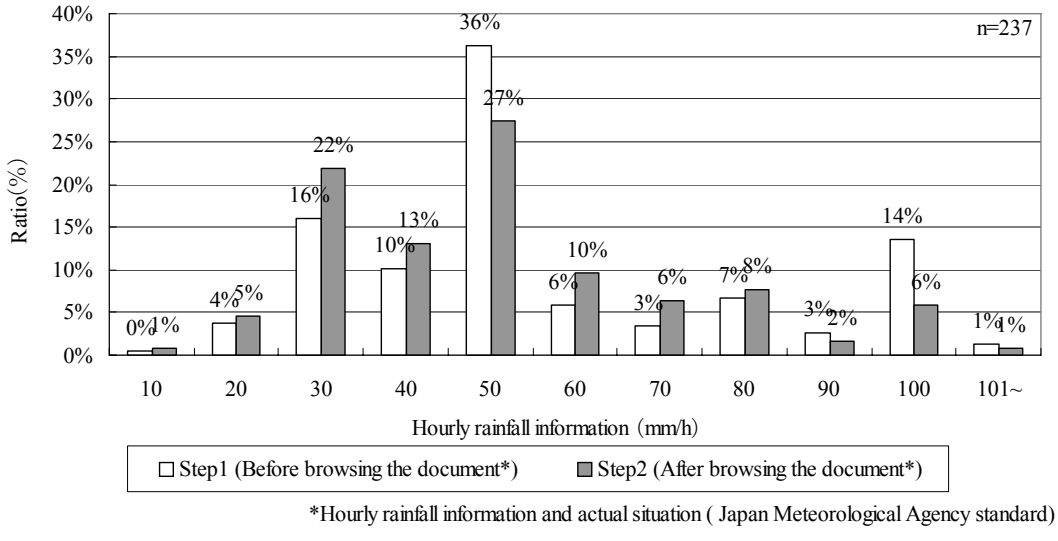


Figure 2. Result about rainfall information.

Analysis on Driver Comprehension about Rainfall Information

Usually, we frequently hear the rainfall information such as "30 mm/h". But some scholar points out that it is not easy for person to image the relationship between rainfall information and actual situation. Therefore, this chapter analyzes the driver consciousness of returning home based on rainfall information for examining above indication in detail.

The question is shown in Figure 1. It asks respondents hourly rainfall information of giving up returning home and getting more information about rainfall in case that downpour happens before returning home. The way of response is 2 steps (See Fig.1). Japan Meteorological Agency standard, it explains the relationship weather terminology and actual situation, for example, 20~30mm/h is the situation of "Getting wet even if putting up umbrella," 30~50mm/h is the situation of "Rainfall as if overturning a bucket".

The result of reply is shown in Figure 2. The ratio of choosing 30~50mm/h is higher than others in both steps. And there is statistically significant difference by t-test (t-statistic is 4.40>1.96). Therefore, there is gap between step1 result and step 2 result, and it shows that the hourly rainfall information of step 2 result is lower than the step1 result one.

Table 2. Relationship between the hourly rainfall information of step1 and step2.

Step1 (mm/h) \ Step2 (mm/h)	10	20	30	40	50	60	70	80	90	100	100-
10											
20		2%	1%								
30		2%	9%	2%	3%						
40			5%	3%	3%						
50			5%	6%	18%	3%	2%	2%			
60				1%		4%					
70					1%		2%				
80					1%	1%		3%			
90								1%	1%		
100			1%	1%	2%	2%	2%	1%		4%	
100-										1%	

※Less 1% is blank

The relationship between the hourly rainfall information of step1 and step2 is shown in Table 2. It points out that the respondents who choose hourly rainfall information such as downpour in step1 tend to change their answer in step2. Especially, in case of that respondents choose over 100 mm/h, there is a trend. And although disaster risk increases in case of over 30mm/h according to Japan Meteorological Agency standard, a number of respondents choose over 50mm/h in step2. It points out that it is important to make respondents realize the relationship hourly rainfall information and actual situation.

Analysis on the Effectiveness of Providing Information on the Way Home

This chapter analyzes the driver consciousness of returning home by information on the way home under downpour, and examines the influence of information content.
The question is shown in Figure 3. The way of question is that respondents choose 1) Return home, 2) Evacuate safety point, and 3) Give up returning home by information A ~ S in case that downpour and severe traffic jam happens on the way home and driver get information A ~ S at separately.

The result is shown in Figure 4. About rainfall information (information A~D), although the ratio of “Return home” is over 70 % in case of less 40mm/h, the ratio of “Evacuate safety point” and “Give up returning home” is tend to increase in case of over 60mm/h. Especially, 90mm/h, the ratio of “Give up returning home” is high (38 %).

About road traffic information (information E~O), information E~L, the ratio of "Evacuate safety point" and "Give up returning home" by the information "Many submerged car" is higher than others. It means that the way of providing specific information about the impact of downpour to driver is effective. Information M~O, the ratio of "Evacuate safety point" and "Give up returning home" is over 90 %, that is, the information is effective.

About time required information, there is a possibility that driver consciousness is different in case of their usual time required for returning home. Therefore, we recognize the trend of usual time required for returning home first. The result of its average is 33 minutes, so we classify the result about time required information into group 1 (less 30 minutes, n=116) and group 2 (over 35 minutes, n=97). The result is shown in Figure 5. It shows that the ratio of "Return home" by four or five times than usual is about 50 % in group 1, but the ratio of "Give up returning home" by four or five times than usual is about 70 ~ 80 % in group 2. Therefore, it shows that the effectiveness of providing time required information is affected by driver's usual time required for returning home.

Question: What do you act when downpour and severe traffic jam happens on the way home and you get information A ~ S at separately. Please select one of the 3 choices "1) Return home, 2) Evacuate safety point, 3) Give up returning home (Return office) ". (All the public transport are not running, and it is easy to return office.)

Information category	Information content	Choice
Rainfall information (Hourly rainfall)	A: 20mm/h	1) Return home 2) Evacuate safety point 3) Give up returning home (Return office)
	B: 40mm/h	
	C: 60mm/h	
	D: 90mm/h	
Road traffic information	E: Traffic jam in the city	
	F: Many floods in the city	
	G: Many road blocked in the city	
	H: Many submerged car due to floods in the city	
	I: Traffic jam ahead	
	J: Too many floods ahead	
	K: Many road blocked ahead	
	L: Many submerged car due to floods ahead	
	M: Danger, evacuate to safer places	
	N: Danger ahead, evacuate to safer places	
	O: Turn back	
Time required information	P: Twice than usual	
	Q: Three times than usual	
	R: Four times than usual	
	S: Five times than usual	

Figure 3. Question about information on the way home.

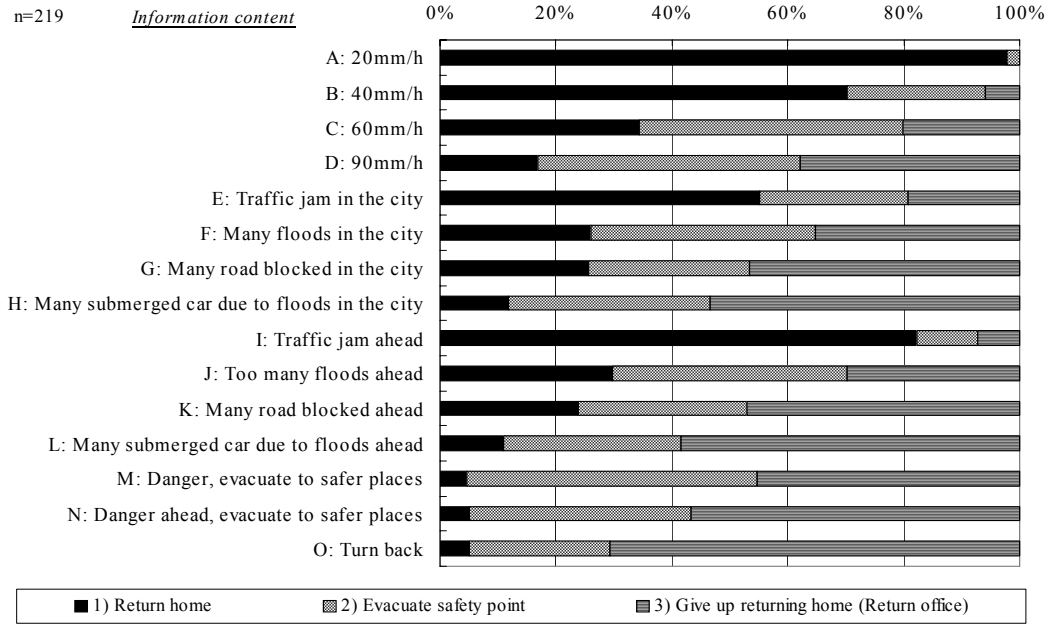


Figure 4. Result about rainfall & road traffic information on the way home.

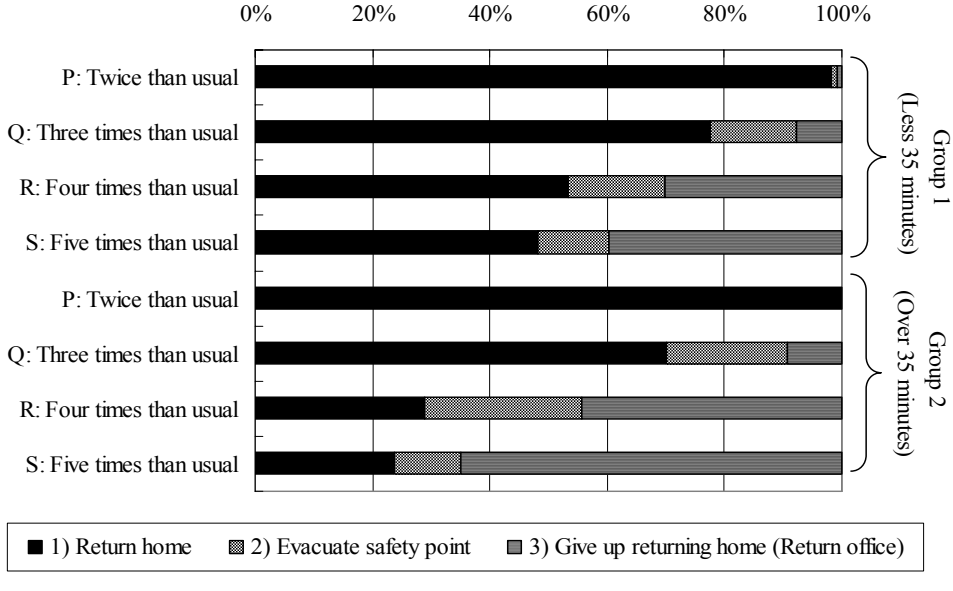


Figure 5. Result about time required information on the way home.

Development of Behavior Model of Returning Home

This chapter develops the behavior models of returning home by Binary Logit Model for analyzing the relationship between the driver consciousness of returning home and information in more detail. The models are 2 types. One type is for analyzing driver attribute and information (Driver attribute model). Another type is for analyzing a lot of information about downpour and driver behavior (Corresponding information model).

Table 3. Driver attribute model

Independent variable	Parameter	t-statistic
Constant terms	-0.623	-1.15
Job	1.235	2.51*
Time required for returning home	-1.341	-1.69**
Peak rush hour	-0.965	-2.01*
Live with elderly or child	1.147	2.79*
Experiencing downpour	-0.481	-0.78
Initial logarithm likelihood	-86.046	
Converged likelihood	-73.551	
McFadden's Rho Squared	0.145	
McFadden's Adjusted Rho Squared	0.110	
Sample size	126	

Driver Attribute Model. About independent variable is a) job (Public employee or company employee = 1, others = 0), b) time required for returning home, c) the hour of usual returning home is peak rush hour or not (17:00~19:00 = 1, others = 0), d) living with "elderly or child" or not (If respondent lives with over 65 years elderly or less 7 years child, it is 1, others are 0), and e) experiencing downpour or not before (If respondents experienced Tokai downpour, 2000, and they didn't go home, or they had taken over 3 hours to go home then, it is 1, others are 0). About dependent variable is information F "Many floods in the city" in Figure 4. The result of "Return home" sets 1, "Give up returning home" sets 0, and the result of "Evacuate safety point" is deleted for improving model's reliability.

Driver attribute model is shown in Table 3. About usual returning home, "Time required for returning home" is -1.341 and "Peak rush hour" is -0.965. It means that the drivers who usually return home in a short time at off-peak period tend to return home under downpour. About driver attribute, "Job" is 1.235 and "Live with elderly or child" is 1.147. It means that the drivers who is not "public employee or company employee" and don't live with "elderly or child" tend to return home under downpour. Therefore, this model can explain the relationship between the driver behavior of returning home and driver attribute by traffic information under downpour.

Corresponding Information Model. About data formation is based on 259 samples (the number of collection, See chapter 2). The result of a respondent is set 19 samples (Information A–S). But "Information P–T" transfer an information content. And "Information A" is deleted because that no one chooses "Return home". About independent variable is driver attribute and traffic & weather information under downpour. The variable of driver attribute is set "Peak rush hour" for considering the influence of peak period. The variable of traffic & weather information under downpour is set dummy variable in case of driver behavior by information B–O.

Time required information (Information P–T) is the thing that the half of time required for returning home by 2–5 (Responding information P–T), because that the questionnaire asks respondents their behavior by information on the way home. And the sample of choosing "Return home" by the half of usual time required for returning home add the data because the object of model generalization. About dependent variable is driver behavior of returning home by "Information B–S" in Figure 5. The result of "Return home" sets 1, "Give up returning home" sets 0, and the result of "Evacuate safety point" is deleted for improving model's reliability.
Corresponding information model is shown in Table 4. About driver attribute, "Peak rush hour" is –0. 176. It means that drivers who usually return home at off-peak period tend to return home under downpour. About rainfall information (Information B~D), the more hourly rainfall is increasing, the more the probability of "Give up returning home" is increasing. About road traffic information (Information E–O), the information of the giving up returning home probability are "I: Traffic jam ahead (–5. 793)," "J: Too many floods ahead (–6. 137)," and "O: Turn back (–5. 916)". About time required information, the more the time is increasing, the more the probability of "Give up returning home" is increasing.

Table 4. Corresponding information model.

Independent variable		Parameter	t-statistic
Constant terms		3.585	19.295*
Peak rush hour		-0.176	-1.85**
Rainfall information (Hourly rainfall)	B: 40mm/h	-1.020	-3.001*
	C: 60mm/h	-2.978	-11.461*
	D: 90mm/h	-4.326	-16.106*
Road traffic information	E: Traffic jam in the city	-2.474	-9.8*
	F: Many floods in the city	-3.791	-15.202*
	G: Many road blocked in the city	-4.140	-16.909*
	H: Many submerged car due to floods in the city	-5.029	-17.685*
	I: Traffic jam ahead	-5.793	-15.359*
	J: Too many floods ahead	-6.137	-17.046*
	K: Many road blocked ahead	-1.085	-3.427*
	L: Many submerged car due to floods ahead	-3.589	-14.357*
	M: Danger, evacuate to safer places	-4.223	-17.058*
	N: Danger ahead, evacuate to safer places	-5.210	-18.027*
	O: Turn back	-5.916	-16.332*
Time required information (minute)		-0.039	-14.566*
Initial logarithm likelihood	-1389.665		
Converged likelihood	-2234.956		
McFadden's Rho Squared	0.378		
McFadden's Adjusted Rho Squared	0.375		
Sample size	3280		

Paired Comparison Analysis between Traffic Information and Weather One

The intended information contents of this study are classified "1) Rainfall information," "2) Traffic and flood information," "3) Submerged vehicles information," and "4) Time required information" This chapter analyzes relative importance of above 4 information by paired comparison analysis.

The question is shown in Figure 6. It asks drivers their relative importance for determining their behavior by 4 information contents under downpour (4 information contents, 7 level evaluation, and six times). About Paired comparison analysis, this chapter adopts Nakaya's method that is developed based on Schffe's paired comparison.

The result is shown in Table 5–6. There is statistically significant difference of the importance of each information contents (Main effect S_a is 258. 69>1. 96), and the importance of each information contents is "2) Traffic and flood information (0. 666)" > "3) Submerged vehicles information (0. 425)" > "4) Time required information (-0. 366)" > "1) Rainfall information (-0. 726)". This result shows that it is effective to provide drivers specific information contents of downpour such as "Many submerged car," "Many floods" for determining their behavior by information contents under downpour.

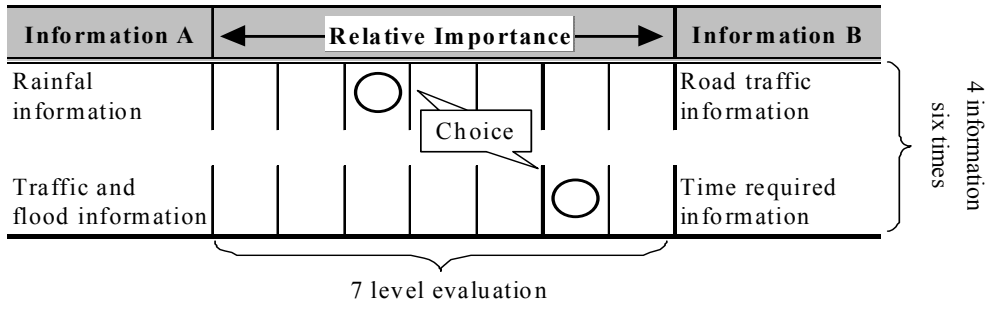


Figure 6. Question about relative importance.

Table 5. Result of paired comparison analysis (1).

Factor	Square sum	Degree of freedom	Unbiased variance	F
Main effect S_a	1166.88	3	388.96	258.69*
Main effect* individual $S_{a(B)}$	3961.12	678	5.84	3.89*
Combination effect S_γ	16.59	3	5.53	3.68*
Error S_ε	1019.41	678	1.50	
All the square sum S_T	6164.00	1362		

n=227 t=4

Table 6. Result of paired comparison analysis (2).

Information	Score X_i…	Priority weight $\alpha_i=x_i.../tN$
1) Rainfall information	-659	-0.726
2) Traffic and flood information	605	0.666
3) Submerged vehicles information	386	0.425
4) Time required information	-332	-0.366
Total	0	0

Conclusions

To summarize our interpretation of the results is as follows.

- Analysis on providing rainfall information, over 30 mm/h cause a high incidence of disaster, but many respondents answer that they return home even if over 50mm/h. In other words, it is not necessarily that drivers can imagine actual rainfall situation from rainfall information.
- Analysis on providing time required information, it revealed that the effectiveness of providing time required information is affected by driver's usual time required for returning home.
- Analysis on effectiveness of providing information during returning home, it revealed that the information of downpour influence such as "There are many submerged vehicles" is effective for driver behavior.
- As a result of developing the behavior model of returning home, it revealed that the relationship between information contents and driver's behavior and the relationship between returning home and driver attribute.
- Analysis on relative importance between traffic information and weather information, it revealed that the information of that it is easy to image downpour such as "Many submerged car" affect driver behavior in returning home.

References

Ushiyama, M. (1999). "A survey of understanding about precipitation information," *Proceedings of First Annual Meeting of Japan Society for Disaster Information,* 143-146.

Ushiyama, M. and Takara K. (2002). "A questionnaire how to collect storm and disaster information." *Annual Journal of Hydraulic Engineering, JSCE, Vol. 45,* 325-329.

Asada, J., Katada, T., Okajima, D. and Kobatake., S. (2001). "A study on the providing the information to make inhabitants understand the refuge information at the time of a flood." *Annual Journal of Hydraulic Engineering*, *JSCE,* 44, 37-42.

Fujita M., and Jun Mitamura. (2004). "Analysis on road traffic situation of travel times and routes of cars under the downpour disaster in Tokai area" *Proceedings of JSCE, 751 IV,* 127-137.

Jun Sakamoto, Motohiro Fujita, Koji Suzuki and Jun Mitamura. (2006). "Development of road flooding situations and animated traffic simulation under the heavy rainfall disaster." *Theory and Applications of GIS, GIS Association of Japan,* 14, 9-19.

Research Committee of Sensory Evaluation. (1973). "Sensory evaluation handbook," *Union of Japanese Scientists and Engineers*.

Behavioral Characteristics of Self-driving Tourists: A Case Study in Kunming

Baohong He[1], Jun Chen[2] and Wei Wang[2]

[1] School of Traffic Engineering, Kunming University of Science & Technology, Kunming 650224, China; Email: haoziyy@sina.com
[2] School of Transportation, Southeast University, Nanjing 210096, China

Abstract

With the advent of Leisure Economy Times and the popularization of cars, besides being used for daily commute and business travel, private cars are playing an important role in recreational activities on weekends and holidays, which is called urban self-driving tourism in China. To understand the behavioral characteristics of this type of tourists, we carried out a survey in Kunming, Yunnan Province, China. We found that it is the individual values and the life style rather than normal travel factors that were the determinants in self-driving tourism. Both travel purpose and travel radius were strong features of recreation. The destinations were generally within ReBAM (Recreational Belt around Metropolis) which was about 500km around the city. The travel distance had the "anti-attenuation" phenomenon. Concerns on transportation, safety and information were the three main discouragements to self-driving tourism.
Key Words: self-driving; travel characteristics; recreation; ReBAM

Research Motivation

The data from the National Bureau of Statistics shows that the number of car ownership reached 24.38 million by 2008. The number of private cars was 19.47 million and took up nearly 80 percent. That was equivalent to 1.5 cars per 100 people (the National Bureau of Statistics 2009). The survey on the travel preference during "National Day Golden Week" done by the National Holiday Office shows self-driving tourism accounted for 30.90% of the short distance recreational travel (which is defined as a trip within the same province) (the National Holiday Office 2008). In developed countries, self-driving tourism is very common and the studies focus on tourists' choices of travel behaviour and travel demand management. For example, Alan Shailes, taking a case of

Cornwall of UK, studied the tourists' attitude and travel strategy to congestion (Alan 2001). Yechezkel Israeli analyzed the influence of self-driving tourism on local transportation in Jerusalem and proposed tourist flow management measures to and within the historical towns (Yechezkel 2004). Janet E.Dickinson studied the nature of the transport problem by analysing residents'perspective and tourists' travel behaviour at a rural tourism destination in the UK (Janet E. Dickinson 2008). In China, the fast growth of self-driving tourism as a non-traditional tourism activity has stimulated research interest in two aspects. One is to, from the perspective of marketing, take self-driving tourism as a tourism product to analyze the differences in type, feature, and problem between the self-driving tourism and traditional travel (Zhai 2003; Chen 2004; Zhou 2005). The other is to take it as a tourism activity to analyze tourists' behaviour features, such as tourists' motive, travel preference, organization and tourism information in terms of ethology, sociology and psychology (Ye 2006; Zhang 2006; Feng 2008). On the whole, most domestic researchers regard it as a tourism activity in holidays and golden weeks and treat self-driving tourists as tourists. They use tourism behaviour analysis. Guan is the pioneer of the study of self-driving (Guan 2005). However, He et al (2005) found that, besides being used for commute and business travel, private cars were used substantially for self-driving tourism and concluded that self-driving tourism would become as common as it was in developed countries with the advent of Leisure Economy Times. Therefore, it is important to understand the behavioral characteristics of these tourists.

Method and Data

He did surveys in Nanjing (He,2005) and Kunming (2006, not published), and found that 58% drivers in Nanjing used a car more than 6 days a week(which could be roughly interpreted that more than a half of the drivers drove on weekends) While, in Kunming, the proportion was as high as 76%. Self-driving tourism has become the most important use only next to commute and business.

Considering the trip purposes--commute and business, many researchers analyzed car users' travel characteristics and the relationship among all kinds of travel modes. However, there are great variations in idea and environment choice of self-driving tourism which is regarded as a way of leisure. Are the findings on travel characteristics of cars in terms of travel modes helpful to understand self-driving tourism? If not, what features does self-driving tourism behavior have?

Based on what is mentioned above, in the sample, the respondents were car users rather than self-driving tourists. Therefore, the survey method was different from those on tourists in tourism destinations. The survey was done in ten automobile sales service shops (4s) in Kunming, in the form of personal interview. The questions involved car users' personal information, travel characteristics (travel purpose, distance, time and frequency) and subjective evaluation on travel. The survey spanned six weeks including two festivals (Labor Day and the Dragon Boat Festival) from April 24th to May 31st in 2009 and obtained 290 valid samples.

Characteristics of Self-driving Tourism

Basic information of drive tourist. Among the respondents, 77% are male, and 23% are female. The average age is 32.4, and 81% of them are from 20 to 40. 84% of them have an associate or higher degree. The average driving experience is 5.1 years, and 55% of them have 3 to 5 years experience. The average monthly income is RMB￥3947, and 32% of respondents' income is within two ranges: the first is from 2000 to 3500 Yuan, and the second is from RMB￥3500 to 5000. The distribution of income group is olive-shaped, and 7% respondents' income is less than RMB￥2000. Generally speaking, most of self-driving tourists are young people with higher income.

47% of the automobiles are economic sedan (which is defined as a car pricing less than RMB￥100,000). People with monthly income less than RMB￥5000 mainly own economic cars, especially the people with monthly income less than RMB￥2000. 76% of the people with monthly income less than RMB￥2000 own economical cars while it is only 31% of the people with income more than RMB￥5000. This suggests that the income plays a more important role in deciding what type of a car to buy than in deciding whether to have tourism trip.

Moreover, civil servants, public institution employees, and professionals are roughly evenly distributed and no centralization is found in occupation distribution.

Travel purpose is recreation. The survey shows 46% self-driving tourists are for the purpose of spending a holiday, 26% for going sightseeing, 9% for visiting friends and relatives, 5% for enriching experience and 4% for others (see Figure 1). It is clear that recreation is a strong motive for unban self-driving tourism. This seems to be different from the finding which shows self-driving tourists' basic consumption pattern is going sightseeing (Feng 2008).

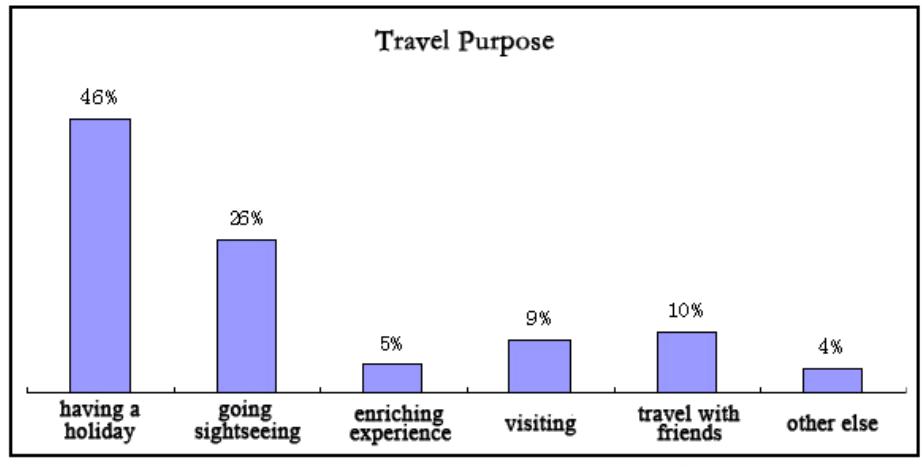


Figure 1. Distribution of travel purpose of self-driving tourism.

The studies in the field of tourism show that sightseeing plays the leading role when the per-capita GDP reaches US$1000, however, it is leisure which meets people's multiple needs that plays the leading role when the per-capita GDP reaches US$2000 and

recreational tourism when US$3000. According to "the 2008 National Economic and Social Development Report", by the end of 2008, the per-capita GDP of Kunming had reached RMB￥25826 which equals to about US$3700 (Kunming Statistical Bureau, 2009). This also indicates the advent of Leisure Economy in Kunming. As urban consumer groups, self-driving tourists' purpose of sightseeing has gradually turned to recreational tourism. The difference is that travel destinations for sightseeing is clear and driving is just a travel mode while drive for recreation is not only a purpose but also a kind of consumer behavior. Because there have been great variations in idea and environment choice, there will be qualitative changes in the cars' own characteristics.

Tourists' travel distance appears the "two-way attenuation" phenomenon. Four questions are asked for analyzing and depicting the distribution of travel distance and the general situation of self-driving tour: the times of self-driving tourism out of the province, the provinces that the destinations locates in, the acceptable one-way travel distance, and trip distance in the year of 2008. It is designed to discuss the distribution of trip destinations and the characteristics of the self-driving distance in larger spatial scale.

In the sample, on average, 63% of respondents have travelled out of the province 1.39 times last year. The destinations are concentrated around the province and 71% are distributed over Sichuan, Guizhou and Guangxi provinces.

According to statistics, 37% of respondents think that the acceptable distance of one-way travel is between300km and 500km; 23% of them think it is between 100km and 300km; only 4% think it is within 100km. In a word, the destinations are generally no more than 500km from Kunming. When it is above 500km, there is a sharp drop in the number of travel. And in this range, the number of tourists grows with the increase of the distance. Based on this, urban self-driving tourism's critical travel distance ranges from 300km to 500km. When beyond the range, tourists' travel distance appears the attenuation phenomenon (see Figure 2). In a study, Wu (2001) defined 500km as a standard distance of urban recreational tourism. It can be concluded that both travel purpose and travel distance reflect the features of typical recreation activities.

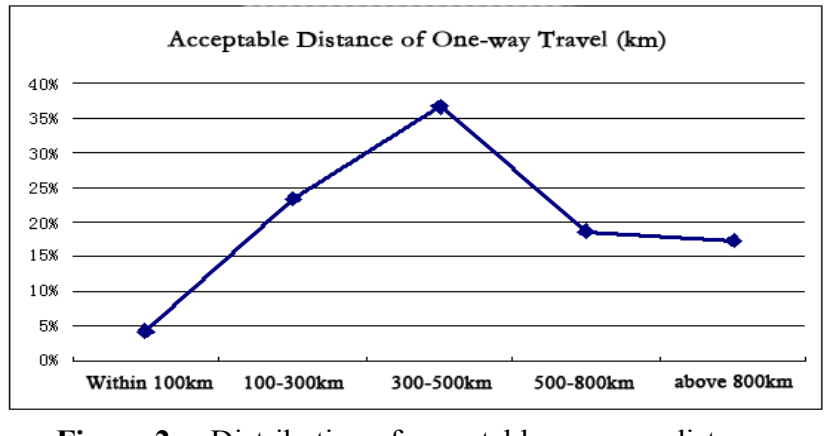


Figure 2. Distribution of acceptable one-way distance

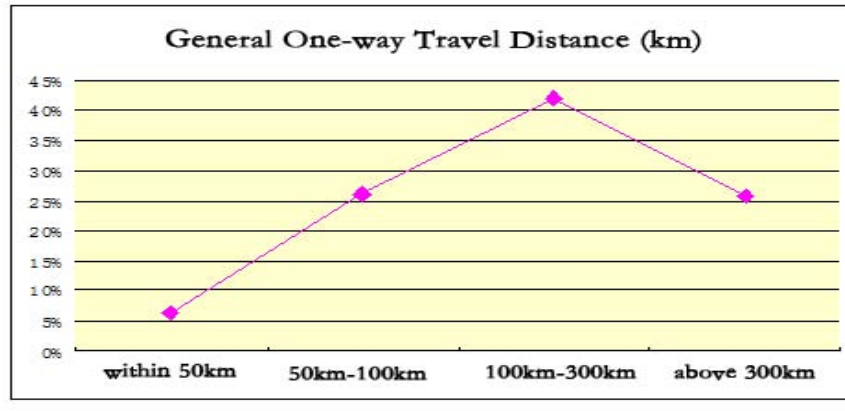


Figure 3. Distribution of general one-way distance

Statistics on travel distance has the same pattern (see Figure 3). 74% tend to choose the destinations within 300km away from the city and only 26% choose destination beyond 300km. Most respondents, nearly 42% prefer the destinations between 100km to 300km, and only 6% choose places within 50km. In the same way, if the distance is below 300km, the number of samples grows with the increase of the distance. It shows the "anti-attenuation" phenomenon. What is mentioned above seems to be different from the findings in Zhang's (2006) study. She predicted that 80% tourists thought the distance of one-way travel was below 500km and the tourists' choices of real travel distance appeared the "anti-attenuation" phenomenon. Moreover, gender is also a factor affecting travel distance. The maximum psychological distance is 498 km for men and 359 km for women. The difference is nearly 150 km.

If the maximum distance that tourists can tolerate is regarded as the maximum psychological distance, the normal one-way driving distance can be regarded as the real tourism distance. Both of the two kinds of distances appear to have the "two-way attenuation" phenomenon, but there is a remarkable difference in the average values, which is 465km for maximum psychological distance, and only 186km for real tourism distance. Wu's study (2001) showed that "91% of residents move within 15km around the city and 60% of residents move within 50km around the city. It can be concluded that due to the increase of mobility, the travel distance of self-driving tourists is longer than the residents. But because of the constraints in destination, traffic, and cash cost, there is a remarkable difference between real tourism distance and maximum psychological distance.

Analysis of the reasons for self-driving tour. For the purpose of analyzing the reasons for self-driving tourism, the questionnaire is designed with seven options, and respondents are asked to choose three of them according to the priority. The study shows that the first reason for self-driving tourism is freedom and mobility, with weight of 547 (see Table. 1). The second reason is personal preference. The third reason is traveling with other people and carrying more stuff. Inconvenient public transportation turns out to be the last reason with weight of 69. The research also shows that 49% of respondents think there are various convenient travel modes to choose from for the

travel, and only 5% think the travel modes are few and inconvenient. He (2005)'s study showed that the need for work, convenience, long distance and trip time were main reasons for use of cars in the city. But these factors are quite trivial in self-driving tourism decision. The three main factors (freedom, habit, traveling with other people) were mainly influenced by individual values and life style, which are different from traditional drive decision. So it can be concluded that as a way of recreation, with the obvious change of idea and surroundings choice, self-driving tourism has changed from a travel mode to a consumption activity and life style, leading to the change of its relationship with public transportation.

Characteristics of weekends tourism. The time choice of self-driving tourists equally lie in all kinds of holidays except for National Day Golden Week, as shown in Figure 4. Weekends constitute most of leisure time of the year, and thus we focus on analyzing the characteristics of tourism on weekends.

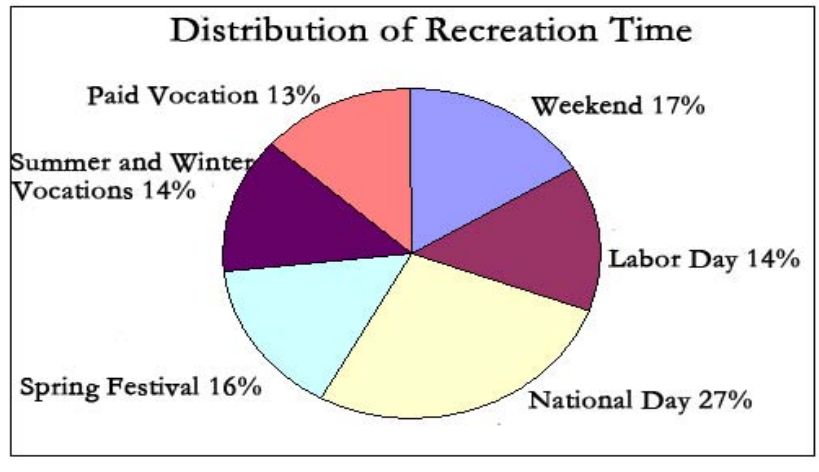


Figure 4. Distribution of general recreation.time

Table 1. Reasons for self-driving tourism

Reasons	Order			Number of sample	Expected value	Order of expected value
	No. 1	No. 2	No. 3			
Freedom and mobility	144	45	25	214	547	1
Travel distance is not too long	25	29	25	79	158	4
Inconvenient public transportation	3	20	20	43	69	6
Carry more stuff	13	40	41	94	160	3
There are the elder or younger	10	34	23	67	121	5
Personal preference	32	41	35	108	213	2
Traveling with other people	23	26	39	88	160	3

Expected value=Number of sample*weight (the value of No.1 is 3, the value of No.2 is 2 and the value of No.3 is 2).

The data shows that average monthly income of people who choose to travel on weekends is RMB￥4208 and 7% higher than the average. Seventy percent of the weekend tourists are self-driving tourism, and is 12% higher than the average of the sample including travelers on week days and weekends. Travel frequency is 12.5 times per year, and is twice of the average level. Most of the travels are recreational, and sightseeing travel only makes up of less than 20%. The proportion of travel for visiting relatives and friends on weekends is higher than the average of the whole sample (see Figure 5).

Concerns on traffic, safety and information become main discouraging factors of self-driving tourism. The "difficulties encountered in the trip" listed by respondents shows that traffic factors play an important role, and the first three difficulties are bad traffic condition in the trip, parking problems in the scenic spots in rush hours, and being unfamiliar with the travel route, as shown in Table 2

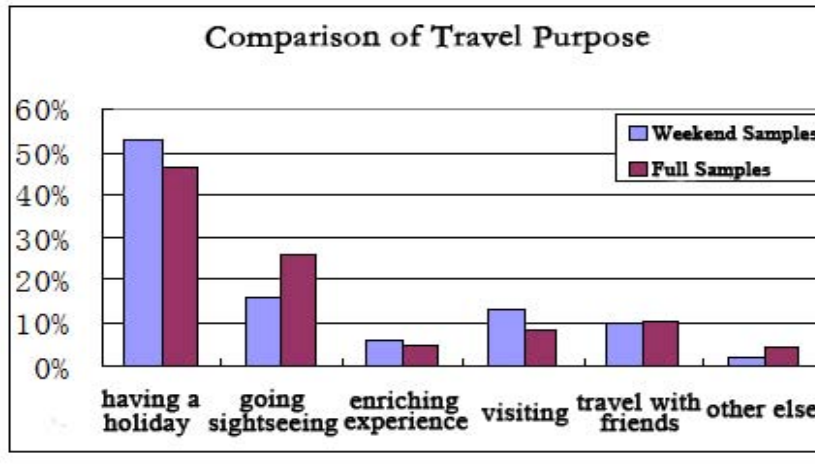


Figure 5. Analysis on comparison of travel purpose

Table 2. Difficulties usually encountered in the trip

Difficulties	Order			Number of sample	Expected value	Order of expected value
	No.1	No.2	No.3			
Bad traffic conditions along the road	105	31	23	159	400	1
Imperfect road facilities along the road	24	32	32	88	168	4
Parking problem in the scenic	47	53	39	139	286	2
Worry about driving at night	11	16	14	41	79	6
Not familiar with travel route	46	54	27	127	273	3
The car break down during travel	12	26	23	61	111	5
Lack of driving experience	5	8	7	20	38	7

The question “most important problems in the trip” listed shows that the first three are traffic, safety and information, followed by cost, distance, congestion and parking supply.

The study on self-driving tourism’s consumption structure shows that the first three are residence, transportation, and food, which add up to 76% of all the costs. The proportion of tickets, shopping and entertainment adds up to 23%. This also suggests that transportation is becoming an important discouragement to self-driving tourism in terms of travel conditions and the costs.

Conclusions

With the increase of car ownership, income and spare time, the self-driving tourism is becoming popular, and more importance is attached to the supply-demand problem of the consequent traffic device. As an important function of family cars, self-driving tourism should be studied in depth. In this paper, based on the empirical study self-driving tourism in the city of Kunming, we conclude:

- The purpose of self-driving tourism is mainly recreation, not the traditional tour, and going out on weekends is getting more common. Great changes have taken place in tourism conception and environment choice, so the characteristic of cars has transformed from a travel mode to consumption activity, which should be paid attention in traffic demand management.
- The distance distribution of self-driving tourism appears to have the "two-way attenuation" phenomenon. The critical distance is 300-500km around the city. The decrease appears when distance exceeds the threshold. Influenced by real tourism conditions such as destination, traffic, and cost, the real tourism distance is different from maximum psychological distance. Gender is also an influencing factor. The maximum psychological distance is 498 km for men and 359 km for women.
- It is individual values and life style rather than normal tourism factors that mainly influence tourism decision. In the light of this statement, when cars are used for tourism, its relationship with public transportation is changed, which needed considering when travel mode choice is studied.
- Traffic, food, and residence make up of 76% of the consumption structure. Traffic, safety and information are mainly restrictive factors of self-driving tourism.
- Most drive tourists are young or middle-aged, with income above the intermediate level of all sorts of occupation. Generally speaking, self-driving tour is becoming more popular and tourists younger.

Acknowledgement

This study is funded by National Key Basic Research and Development Plan (973 Plan) (Project ID: 2006CB705500), National High Tech Research and Devleopment Plan (863 Plan) (Project ID: 2008AA11Z201), New Century Elitist Support Plan of the

Department of Education (Project ID: NCET-07-0176), and Yunnan Provincial Department of Education Science and Technology Foundation (Project ID: 08Z0018）.

References

Alan Shailes, Martyn L. (2001). "Tourists' travel behaviour in response to congestion: the case of car trips to Cornwall, United Kingdom." *Journal of Transport Geography,* 9, 49-60.

Baohong He, Jun Chen, Wei Wang. (2005) "Survey and Analysis of the Trip Characteristic of Private Cars—with Nanjing as Example." *Urban Planning Forum,* 4(158), 83-87.

Bihu Wu. (2001)`*Plan of Regional Tourism*, China Travel & Tourism Press, Beijing.

Hongzhi Guan, Jie Shao, Yajie Li, Hongling Bai. (2005). "A Fundamental Approach to Analyze the Driving Demand." *Journal of Changsha University Beijing University of Technology,* 31(2), 151-154.

Hui Zhou. (2005). "The Characteristics of Market Demand for City Self-Driving Tour with the Case of Changsha City." *Journal of Changsha University,* 19(6), 25-27.

Janet E.Dickinson, Derek Ronnins,(2008). "Rrpresentation of tourism transport problems in a rural destination" *Journal of Tourism management,* 29, 1110-1121.

Kunming statistical bureau. (2009). "Statistical Communiqué of the People's Republic of Kunming on the 2008 National Economic and Social Development." http://www.km.gov.cn/structure/xwpdlm/tjxxxx_100459_1.htm(April 2, 2009)

Qiankang Chen. (2004). "A Study on the Market Development of Self-Driving Tours." *Tourism Tribune,* 19(3), 66-71.

Shuhua Feng. (2008). "Study on Behavioral Characteristics of Self-driving Tourists and their Spatial Effect." *Tourism Tribune,* 23(9), 34-38.

The National Holiday Office. (2008). "Analysis Report of tourists' travel preference during Golden Week". http://www.ljta.gov.cn/FNADMIN/ReadNews.asp?NewsID = %203290(September 24, 2008).

The National Bureau of Statistics. (2009). http://www.stats.gov.cn/tjfx/ztfx/qzxzgcl60zn/t20090923_402589513.htm(September 23, 2009)

Xiangkun Zhai(2003). "Strategic thinking to develop self-driving tour in China." *Journal of Beijing International Studies University,* 5(117), 54-60.

Xiaoyan Zhang, Shan-qin Zhang, Xung MA. (2006). "A Study on the Behavior of Self-Driving Tourists in China." *Tourism Tribune,* 9(21), 31-35.

Yechezkel Israeli, Yoel Mansfeld. (2004). "Tourist flow management to and within historical towns: the case of the old city of JERUSALEM", WTCR.

Zhantu Ye (2006). "The Analysis of Self-driving Tourists' Travel behaviour Characteristic —A case study of Guangzhou." Zhong Shan University.

Route Choice Analysis of Bus-use Commuters in Yangon, Myanmar

Hironori Kato[1], Akihiro Inagi[2] and Phyo Thet Thet Htun[3]

[1] Associate Professor, [2] Master Student, Department of Civil Engineering, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan, Email: kato@civil.t.u-tokyo.ac.jp
[3] Doctoral Student, Department of Civil Engineering, Yokohama National University

Abstract

This paper surveys the travel behavior of bus-use commuters in Yangon, Myanmar and analyzes the bus route-choice models with the collected data. The revealed preference survey combined with the stated-preference survey was implemented in the western and southern parts of Yangon City. 1,240 choice data of 169 individuals were collected including 169 RP choice data and 1,080 SP choice data. Then, the bus route-choice is analyzed with a binary Logit model. After checking the reliability of SP data by estimating the SP model, the RP/SP combined model is estimated. Additionally, the RP/SP model incorporating the monthly income is also estimated. They show that the bus-use commuters prefer the high-quality buses to the truck bus. They also show that the estimated value of travel time is about 10.1 Kyats/minute and it increase as the monthly income increases.
Key Words: Urban bus, route choice, SP/RP combined model, Yangon, Myanmar

Introduction

Yangon is the most important commercial center of Myanmar with about 4.1 million people in the area of 598.76 km^2 (United Nations, 2007). There are mainly four types of transportation modes available in Yangon: private car, rail, taxi and bus. The vast majority of urban transportation relies on the bus transportation in Yangon. According to Zhang et al. (2008), the modal share of bus transportation in Yangon reaches 84%. This is because of the difficulties of the car ownership, the low quality of rail service, and the regulation of motorbike and para-transit service in the central business district.

However, as Kato et al. (2010) points out, the quality of urban bus service in Yangon is unacceptably bad. The most of buses are second-handed and ill-maintained. They often have the engine troubles even during the operation. They are inconvenient and uncomfortable due to the ill-operated service. No time table and no route map are shown to the public. Not only the visitors but also the local people have the difficulties to use the local bus system. Although many bus routes are operated in a wide network of the City, they are sometimes over-crowed due to the lack of vehicle's capacity. The traffic congestion caused by the lack of road capacity leads to the chronic delay and unpunctuality of bus service particularly during the peak hours. The bus users have to wait for a long time at the bus stops to wait a next bus. The bus stops or terminals are also not well-designed. Some bus passengers must walk several minutes to transfer the buses at the bus stops/terminals. As the urban area grows, the travel distance has been also longer. Many local people have to travel long in the congested vehicles during the peak hours. To improve the quality of urban bus service is critically important in Yangon.

This paper surveys the behavior of local bus users in Yangon. The bus-use commuters are surveyed because the most serious problems occur mainly during the morning peak hours. Then, their bus route-choice is analyzed with the collected data. The bus route-choice model incorporates the main factors of bus service including the travel time, waiting and transfer at the stops/germinal and travel cost. The accessibility of bus users can be evaluated with this model.

The paper is organized as follows. Next section summarizes the characteristics of bus transportation service in Yangon. Then, section 3 presents the local survey on bus-use commuter's behavior. The bus route-choice model based on RP/SP combined model and its estimation results are shown in section 4. Finally, the achievements are discussed.

Bus Transportation Service in Yangon

There are the two types of local bus services in Yangon: the bus service given by bus companies and the bus service controlled by the bus control committees (BCCs). The bus companies are the large-scale private bus operators who own the bus vehicles, hire the drivers/conductors and operate the urban bus service in Yangon. The two bus companies are now supplying the bus service in Yangon: the Golden City Link Co. (GCL) and the Union of Myanmar Economic Holdings Limited (UMEHL). The BCCs are the non-profit organizations which control the small-scale individual bus owners. The individual bus owners own the bus vehicles and lease the bus vehicles to the drivers and conductors. The individual bus owners must belong to one of the BCCs. There are a number of individual bus owners in Yangon.

There are the two types of bus fare systems in Yangon: a distance-based fare system and a constant fare system. First, the distance-based fare system is introduced into the bus routes controlled by the BCCs except for some special bus routes. A unique

distance-based fare system is shared among all the BCCs: the minimum fare is 50 Kyats; the fare rises by 50 Kyats for every 5-mile after 5-mile travel; and the maximum fare is 250 Kyats. Second, the GCL has the constant fare system. Its fare is 300 Kyats per ride. Third, the UMEHL has the two types of fare systems. The one is the constant fare system which is applied to the routes where the air-conditioned city bus service is provided. Its fare is 200 Kyats per ride. The other is the distance-based fare system which is the same as that of BCCs. Generally, the bus fare is collected by the conductors serving in the bus vehicle or at the entrance of the vehicle. Note that the GCL exceptionally introduces the ticket system, in which the bus conductor sells the tickets for collecting the fare.

No time table of bus operation is available in Yangon. Although the rough information about the service frequency is widely known by the local people, it is not presented explicitly anywhere. As there is no information about the schedule of the first and last bus service of the day, they are not guaranteed at all. As the bus number is written only in Myanmar language, it is almost impossible to identify the destination for the people who have no experience to use the bus. As no information about the bus stop is presented, the visitors have no way to know the name of bus stop. Additionally, no information about the fare table is presented at the bus stops. The bus operation is also not well-controlled in Yangon. In practice, only the cycle time of bus operation is monitored by the BCC at the control gate.

According to Kato et al. (2010), the local buses can be categorized into the following six types from the viewpoints of their function, design and capacity: city bus (air-conditioned); city bus (non-air-conditioned); mini bus; Dyna/Canter; Hilux; and other. Dyna/Canter and Hilux are the trucks which are redesigned into the passenger-use by adding the roofs and seats on the body. They are widely called the “truck bus” by local people in Yangon. The majority of buses in operation are the truck buses. The vehicles controlled by the BCCs are mainly the small-scale bus vehicles, including the Mini bus, Dyna/Canter, and Hilux.

The average length of a bus route is about 32 km. Note that the distance from the northern end to the southern end or the distance from the eastern end to the western end of Yangon City is about 30 km each. The most of bus routes run directly between the suburban areas and the central business district. In other words, the bus network in Yangon is characterized as a “point-to-point” network. In many bus routes, the service frequency during the peak hours in the morning or in the evening is higher than the service frequency during the off-peak hours. For example, according to our interviews, the MaThaTa BCC operates the bus service every 5 minutes from 5:00 a.m. to 7:00 a.m.; every 3 minutes from 7:00 a.m. to 9:00 a.m.; every 7 minutes from 10:00 a.m. to 3:00 p.m.; and every 3-5 minutes from 5:00 p.m. to 7:00 p.m.

Local Survey on Bus-use Commuters in Yangon

The study team including the authors made the local survey on the urban bus-use traveler's behavior. The survey collects the data of sample-based bus-route-choice in Yangon City. The area where the survey was implemented is mainly the southern and western parts of Yangon City including the central business district (Lanmadaw township, Latha township, Pabedan township, Kyauktada township and Pazun Daung township), Ahlone township, Bahan township, Kyee Myin Daing township, Kamaryut township, Hlaing township, Hlaing Tharyar township, Insein township, Mayangone township, Mingalardon township (south area), North Okkalapa township, Sanchaung township, Tarmway township and Yankin township. These areas cover the two main roads and a number of bus routes connecting the sub-urban areas with the central business district.

The questionnaire sheets in the survey request the respondents to answer: (1) the daily commuting behavior in the morning including the origin and destination; departure time from home and arrival time to destination; travel mode used and the reason to choose it; routes used including the bus number and the reason to choose it; travel time and travel cost; average waiting time at bus stop; available alternative routes; (2) individual attributes including personal information such as age, gender, job, household income and the number of family members; the addresses of home and workplace; and (3) travel preference data including the priorities among the different service factors; customer satisfaction regarding the current bus system including fare, travel time, frequency and other service; and the stated preference (SP) choice among the hypothetical routes. As for the stated preference survey, the respondents are requested to choose the best route among the given route choice set from their home to their workplaces. The 16 types of hypothetical choice problems are prepared in advance. A surveyor randomly selects the two or three problems and asks them to a respondent. The travel time, travel cost, transfer time, waiting time and access/egress time and the bus type of each route option are shown to the respondent in the stated preference question.

The bus travel survey takes 10 to 15 minutes per respondent. The bus stops where the bus travel surveys were made are shown below: Bus stop of line No.48 and Pauktaw; Insein Garden; Bying Naung Junction and Golden bus line No.2 bus stop near that; and Thamine Junction. They are selected mainly because there are two or more bus routes from them to the central business district. The interviews are made during peak hours in the evening. This is first because we wanted to collect the bus-use commuters. The reason for collecting the data of commuters is that the major traffic problems regarding the urban bus transport are caused by the commuter's traffic during the peak hours of morning or evening on weekdays. The second reason for making the interviews in the evening is because we consider that the commuters do not accept the interviews in the morning due to their time constraints. The surveyors selected randomly the bus passengers who are getting off the bus at the bus stop/terminal and interviewed them with the paper-based questionnaire sheets.

The study team hired 20 local surveyors, mainly the master-course students of the Yangon Technological University. The surveys were implemented on November 9 to 21, 2008. They were all the non-rainy weekdays. 593 individuals were interviewed and 2,379 route-choice results were collected. 1,389 route-choice data including the RP data of 169 of individuals and the 1,220 SP data of 465 individuals are valid among them. The share of RP-based data is low among all the route-choice data because many of respondents do not have the alternative bus route from their home to their workplaces in their daily commuting.

The characteristics of survey respondents are summarized in Table 1. First, the male respondents are more than the female respondents. This reflects the gender share of local commuters in Yangon. The respondents of 18 to 59 years old are dominant. This is because only the commuters are interviewed. The share of age groups from 18 to 59 may reflect the age distribution of population in Yangon. About half of the respondents are the employees at the private companies while one third of the respondents are the public servants. The share of self-employees is low probably because the most of self-employees do not commute or because they do not use the buses for commuting. The respondents of 30,000 – 60,000 Kyats are the most among the income groups. The higher the income is, the lower the bus-use commuters are. The average monthly income of respondents is 85,968 Kyats. As Zhang et al. (2008) shows, the increase in income will bring about a potentially large increase in car usage and consequently reduction in transit systems. About 40 % of respondents arrive at their workplaces during 8:00-8:59. This is mainly because the working time starts at 9:00 in many offices in Yangon. However, interestingly, the respondents arriving at destinations during 7:00-7:59 or 9:00-9:59 are also observed. This may be because they shift the departure time to avoid the traffic congestion during the peak hours in the morning.

Estimation of Bus Route-choice Model

Bus Route-choice Model. We estimate the bus route choice model with the data collected in the bus travel survey. A multinomial Logit (MNL) model is used to analyze the route choice of bus travelers. MNL model is one of the discrete choice models (Ben-Akiva and Lerman, 1985). In the mode, it is assumed that a consumer chooses an option discretely by maximizing his/her utility with the random factor. The conditional indirect utility function of a route from the one origin to the other destination is formulated as follows:

$$U^n_{od,r} = V^n_{od,r} + \varepsilon^n_{od,r} \tag{1}$$

where $V^n_{od,r}$ is the universal component of the indirect utility of an individual n when the rth route is chosen from zone o to d, and $\varepsilon^n_{od,r}$ is the error component of the utility function following the independently and identically distributed (iid) Gumbel.

Table 1. Characteristics of survey respondents (N=169).

Variables	Categories	Number of observations	Percentage (%)
Gender	Male	109	64.5
	Female	60	35.5
Age	Under 17	1	0.6
	18 - 29	61	36.1
	30-39	49	29.0
	40-49	35	20.7
	50-59	22	13.0
	Over 60	1	0.6
Occupation	Private company	84	49.7
	Public servant	57	33.7
	Self-employment	28	16.6
Monthly income (Kyats)	Under 30,000	23	13.6
	30,000-60,000	65	38.5
	60,001-90,000	41	24.3
	90,001-120,000	15	8.9
	120,001-150,000	10	5.9
	Over 150,001	15	8.9
Arrival time in the morning	-6:00	3	1.8
	6:00-6:59	11	6.5
	7:00-7:59	42	24.9
	8:00-8:59	69	40.8
	9:00-9:59	35	20.7
	10:00-10:59	8	4.7
	11:00-	1	0.6

Then, the probability of choosing the rth route is chosen from zone o to d for individual n is derived as:

$$p_{od,r}^{n} = \frac{\exp\left(\lambda V_{od,r}^{n}\right)}{\sum_{r' \in R_{od}^{n}} \exp\left(\lambda V_{od,r'}^{n}\right)} \tag{2}$$

where $p_{od,r}^{n}$ is the probability of choosing the rth route is chosen from zone o to d for individual n and λ is a scale parameter corresponding to the Gumbel distribution with $\lambda^2 = \pi^2 / 6\sigma^2$ (σ^2 is the variance of the Gumbel distribution) and zero mean.

The conditional indirect utility function of each route is assumed to be a linear function of explanatory variables as:

$$V_{od,r}^{n} = \sum_{m} \kappa_m X_{m,od,r}^{n} \tag{3}$$

where κ_m: unknown parameters and $X_{m,od,r}^{n}$: the mth explanatory variable of the rth route from zone o to d. Since the unknown parameters λ and κ_m cannot be estimated independently, we estimate the unknown parameters ρ_m by using the following utility function:

$$\lambda V_{od,r}^{n} = \sum_{m} \rho_m X_{m,od,r}^{n} \tag{4}$$

The unknown parameters are estimated by maximizing the likelihood function with the individual data collected in the bus travel survey. The likelihood function is defined as:

$$L(\boldsymbol{\rho}) = \prod_{od} \prod_{r} \prod_{n} \left(p_{od,r}^{n}\right)^{\delta_{od,r}^{n}} \tag{5}$$

where $\delta_{od,r}^{n}$ is 1 if individual n chose the rth route is chosen from zone o to d; and 0 if not.

RP/SP Combined Model. The data collected in our survey includes RP data as well as SP data. SP data has more biases than RP data. For example, the respondents cannot imagine the hypothetical options appropriately. They may feel confused when answering questions due to bad explanation by the surveyors or improper ordering of questions. On the other hand, the SP data is useful to forecast the choice of a new option or to analyze the sensitivity of the level-of-service which is important from the transportation policy but which cannot be observed in reality. (Hensher, 1994; Louviere et al., 2000) To model the individual's choice behavior based on SP data, an RP/SP combined model is applied here (Ben-Akiva and Morikawa, 1990). This model is especially used to correct SP reported biases by introducing the RP information.

Let the conditional indirect utility functions of RP data and SP data be $U_{od,r}^{n,RP}$ and $U_{od,r}^{n,SP}$ respectively. By following equation (1), they are defined as

$$U_{od,r}^{n,RP} = V_{od,r}^{n,RP} + \varepsilon_{od,r}^{n,RP} \tag{6}$$

$$U_{od,r}^{n,SP} = V_{od,r}^{n,SP} + \varepsilon_{od,r}^{n,SP} \tag{7}$$

The RP/SP combines model assumes that SP utility function $U_{od,r}^{n,SP}$ has a different variance $\sigma_{od,r}^{n,SP}$ of the error term $\varepsilon_{od,r}^{n,SP}$ from the variance $\sigma_{od,r}^{n,RP}$ in RP utility function as follows:

$$\sigma_{od,r}^{n,SP} = \mu^2 \sigma_{od,r}^{n,RP} \tag{8}$$

where μ is an unknown scale parameter. As it is expected that there exist more biases in SP data than in RP data, the scale parameter μ is usually less than 1. We again assume that the error component of the utility function follows the independently and identically distributed (iid) Gumbel. Then, the probabilities of choosing an option in RP data and SP data are derived respectively as follows:

$$p_{od,r}^{n,RP} = \frac{\exp\left(\lambda V_{od,r}^{n,RP}\right)}{\sum_{r' \in R_{od}^{n,RP}} \exp\left(\lambda V_{od,r'}^{n,RP}\right)} \tag{9}$$

Table 2. Estimation results of SP models.

Variable	Unit	Model SP1		Model SP2	
		Coefficient	t-statistic	Coefficient	t-statistic
In-vehicle time	Minutes	-0.0345	-7.62	-0.035	-7.70
Waiting time	Minutes	-0.0366	-2.37	-0.0381	-2.40
Transfer time	Minutes	-0.0102	-0.75		
Dummy for transfer				-0.1253	-1.02
Access and egress time	Minutes	-0.0578	-3.44	-0.0584	-3.50
Travel cost	Kyats	-0.0041	-3.70	-0.0043	-3.82
Dummy of bus type		0.432	4.48	0.4315	4.47
McFaddens's Rho-squared			0.0447		0.0449
Number of observations			1080		1080

$$p_{od,r}^{n,SP} = \frac{\exp\left(\mu\lambda V_{od,r}^{n,SP}\right)}{\sum_{r'\in R_{od}^{n}} \exp\left(\mu\lambda V_{od,r'}^{n,SP}\right)} \tag{10}$$

The unknown parameters μ and $\boldsymbol{\rho}$ can be estimated by maximizing the following likelihood function:

$$L^{RP/SP}(\boldsymbol{\rho},\mu) = \left(\prod_{od}\prod_{r}\prod_{n}\left(p_{od,r}^{n,RP}\right)^{\delta_{od,r}^{n,RP}}\right)\left(\prod_{od}\prod_{r}\prod_{n}\left(p_{od,r}^{n,SP}\right)^{\delta_{od,r}^{n,SP}}\right) \tag{11}$$

Estimation Results. First, we estimate the MNL model only with the SP data to test the reliability of SP data. This is because the SP survey has more biases than RP survey although the respondents show the high interests in our survey. The individual's choice set is assumed to include the two routes. This is because the respondents answered the choice of two alternatives in both RP and SP surveys. The explanatory variables we use are in-vehicle time, waiting time, transfer time, dummy for transfer, access and egress time, total travel time, travel cost, and dummy of bus type. The dummy for transfer is defined as 1 if the route includes one or more transfers and 0 if not. The dummy of bus type is defined as 1 if the bus operated in the route is "city bus" and 0 if not. The estimated results of the four models are shown in Table 2.

The signs of all parameters are fine. Most of the parameters for level-of-service variables are statistically significant. However, the transfer time or the dummy for transfer is not significant. Probably the respondents found the difficulties to find the difference of transfer time among the given choice options. However, in general, it is expected that the respondents answer the reasonable choice after understanding the survey.

Table 3. Estimation results of RP/SP combined models.

Variable	Unit	Model SP/RP1		Model SP/RP2		Model SP/RP3	
		Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic
In-vehicle riding time	Minutes	-0.0677	-3.24	-0.055	-3.17		
Waiting time	Minutes	-0.0539	-1.83	-0.0518	-2.06		
Transfer time	Minutes	0.0271	0.77				
Dummy for transfer				-0.0631	-0.25		
Access and egress time	Minutes	-0.0957	-2.70	-0.0875	-2.64		
Total travel time	Minutes					-0.0547	-3.36
Travel cost	Kyats	-0.0072	-2.84	-0.006	-2.94	-0.0054	-3.03
Travel cost/income							
Dummay of bus type		0.8655	3.04	0.6928	2.78	0.5899	2.89
Scale parameter		0.4133	2.44	0.5746	2.57	0.7675	3.14
Value of in-vehicle riding time	Kyats/minute		9.40		9.17		
Value of total travel time	Kyats/minute						10.13
McFaddens's Rho-squared			0.046		0.046		0.059
Number of observation			1240		1240		1240

Table 4. Estimation result of RP/SP combined model incorporating monthly income.

Variable	Unit	Model RP/SP4	
		Coefficient	t-statistic
Total travel time	Minutes	-0.0625	-3.73
Travel cost	Kyats		
Travel cost/income		-186.3	-2.40
Dummay of bus type		0.6508	3.08
Scale parameter		0.6155	3.41
McFaddens's Rho-squared			0.055
Number of observation			1240

Then, we estimate the RP/SP combined model. As for the RP data, the explanatory variables of unused route are prepared based on the observed data. The estimation results are shown in Table 3. In the same way as the results of SP model estimation, the transfer time or the dummy for transfer is not significant in Model RP/SP1 and Model RP/SP2. Additionally, the sign of transfer time in Model RP/SP1 is positive. The Model RP/SP 3, in which the total travel time is used, has the best fitness among the various models we tried to estimate. The estimated scale parameters are also significant. They are less than unity. This means that the variance of error term in the utility function of RP data is significantly smaller from that of SP data as we expected. The estimated parameter of dummy for bus type is positive. This means hat the bus-use commuters prefer the city-bus to the other types of bus. This reflects that the local people want to use the comfortable vehicle rather than the truck bus.

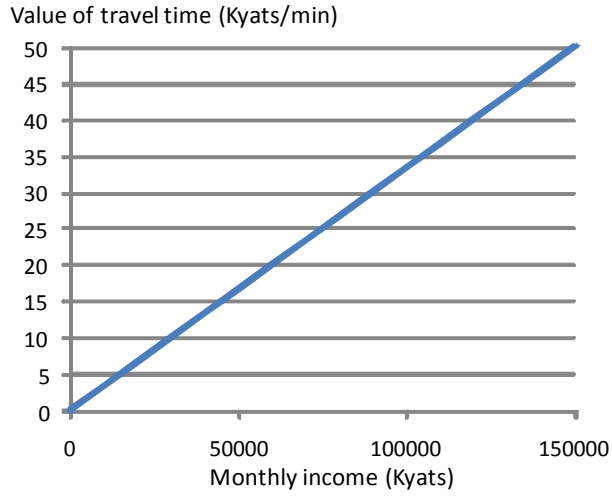


Figure 2. Monthly income vs. estimated value of travel time.

The estimated value of travel time is 10.13 Kyats/minute in Model RP/SP3. Zhang et al. (2008), using the SP-based modal-choice data in Yangon shows that the estimated value of travel time vary from 3 to 10 Kyats/minute which depend on the individual's income level. Then, we also estimate the RP/SP model incorporating the income. The travel cost/income is newly introduced instead of travel cost. This reflects the hypothesis that the route is chosen on the basis of not the travel cost directly but the share of travel cost among the total income. Table 4 shows the estimated parameters. Although the fitness of the model is worse than Model RP/SP3, the estimated parameters satisfy the expected conditions. All of the parameters are also significant. Then Figure 2 shows the monthly income versus estimated value of travel time based on the Model RP/SP4. This indicates that the value of travel time varies from 0 to 50 in the range of 0 to 150,000 Kyats of monthly income. As the monthly income increases, the value of travel time is also higher. The estimated value of travel time may be higher than the estimated values shown by the previous study (Zhang et al., 2008). This is probably because the respondents in our survey are bus-use commuters only. The commuters mean the workers who earn the salary from some business. The workers have higher value of travel time than non-workers because the value of travel time is higher theoretically as the wage rate is higher. The previous study includes the students in addition to the workers and they may reduce the estimated value of travel time. Additionally, it is expected that the bus users earn higher income than the rail users. This is because the rail service in Yangon is so poor that the lower-income people tend to use it. The previous study includes the rail users whereas our study does not.

Conclusions

To our best knowledge, our study is the first attempt to analyze the route-choice behavior of bus users in Yangon. The estimated models are expected to contribute to bus operators, bus regulators and transportation planners. First, the bus operators can improve the business plans based on the detailed information of bus-use demand.

Second, the bus regulators can also take the preference of bus users into considerations when they permit the bus route or bus vehicle to the bus operators. Third, the impacts of transportation policies on bus transportation demand can be forecasted. For example, the transportation policy for reducing the traffic congestion increases the bus service speed and this influences the bus-route choice of local people. Finally, the transportation policy regarding the bus transportation can be evaluated with the model. For example, the improvement of accessibility caused by transportation policy can be calculated as the monetary benefit by using the estimated model. The log-sum values derived from the multinomial logit model represents the traveler's welfare. This can be used in the benefit calculation during the process of cost-benefit analysis.

References

Ben-Akiva, M., and S. Lerman (1985). "Discrete Choice Analysis: Theory and Application to Travel Demand." MIT Press.

Ben-Akiva M, Morikawa T. (1990). "Estimation of mode switching models from revealed preferences and stated intentions." *Transportation Research Part A*, 24(6), 485–495.

Hensher D. A. (1994). "Stated preference analysis of travel choices: the state of practice." Transportation,21,(2), 107–133.

Institute Sultan Iskandar, and Malaysian Industry Government Group for High Technology (2005). "Yangon Strategic Development Plan, Interim Report."

Kato, H. (2009). "Improvement of Urban Bus Transport System in Yangon, Myanmar, Final Report." *ASEAN-Japan Transport Partnership Project*, 1–150,

Japan Transport Cooperation Association, Prepared for Ministry of Land, Infrastructure, Transport and Tourism of Japan.

Kato, H., Saito, N., Inagi, A., and Myint U. A. (2010). "Regulatory framework and operational system of urban bus transportation in Yangon." *Myanmar, Transportation Research Board 2010 Annual Meeting, January 2010.* (accepted)

Louviere J. J., Hensher D. A.,and Swait J. D. (2000) *Stated Choice Methods: Analysis and Application*, Cambridge University Press.

United Nations (2007) World Urbanization Prospects.

Zhang, J., Fujiwara, A., and Thein, S. (2008). "Capturing traveler's stated mode choice preferences under influence of income in Yangon city." Myanmar, *Journal of Transportation Systems Engineering and Information Technology*, 8(4), 49-62.

An Inhabitant Travel Time Distribution Model

Junjuan Liu, Fenyi Dong and Bingjun Li

School of Information and Management Sciences, Henan Agricultural University, Zhengzhou, Henan 450002, China; PH (86)371-63555710; Email: zzjjliu@126.com, dongfenyi@163.com, zzlbjun@163.com

Abstract

In order to study the law of inhabitant travel time distribution, the model for the constraints of utilities and targeted information entropy maximization was proposed. Probability theory of Logit model and maximum information entropy principles were used to build the planning model. Through solution of the planning model, model of inhabitant travel time distribution was set up. The frequency distribution of travel time of work, school, and shopping about Suzhou City has respectively 2,3,4, and 5 polynomials simulated, and suggests using 4 or 5 times the polynomial to simulate travel time distribution. The results of simulations show that travel time distribution of work, school, shopping are accord with the model.
Key Words: inhabitant travel time distribution; utility theory; maximum information entropy theory

Introduction

Travel distribution influences urban transport planning, land use planning, and urban planning. Researchers accumulate a lot of harvests on travel distribution models, such as the growth coefficient model, the gravitation model, and the exponential model (John and Lester 2006; Murchland 1966; Hallefjord and Jornsten 1984). These models, at the previous urban planning and urban transport planning, have played important roles. Recently, the travel distribution model has new development, some scholars set up a simulated electron cloud travel distribution model based on electron probability density function of hydrogen atoms (Dianhai et al. 2007), using the maximum entropy principle set up a travel distribution model (Ronghan and Dianhai 2005), the finite Weibull mixture distribution model for modeling value of time distribution was proposed (Lijun 2009) and some paper about distribution of travel time(Hollander and Liu 2008; Paterson and Rose 2008). These models named the

aggregate model from the point of data, and other corresponding type models named the non-aggregate model (Pengpeng et al. 2006; Li and Ziliaskopoulos 2002). There are advantages and disadvantages each other of two types models. It is the struggling direction that these two type models how to effectively combine and get closer to the actual problem. This paper tries to set up a travel distribution model based on the utility theory and the maximum entropy theory, and the two theories belong to the disaggregate type and aggregate type.

Travel time is an important index of the characteristics of the inhabitants' travel. Study of the character and the law of travel time have guiding significance on the transportation system planning, construction, and management.

Model Theory

Suppose the utility function is $u(x)$, $u(x)$ is a abstract function, x is variable and travel time in this paper, $u(x)$ and x has relation

$$u(x) \propto x \tag{1}$$

and $f(u)$ is image function of $u(x)$, the connection is

$$f(u) \propto u(x) \tag{2}$$

Suppose there are m projects, in the basis on the Logit model of the maximize utility on individual, the probability of chosen project i is

$$p_i = \frac{f(u_i)}{\sum_{i=1}^{m} f(u_i)} \tag{3}$$

And so P_i and $u(x)$ is

$$p_i = \frac{f(u_i)}{\sum_{i=1}^{m} f(u_i)} \propto \frac{u_i(x)}{\sum_{i=1}^{m} u_i(x)} \tag{4}$$

and gains information entropy is

$$H_i(x) = -p_i \ln p_i = -\frac{f(u_i)}{\sum_{i=1}^{m} f(u_i)} \ln \frac{f(u_i)}{\sum_{i=1}^{m} f(u_i)} \propto -\frac{u_i(x)}{\sum_{i=1}^{m} u_i(x)} \ln \frac{u_i(x)}{\sum_{i=1}^{m} u_i(x)} \tag{5}$$

Build the planning model of the maximum information entropy in system,

$$\max \; H = -\sum_{i=1}^{m} p_i \ln p_i \tag{6a}$$

$$\text{s.t.} \quad f(u_i) = p_i \sum_{i=1}^{m} f(u_i) \tag{6b}$$

$$\sum_{i=1}^{m} p_i = 1 \tag{6c}$$

Solve this planning model using the Lagrange multiplier method, and use Lagrange multiplier μ_1 and multiply Eq. 6b, $(\mu_2 - 1)$ multiply Eq. 6c, these two formulas use addition and subtraction from Eq. 6a, receive Lagrange function is

$$L = -\sum_{i=1}^{m} p_i \ln p_i - \mu_1 \left(p_i \sum_{i=1}^{m} f(u_i) - f(u_i) \right) - (\mu_2 - 1) \left(\sum_{i=1}^{m} p_i - 1 \right) \tag{7}$$

Using the Lagrange function to P_i differential, the extremism condition is

$$\frac{\partial L}{\partial P_i} = -1 - \ln p_i - \mu_1 f(u_i) - \mu_2 + 1 = 0 \tag{8}$$

$$p_i = e^{-\mu_2 - \mu_1 f(u_i)} \tag{9}$$

$$\text{order} \quad g(f) = -\mu_2 - \mu_1 f(u) \tag{10}$$

as Eqs.1 and 2,

$$g(f) \propto f(u) \propto u(x) \propto x \tag{11}$$

Eq. 11 shows that transfer relation between these functions and variables, as the character of transfer, $g(f)$ can extract the function on x, order the result is $v(x)$, so $v(x)$ and $g(f)$ have the same expression effect, simply has different variable in the two functions, and so Eq. 9 can be expressed as:

$$p_i = e^{v(x_i)} \tag{12}$$

Customarily, with the polynomial description, for it has better describing ability, when no clear expression of $v(x)$ in Eq.12, so $v(x)$ and Eq. 12 can be expressed as:

$$v(x) = \lambda_0 + \lambda_1 x + \lambda_2 x^2 + \cdots + \lambda_n x^n \tag{13}$$

$$p_i = \exp\left[\lambda_0^i + \lambda_1^i x_i + \lambda_2^i x_i^2 + \cdots \lambda_n^i x_i^n\right] \quad (i = 1,2,\cdots,m; j = 1,2,\cdots,n) \tag{14}$$

x_i represents travel time in the above formulas. Travel time of work, school, and shopping has been discretized m values from x1, x2,…, to xm, and it gradually increases. For example, if less than 10 minutes choose 7 minutes, 10~20 minutes choose 15 minutes and so on. λ_j^i is the unknown parameter, both n and x_i would influence λ_j^i, p_i is probability of x_i. Eq. 14 is a probability model based on utility and the maximum information entropy.

Case of the Travel Time Distribution

It is different for travel time distribution of work, school, or shopping (Jian 2004). Research travel frequency for travel times, validate the model, and calibrate parameters. Suzhou is a city located in the Jiangsu province in China, using investigative data of Suzhou to seek the law of travel time, and sample size of work, school, shopping is, respectively, 2998, 964, 1224. In the case, class mid-value can be chosen as follow: less than 10 minutes choose 7 minutes, and 10~20 minutes, 20~30 minutes, 30~40 minutes, 40~50 minutes, 50~60 minutes, more than 60,minutes is respectively choose 15 minutes, 25 minutes, 35 minutes, 45 minutes, 55 minutes, 70 minutes and so on. Table 1, Table 2, and Table 3 is, respectively, travel time distribution of work, school, and shopping.

Using Eq. 14 calculation the fitting coefficients based on investigation data of Suzhou (fact distribution in Tables 1-3), and calculate the test value of Q, validate reliability of the model. From the curve of Figures 1-3(fact distribution), the polynomial is clearly not fitting the number 1, the following polynomial numbers are 2, 3, 4, and 5. The results can be verified in Tables 1-3 and Figures 1-3.

Seeing from Tables 1-3 and Figures 1-3, the frequency distribution of travel time of work, school, shopping has 2, 3, 4, and 5 times polynomial simulated, found that the higher number of polynomials and the better simulation, especially the numbers 4 and 5. It is suggestion that the 4 or 5 times polynomial simulate the frequency distribution of travel time to work, school, and shopping.

Table 1. Travel time distribution for work (time:min).

travel time	Fact distribution	Theory distribution (n=2)	Theory distribution (n=3)	Theory distribution (n=4)	Theory distribution (n=5)
<10	0.1835	0.3095	0.2129	0.1784	0.1808
10~20	0.3662	0.2291	0.2955	0.4065	0.3907
20~30	0.2845	0.1523	0.2279	0.2377	0.2472
30~40	0.0720	0.0977	0.1113	0.0850	0.0858
40~50	0.0437	0.0604	0.0468	0.0406	0.0388
50~60	0.0310	0.0360	0.0230	0.0314	0.0323
>60	0.0190	0.0155	0.0207	0.0190	0.0190
total	1.0	0.9005	0.9381	0.9986	0.9944
χ^2		0.8684	0.3102	0.0584	0.0536
Q		0.9291	0.9581	0.9712	0.8168
Test		pass	pass	pass	pass

Table 2. Travel time distribution for school (time:min).

travel time	Fact distribution	Theory distribution (n=2)	Theory distribution (n=3)	Theory distribution (n=4)	Theory distribution (n=5)
<10	0.3963	0.6068	0.3884	0.3922	0.3905
10~20	0.3320	0.2625	0.3554	0.3493	0.3540
20~30	0.1888	0.1031	0.1667	0.1663	0.1641
30~40	0.0456	0.0459	0.0537	0.0545	0.0543
40~50	0.0197	0.0232	0.0171	0.0172	0.0175
50~60	0.0073	0.0133	0.0078	0.0076	0.0076
>60	0.0104	0.0073	0.0103	0.0103	0.0103
total	1.0	1.0620	0.9994	0.9975	0.9983
χ^2		0.4038	0.0258	0.0255	0.0252
Q		0.9822	0.9989	0.9873	0.8738
Test		pass	pass	pass	pass

Table 3. Travel time distribution for shopping (min).

travel time	Fact distribution	Theory distribution (n=2)	Theory distribution (n=3)	Theory distribution (n=4)	Theory distribution (n=5)
<10	0.4191	0.5675	0.4747	0.4032	0.4083
10~20	0.3113	0.2393	0.2701	0.3628	0.3490
20~30	0.1642	0.0978	0.1186	0.1233	0.1281
30~40	0.0302	0.0491	0.0523	0.0408	0.0411
40~50	0.0278	0.0303	0.0268	0.0235	0.0225
50~60	0.0237	0.0230	0.0185	0.0247	0.0254
>60	0.0237	0.0223	0.0256	0.0236	0.0236
total	1.0	1.0292	0.9866	1.0018	0.9980
χ^2	0.4191	0.5675	0.4747	0.4032	0.4083
Q	0.3113	0.2393	0.2701	0.3628	0.3490
test		pass	pass	pass	pass

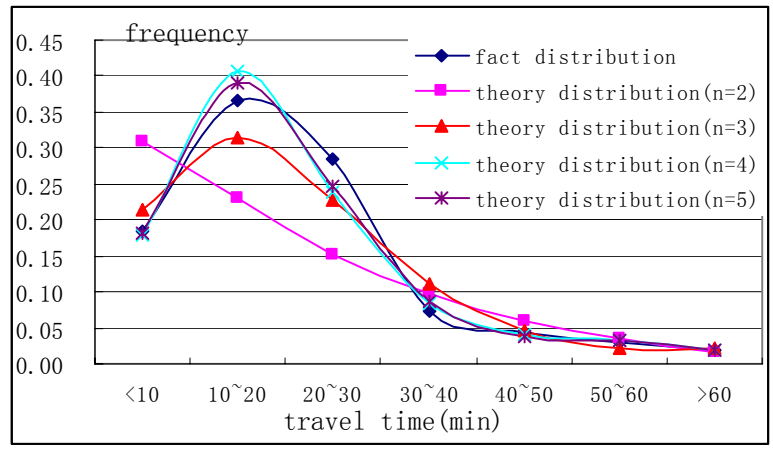


Figure 1. Travel time distribution for work.

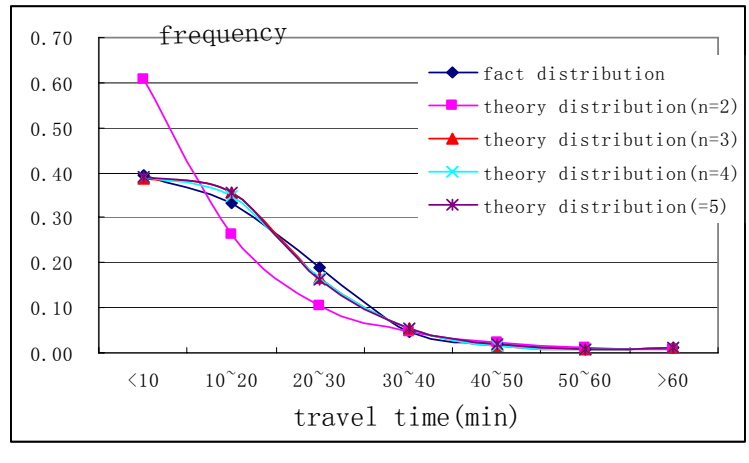


Figure 2. Travel time distribution for school.

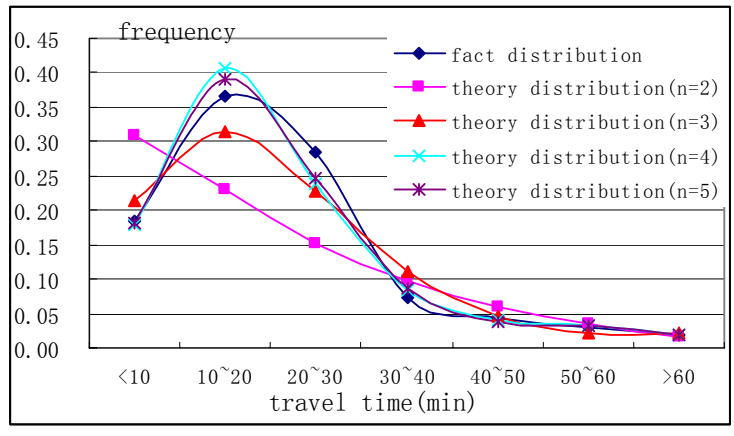


Figure 3. Travel time distribution for shopping.

Conclusions

There is a strength and effectiveness about the utility theory and information entropy theory when using them to describe and solve problems. This paper set up the planning model using an organic combination. The frequency distribution of travel time about work, school, shopping has 2, 3, 4, and 5 times polynomial simulated, and suggest the 4 or 5 times polynomial to simulate travel time distribution. From the modeling principle, this distribution model has great universal applicability, it can not only describe the distribution of travel time, but also describe the distribution of travel distance, and can be applied to sociology, economics, and so on. The effectiveness of the distribution model of travel time has proven with inhabitant travel time of work, school, and shopping of Suzhou City.

References

Dianhai, W., Chao, J., and Ronghan, Y. (2007). "Electron cloud model of residential travel distribution." *Acta Physica Sinica*, 56(7):3642-3648.

Hallefjord, A. A., and Jo¨rnsten, K.(1984). "Gravity models with multiple objectives - theory and applications." *Transportation Research part B,* 20,19-39.

Hollander, Y., and Liu, R. (2008). "Estimation of the distribution of travel times by repeated simulation." *Transportation Research Part C,*16(2): 212-231.

Jian, L., and Wei, W. (2004). "Study of resident trip time consumption characteristic." *Journal of Highway and Transportation Research and Development*, 21(10):102-104.

John, S. M., and Lester, A. H. (2006). "Assessing the utility of private information in transportation planning studies: a case study of trip generation analysis." *Journal of Socio-Economic Planning Sciences*, 40(2):94-118.

Li, Y., and Ziliaskopoulos, T. (2002). "A combined model for time - dependent trip distribution and traffic assignment." *Paper Presented at the 2002 Annual Meeting of the Transportation Research Board (CD - ROM)*,Washington, D.C.

Murchland, J. D.(1966). "Some remarks on the gravity model of traffic distribution and equivalent maximizing formulation." *Report LSE-TNT*-38.

Paterson, D. and Rose. G. (2008). "A recursive, cell processing model for predicting freeway travel times." *Transportation Research Part C,16(4)*: 432-453.

Pengpeng, J., Hua-pu, L., and Lang,Y. (2006). "Disaggregate traffic mode choice model based on combination of revealed and stated preference data." *Tsingua Science and Technology*, 11(3):351-356.

Ronghan, Y., and Dianhai, W. (2005). "Entropy models of inhabitrant trip distributions and parameters calibration." *Journal of Traffic and Transportation Engineering*, 2(4):106-110.

Yu, L. (2009). "Finite weibull mixture distribution model of VOT distribution forecasting." *China Journal of Highway and Transport*, 22(4):96-101.

Traffic Cognitions, Behaviors, and Education of Child Pedestrians: A Case Study in Kunming, China

Yaqin Qin[1], Jian Xiong[2], Fengxiang Guo, Huasen Wan and Xianguang Jia

[1] Associate Professor, Faculty of Transportation Engineering, Kunming University of Science and Technology, Yunnan China 650224, Email:qyq_email@foxmail.com
[2] Professor, Faculty of Transportation Engineering, Kunming University of Science and Technology, Yunnan China 650224, Email: xjebox@163.com

Abstract

There are over 18,500 children below the age of 14 that died from traffic accidents every year in China. The death rate in China is 2.5 times more than that in Europe and 2.6 times compared to the USA. Researchers have shown that half of children's traffic accidents were caused by their own negligence. One of countermeasures is to enhance the children's traffic cognition which can not only reduce the traffic accidents in childhood, but also help them become good drivers when they grow up, and therefore improve the traffic system of the society.This article designed a questionnaire of child pedestrian's traffic knowledge, cognition, behavior, and attitude from them and their parents which suits for a developing country's city—Kunming (in China). Based on the data from 122 primary school-aged children of the city, the child pedestrian's traffic cognition and behaviors were analyzed by statistics including Pearson, Spearman's rho and Kendall's rank correlation. The results showed that there was a considerable difference between the child pedestrian's traffic cognition and behaviors ($r=0.805$, $p=0.005$), and a significant correlation between children and parents behaviors (Kendall: $rk=0.079$, $P=0.032$; Spearman: $rs=0.086$, $P=0.038$). Due to the weakness in the traffic safety education from parents, school, and government for child pedestrians, some feasible educational and practical measures on educating children's traffic safety were put forward for cities in developing countries.

Key Words: traffic safety education; cognition; behavior; child pedestrian

Introduction

With the development of traffic, children are primarily injured or killed as pedestrians who are active road-users. The WHO and UNICEF (2008) report on child injury preventions shows that road traffic crashes are the leading cause of death among 10-19 year olds globally. More than 260,000 children and teenagers die from a road traffic injury each year. That is about 718 children per day. While the accident statistics in China demonstrate that there are over 18,500 children below the age of 14 who die from traffic accidents every year. The death rate of these children's traffic accidents in China is 2.5 times more than that in Europe and 2.6 times compared to the USA (Jia, F. 2009). Traffic accidents for children have become the second leading cause of death, after drowning. It's the sixth leading cause of death in children in China. Chinese Centre for Disease Control and Prevention survey shows that children are more prone to pedestrian accidents because walking is their main way travel.

WHO (2008) reports that in high-income countries most children are killed due to the use of vehicles (exception like the Netherlands where bicycles are in common use), while in low and middle-income countries they are usually pedestrians or cyclists. It is reported that child pedestrian injury rate is highest in Africa and Asia where it is common for people to walk along roads, particularly among 5-14 year-olds. Traffic management's analysis revealed that most child road traffic accidents occurred in two main circumstances: Children suddenly appeared in the motor lane, or children crossed in front of moving vehicles. Half of these accidents were caused by children. Schools are a significant generator of localized congestion, with morning and afternoon peaks similar to that which is seen in commuting behaviors. This congestion causes problems with traffic flow on the streets near schools and may create hazardous conditions for children and adults travelling by non-motorized means. A report on the relative risks of school travel found that pedestrians and bicyclists travelling to school have the highest rate of injury and fatality on a per-mile basis (Transportation Research Board 2002). School-age children often experience injuries when they are away from home in the company of peers. In fact, it has been suggested that elementary school children are at greatest risk of injury when they are with their peers (Sally and Barbara 1997).

Barbara et al. (2010) examined how children's risk and safe behaviors developed from their knowledge, cognitions and attitudes by psychometric assessment of the instruments. It is the main reason that children's awareness of traffic safety and security of the relative lack of common sense led to multiple traffic accidents involving children. Children often do not fully understand the potential results of traffic incidents (City of Palmedium.Califolia, 2010). If children do not understand basic traffic safety, they may quickly make movements which cannot be anticipated by drivers.

There are many literatures about children's traffic safety, cognition, and behavior in developed countries because they have encountered children's traffic problems

earlier. Attributing to these large basic researches for children's traffic problems, the rate of traffic accidents of children is lower than in developing countries. With a rapid increase in the level of motorization in China, people's daily travel has become very convenient, but the accompanying road safety issues are also increasingly severe, especially children's road safety issues. To the best of our knowledge, China has only a handful of studies regarding the prevalence and characteristics of traffic safety conditions of adolescents from epidemiology in big cities (Li et al. 2002; Sun et al. 2007; Li et al. 2009). Whereas few studies have focused on child pedestrians traffic cognitions and habits from transportation. SAFEKIDS CHINA (2009, available from: http://www.safekidschina.org/news.html) and FedEx investigated the safety of walking to school of with 1850 students in Shanghai, Beijing, and Guangdong, and found some severe traffic safety problems still exist. As a developing country, the traffic situation has its particularity in China. To improve children's traffic safety, the knowledge, cognition, and behavior of them should be studied in the first in these traffic environments.

Methodology

Questionnaire Design. The questionnaire is designed to take about 15 minutes to complete according to the children's characteristics. In order to account with the children's traffic situation in China, all questions we selected are specific to the urban cities in China. For example, many parents accompany kids from and to school instead of them using school buses; there are many electric bikes on roads, many schools near to the urban main roads, and the complex order on bus stops. We designed the questionnaire after a full investigation about the traffic latent hazards of child pedestrians in Kunming.

The survey consists of 31 questions divided into four parts. The first part concerns demographic issues including age, grade, gender, and school attended.

The second part is about basic traffic information, such as the distance, school trip mode, parental companionship, the degree of knowledge of traffic markings, and the medium of acquiring the traffic safety knowledge.

The third part is intended to capture respondents' traffic knowledge, cognitions, experience, habits, and attitude (shown in Table 1). Respondents are going to be asked to select the corresponding answer from several options according to their knowledge. This part is the most important one, and the 18 questions pertain to all kinds of behaviors children have on their way to school. At the same time, the respondents are also asked to distinguish safety and dangerous behaviors.

The last part is intended to capture the children's cognitions about their parents' behaviors and attitudes toward traffic safety (shown in Table 2). Participants were asked to choose the most corresponding answer. The questions include 9 items, and 6 items are corresponding to the items of children. (Table 1: CB5, CB9, CB1, CB10, CB7, CB3 to Table 2: PB1–PB6).

Table 1. Cognition, behavior, attitude, and knowledge of items in questionnaire.

Cognition /Behavior Items	Description of the Cognition and Behaviors of Children
CB1/CB2	Chasing with classmates, listening to music, playing games, read comic books or playing mobile phone when walking on the road
CB3/CB4	How to cross the street at the intersection with traffic lights or not
CB5	How to cross roadside sidewalk when it is blocked
CB6	How to cross the street with a green light and encounter a vehicle turning right
CB7	How to get on a bus at a crowded bus stop
CB8	How to get on and off the vehicle when parents parked the vehicle on the roadside
CB9	How to cross the road when it is hardly late for school and encounter a red traffic light.
CB10	Will you choose to across the central barrier to the school gate directly
Attitude Items	Description of the Attitudes of Children
CA1	How safe do you think electric bikes are
CA2	The attitude of the driving motor vehicles on the road in front of school gate
CA3	The attitude of violating the traffic regulations
CA4	The attitude of a need for learning and training road safety knowledge
Knowledge Items	Description
R1	Medium of acquiring the traffic safety knowledge and sorted
R2	Safety sorted to travel ways from and to school and sorted
R3	Parental impact on children's behaviors
R4	Times of traffic accidents of the household because of violating the traffic regulations speed of driving or riding

Participants and Procedure. The surveys were conducted with some primary school-aged children to ensure the questions were understandable. The inadequate questions or answers were deleted. Finally, 122 primary school children in grades 2-6 from Kunming have taken part in to fulfil this survey. The two schools are located in the area with large traffic flows and the city main streets nearby. After class, research assistants distribute the questionnaires and assist the students with some instructions before they start. Afterward, students should complete ALL of the questions on BOTH sides of the paper (a total of 31 questions). They should NOT go back to change the answer after reading a subsequent question. Either, students should NOT provide their name or any other identification on the surveys except for responsing to the first four questions. When working on the questionnaire, the research assistant should make the children at ease, and ensure that the children do not consider that the questionnaire is an exam.

Table 2. Parental behaviors and attitudes items in questionnaire.

Parental Items	Description of the parental behaviors and attitudes
PB1	Walking on crosswalk
PB2	Running the red light
PB3	Driving or riding against the traffic on motor or non-motor vehicle road
PB4	Walking across the central fence
PB5	In accordance with the order line of traffic
PB6	Speed of driving or riding
PB7	Attitude to allowing child to sit in front of co-pilot position
PB8	Attitude toward the wearing of seatbelts
PB9	Attitude toward the wearing of seatbelts

Analytical Methods. In this study, the original data was clustered by grade in all. The Bivariate Correlation analyses were used to figure out the relationship between behavior and cognition. Pearson's correlation coefficient is a measure of linear association. Two variables can be perfectly related, but if the relationship is not linear, Pearson's correlation coefficient is not an appropriate statistic for measuring their association. In order to assess the degree of correlation existing between the behaviors of children and their parents, Spearman's rho and Kendall's rank correlation coefficients are used in the data analysis.

Results and Discussion

Travel Trip to School. In this investigation, 49.18% of 122 students make their trips to school via walking, bike, and electric bike; private automobiles are the other dominant mode, accounting for 27.02% of trips. Buses take 23.8%. Data on travel mode revealed that most students participated in the complex traffic environments. In addition, according to survey data the proportion of parental companionship is 68%, which also shows that parents worried about their child's traffic safety exclude the further distance to school.

Difference between Cognition and Behavior. The correct cognition rate (CCR) and safe behavior rate (SBR) of samples under 10 kinds of traffic conditions (CB1-CB10) should be seen in Table 3. From the statistical data in Table 3, we know that there is a difference between CCRs and SBRs. The SBRs are always less than CCRs, which reveal there are some reasons leading to the mismatch of children's road safety awareness and behaviors. These data also show that the children's knowledge of traffic is step on the surface of cognition. So there is a considerable difference in the cognition and behavior. Table 4 shows that the correlation between CCR and SBR is significant, and the difference is 17.53%.

Table 3. Statistics of the correct cognition rate and safe behavior rate.

	CB1	CB2	CB3	CB4	CB5	CB6	CB7	CB8	CB9	CB10
CCR(%)	78.45	79.17	77.14	90.60	81.03	62.71	86.92	99.16	93.75	84.49
SBR(%)	73.28	70.83	54.29	68.38	56.03	38.99	63.55	90.76	67.86	74.14

Table 4. The Statistics and correlation of SBR and CCR.

	Mean	Std. Deviation	Pearson Correlation	Sig. (2-tailed)
CCR	83.34%	10.20%	.805**	.005
SBR	65.81%	13.89%		

**. Correlation is significant at the 0.01 level (2-tailed).

Table 5. The parental safe behavior rate in daily trips.

Parental behavior	PB1	PB2	PB3	PB4	PB5	PB6	PB7	PB8	PB9
SBR (%)	56.56	66.67	68.33	73.98	49.17	86.32	46.22	66.39	78.81

Child pedestrian Traffic Habits. Based on the behavior data CB1 to CB10, we have also found that children paid less attention to their own safety. They are lack of the awareness of self-protection and pay less attention to traffic safety rules. This reveals the child pedestrians' traffic habits are not so good because of their physiological and psychological characteristics. Our results, although based on a fairly small student sample, confirm that children's behavior is different from adults. Children have difficulty identifying speed and distance, dealing with multiple distractions, determining right from left, and believing the safest way to cross a street is to run. All these are bad traffic habits.

Parental Traffic Behavior. In the usual nine traffic conditions, the data show that the parents' behavior is not good in daily trips (see from Table 5). The highest safe behavior rate is only 86.32%, and the lowest is 46.22%, These rates show the weakness of parental cognition of traffic safety. Because the questionnaire is answered by children not parents, the parental behavior items are answered from the child's point of view regarding their parents' daily trips. So this can be shown that parents have a major role their in children's cognition and behaviors. The parents do not only give their children a model for life but also a model for driving style that reflects both cognitive and motivational factors.

Correlation between Children's and Parental Behaviour. In order to find the relationship between children and parental traffic behavior, items items about their traffic behavior styles were rated on a 4-point scale, ranging form "never" (1) to "often" (4), which represents: 1～Patient, 2～Anxious, 3～Careless, and 4～Hostile. The mean, std. deviation, and correlation coefficients are shown in Table 6. There is a similarity in total parents and children traffic behavior style (mean=1.750 and 1.7214), which style is between patient and anxious. Further tests revealed that children' traffic behaviors are linear correlate to their parents in Item2, Item3, Item 4,

but are not in Item1, Item5, and Item 6. We can conclude that there is non-linear correlation in Item1, Item5, and Item 6. In total six items, the correlation coefficients are 0.79 and 0.86. It shows there is significant correlation between children and parents.

Table 6. Correlation of six items between children and parental behavior.

	Mean	Std. Deviation	Kendall's tau_b		Spearman's rho	
			Correlation Coefficient	Sig. 2-tailed	Correlation Coefficient	Sig. 2-tailed
CB_Total	1.750	1.0034	.079*	.032	.086*	.038
PB_ Total	1.7214	1.10172				
CB_Item1	1.438	.9039	-.018	.845	-.018	.863
PB_Item1	1.7083	1.02405				
CB_Item2	1.7129	1.12549	.415**	.000	.459**	.000
PB_Item2	1.5702	.92038				
CB_Item3	1.5660	.99541	.240**	.008	.257**	.009
PB_Item3	1.6218	1.04144				
CB_Item4	2.1771	.88252	.183*	.048	.203*	.049
PB_Item4	1.4298	.84485				
CB_Item5	1.6211	1.03327	.050	.584	.059	.572
PB_Item5	2.7833	1.34841				
CB_Item6	2.0000	.87343	.084	.381	.088	.397
PB_Item6	1.2051	.53399				

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Countermeasures for Educating Traffic Cognition and Behaviors of Child Pedestrians

Education can improve skills, behaviors, and attitudes. A systematic review of safety education for pedestrians found an improvement in behaviors and attitudes that predispose people to risk-taking (WHO and UNICEF 2008).The accident statistics clearly suggest that measures aimed at children at early school age should target the problems they have when crossing the road as pedestrians (Flavell 1992). Traditionally, we thought that children must be taught how to adapt their behavior to the demands in a motorized society and the road safety measures for children should focus largely on road safety education. Meadows (1993) claims that children can be trained to achieve cognitive abilities normally gained at a later stage in life, and further, that children at one age can be as competent at this activity as an older, untrained child. So we can conclude that childhood is the pretty time to build a positive character and to develop good traffic habits in life.

There are many countermeasures to educate children's safe traffic cognitions and behaviors in developed countries. One of the most effective education programs is "The Children's Traffic Club," it is a successful, award winning road safety

programmer, too. Children's Traffic Club began in Scandinavia. In Norway, after running for more than a decade, club members had on average a 20% lower casualty rate. In Sweden, casualty rates for 4–6 year olds were reduced by 77%.(see form :http://www.powys.gov.uk) In the UK, the children's road safety education throughout the whole process of growth, from the children walking and playing safety education into the training of young people riding bicycles. A study found that the overall reduction in the rate of road casualties by 12% of Club members, the injury rate while walking by 4%, and the casualty rate crossing the street by 20% (see from: http://www.trafficclub.co.uk). Finland has a "Children's Traffic City" in Helsinki to educate the children's knowledge and cognitions of road safety." Children's Traffic City" simulates a variety of transportation facilities including not only road sections, intersections, sidewalks, bike lanes and zebra crossings, but also the various traffic signs and traffic lights. Safe Routes to Schools is a popular program spreading across Canada and the U.S. which promotes walking and biking to school through education and incentives that show how much fun it can be (available from: http://www. Saferoutesto schools .org/ index.shtml). The studies of Tracy E. M. (2007) and Marlon G. B (2005) about California Safe Routes to School (SR2S) program shows the effectiveness of SR2S in increasing walking or bicycling to school.

While in China, although there are several children injured program in big cities such as Beijing, Shanghai, and Guangzhou, but based on the results of the questionnaire in Kunming, we found that the traffic safety education for child pedestrians was very weak. The weakness is from parents, school, and government. Specifically, the parents, despite their serious stress on their children's safety, are not bound by their own safe actions, which have been a bad influence on their children's safety awareness. At the same time, traffic safety education in schools is fragmented, unplanned, and non-systemic, which results in the lack of traffic safety education of children in their societal education responsibilities. We also found the drumbeating of traffic safety education for children is less in Kunming. There was little traffic safety education online games or other multimedia associated with the traffic characteristics of developing countries. Both result in that there has almost had no repetitive stimulation on traffic safety behaviors, leading to incorrect or unsafe behavior of a child although their traffic safety awareness. The children have no real cognitions to traffic safety. In addition, the government investment is lacking in children's safety education, which resulted in an important reason for the weakness.

In response to the above weakness, the society system should be integrated to take countermeasures and enforce the traffic safety education from the family, school, and government. In the children's safety education, the government should take into account that children are the special groups involved in the road traffic system, the design of transportation facilities should be suitable for children's cognitive behavioral characteristics, and provide support and input in the establishment of children's traffic clubs or other educational projects. Schools should combine with the government, society, and the family, in educating the children in a variety of fun and exciting ways to help them remember the important aspect of road safety. At the

end, the children can understand the importance of compliance with the social orders besides school orders. Parents play a most important role on children's cognition of traffic safety. They should fully understand the importance of traffic safety education, using every opportunity to let children know and understand the various traffic signs, and traffic safety knowledge to enable children to create a good awareness of road safety and self-protection awareness. Teaching children about traffic safety requires repetitive stimulation and their own good traffic behavior including behaving responsibly as a pedestrian, bicyclist, or driver. Through these measures to the systematic traffic safety education for children, the level of the children as well as all the people's traffic safety cognitions will be raised, and the traffic environment will be improve. These will play a positive role in promoting the society traffic. The education of children's good traffic cognition can not only reduce the traffic accidents in childhood, but also participate correctly in driving and other traffic involvements after the children grow up, as well as improve the traffic system of the society.

Conclusions

In this paper we have analyzed the questionnaire data from primary student's behavior and cognition in traffic. Our results are consistent with the general conclusion of previous research where field investigations have shown that there is a considerable difference between the behavior and cognition of child pedestrians and adult pedestrians. Some experiences of developed countries were used as a reference in this paper, and some traffic safety education countermeasures were put forward for cities in developing countries.

Further work should look more closely into the mechanisms of the relationship between cognitions and behaviors of child pedestrians' traffic habits.

Acknowledgment

The authors appreciate the help of graduate students in the traffic simulation lab of KMUST, the Wucheng and Panlong Primary Schools in Kunming.

References

Barbara, A. M., Michael, C., Benjamin K. B., Elizabeth, O., Mary, C., Jeffrey, T., Abhaya, K., Nazilla, K., Ralpsh, M., and Tsegaye, B. (2010). "Development of the BACKIE questionnaire: A measure of children's behaviors, attitudes, cognitions, knowledge, and injury experiences." *Accident analysis and prevention,* 42, 75-83.

City of Palmedium, Califolia. (2010). "Teaching Children about Traffic Safety." *http://www.cityofpalmdale.org/departments/traffic/pamp_11.html*>.

Duperrex, O., Bunn, F. and Roberts, L. (2002). "Safety education of pedestrians for injury prevention: a systematic review of randomised control trials." *British Medical Journal*,324:1129.

Flavell, J. H. (1992). "Cognitive-development — past, present, and future." *Developmental psychology*, 28, 998-1005.

Jia, F. (2009). "Seven elements of caring for the future - away from the children's traffic accidents." *http://news.cheshi.com/keji/200905/107925.shtml*.

Li, Y., Lu, W. and Zhou S. F. (2002). "Study on the traffic safety consciousness of school students in Shanghai." *Chinese Journal of Environmental and Occupational Medicine,* 19(3),155-159.

Li, Y.C., Jin, H.Q., Tao, X.Y., Zhan, S.L., and Hu, C.L., (2009). "Knowledge, attitude and behavior on road safety among middle school students in JiNan." *China Journal of School Health.*, 30(3), 232-234.

Marlon, G..B., Craig, L., Kristen, D., Tracy, M. and Mariela, A. (2005). "Evaluation of the California safe routes to school legislation." *American Journal of Preventive Medicine*, 28(2S2), 134-139.

Meadows, S. (1993). "The child as a thinker: The development and acquisition of cognition in childhood." *London: Routledge*.

Sally, C. and Barbara, A.M. (1997). "The influence of peers on children's judgments about engaging in behaviors that threaten their safety." *Journal of applied developmental psychology,* 18, 547-562.

Tracy, E. M. (2007). "The relative influence of urban form on a child's travel mode to school." *Transportation Research Part A*. 41, 69-79.

Transportation Research Board. (2002). "The relative risks of school travel: a nationwide perspective and guidance for local community risk assessment." *Special Report 269,*.http://gulliver.trb.org/publications/sr/sr269.pdf.

WHO and UNICEF. (2008). *World report on child injury prevention,* Hanoi, Switzerland.

Sun, Y., Shen, X. C., Qiao, Z. Y., Luo, H. P., Wang, Z. W., and Wang, S.M. (2007). "Awareness of traffic safety among primary and middle school students in Pudong New District in Shanghai." *China Journal of School Health,* 28(8), 685-687.

Comparison of Urban Link Travel Time Estimation Models Based on Probe Vehicle Data

Fangfang Zheng[1] and Henk van Zuylen[2]

[1] PhD candidate, Department of Transport and Planning, Delft University of Technology, Stevinweg 1, Delft, 2628CN, the Netherlands; College of traffic and transportation, Southwest Jiaotong University, Erhuanlu Beiyiduan, Chengdu, 610031, P.R.China; Email: f.zheng@tudelft.nl
2 Prof., Department of Transport and Planning, Delft University of Technology, Stevinweg 1, Delft, 2628CN, the Netherlands; Hunan University, Changsha, Hunan, P.R. China, Email: H.J.vanZuylen@tudelft.nl

Abstract

The information collected by probe vehicles on the urban signalized network is widely used for traffic monitoring. Traffic conditions could be inferred based on the data collected by these probe vehicles. In recent years, the attempts to estimate link travel time based on probe vehicle data (e.g. positions, time stamps and speeds) are arising. However, due to the low polling frequencies (e.g. 1min or 5min), travel times recorded by probe vehicles provide only partial link or route travel times. In this paper, a three-layer Artificial Neural Network (ANN) model is proposed to estimate the complete link travel time for each individual probe vehicle. The information including positions, time stamps and speeds is input into the model and the output is the complete link travel time. The model is evaluated using the data from vissim simulation model. The results are compared with those from Hellinga's model and distance-proportion model. The evaluation results show that the proposed ANN model performs much better than Hellinga's model and distance-proportion model both in undersaturated conditions and oversaturated conditions.
Key Words: urban link travel time; probe vehicle data; Artificial Neural Network

Introduction

Literature Review. Probe vehicles are widely used to collect information for traffic monitoring in Advanced Traveler Information Systems (ATIS). Based on the information (e.g. positions, time stamps and speeds) collected by the probe vehicles,

travel times can be estimated and disseminated to the travelers. Hence, the application of probe vehicle data has received considerable attention.

In recent years, model-based approaches and data-driven approaches have been developed to estimate travel time based on Probe Vehicle Data (PVD). (Liu 2009) proposed a model to estimate arterial travel time by tracing a virtual probe vehicle along an OD route with multiple intersections. High resolution 'event-based' data including every vehicle actuations over loop detector and every signal phase changes is needed to determine three possible maneuvers (acceleration, deceleration and no-speed-change) of the virtual probe. The model works quite well with very low estimation error of 1.8%. Instead of solely utilizing probe vehicle information, (Bhaskar 2009) proposed a model to estimate the average travel time on signalized urban networks by integrating the traditional cumulative plots with probe vehicle information. In undersaturated conditions, virtual probe vehicle information is used to correct the cumulative plots and reduce the relative deviations in the cumulative plot with the strong assumption that the travel time of the virtual probe vehicle is free-flow travel time of the link, while in oversaturated conditions real probe vehicle information is needed.

Basically, the traffic data collected by probe vehicles couldn't provide direct information about flow, density and average speeds which are usually the input for model-based approaches. Therefore, data-driven approaches seem most appropriate for this situation,e.g. regression based(CHAN 2008), pattern recognition based (Li 2002; Robinson 2005) and neural network based models ((Van Lint 2005; Liu 2006; Liu 2007). Neural network based models are more robust than regression models as they utilize the data to build the model structure as well as its parameters. It has been applied in different engineering discipline and has been proven to be a very powerful method of mathematical modeling, especially in modeling nonlinear relationships.
In this paper, a three-layer Artificial Neural Network (ANN) model is proposed to estimate the complete urban link travel time. For each individual probe vehicle, time stamps, speeds and positions on links at fixed polling moments are input to the neural network model and output is the travel time of this probe vehicle traversing a single complete link. In section 2, the detailed description of the ANN model is given. In section 3, the performance of the proposed model is evaluated using the data from a simulated urban network. In section 4, some sensitivity analysis of the input information on performance of the ANN model is presented. Finally, some conclusions and future research are presented.

Link Travel Time Allocation. Travel times collected by probe vehicles do not originate from a single complete link but are experienced by probe vehicles from a certain position on one link to a certain position on another link. These can be categorized into three types as illustrated in Figure 1.

P_1, P_2, P_3, P_4 are positions on the corresponding links and t_1, t_2, t_3, t_4 are time stamps. $t_{1,decomposed}$, $t_{2,decomposed}$, $t_{3,decomposed}$, $t_{4,decomposed}$, represent the reallocated link travel times based on the travel times collected by probe vehicles. The complete link travel

time here is defined as the time difference between the time instant when the vehicle passes the upstream stop line and the time instant when the vehicle passes the downstream stop line.

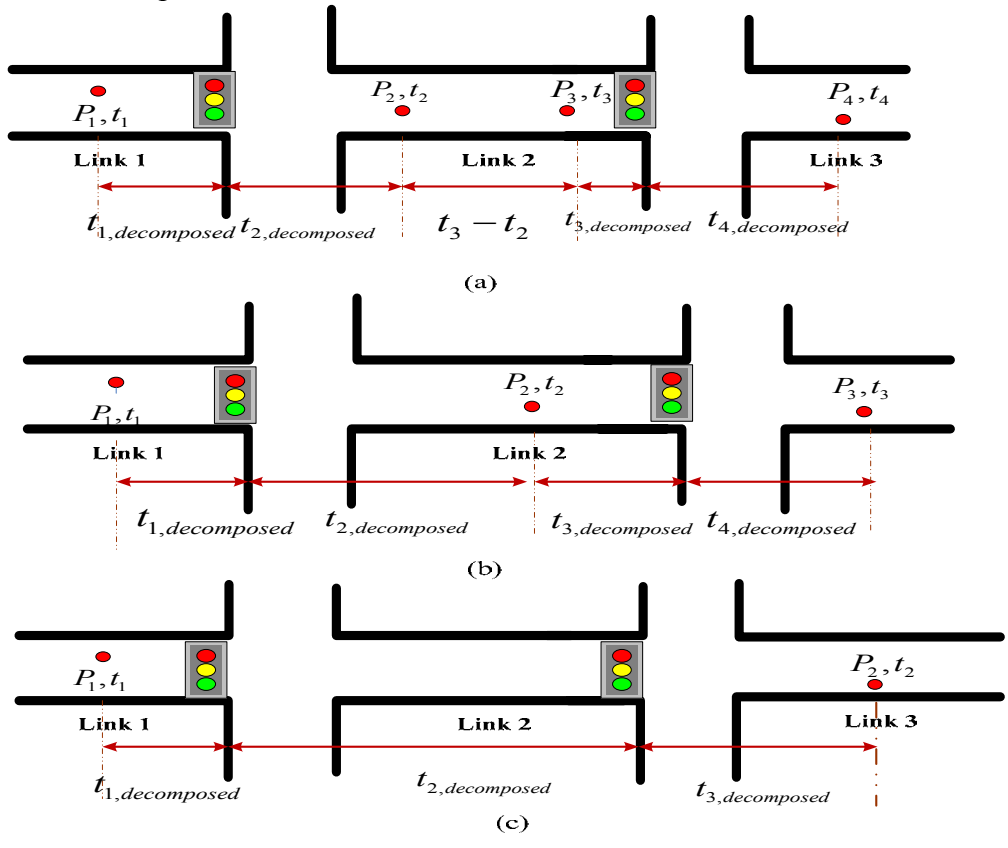


Figure 1. Sketch of assignment of travel times between recorded positions to the link in the middle.

Type 1: The reported positions are on the same link (e.g. Link 2) as shown in Figure 1(a), the complete travel time of link 2 is composed of three parts:

$$T_{L2} = t_{2,decomposed} + t_3 - t_2 + t_{3,decomposed} \tag{1a}$$

For this case, the link is long or the traffic condition on the target link is likely to be congested or vehicles need to wait for the red time since the probe vehicle experiences long travel time (at least longer than the sampling interval) on this link.

Type 2: The first and second reported positions are on adjacent links shown in Figure 1(b), then the travel time of link 2 is estimated as:

$$T_{L2} = t_{2,decomposed} + t_{3,decomposed} \tag{1b}$$

Type 3: At least one full link is existing between two consecutive reported positions illustrated in Figure 1(c), the travel time of link 2 is:

$$T_{L2} = t_{2,decomposed} \tag{1c}$$

For this case, the traffic condition on the target link is likely to be free flow or undersaturated since the probe vehicle experiences short travel time.

Now the question is how to reallocate travel times into individual links only based on probe vehicle data. In the following, an artificial neural network model and an analytical model are discussed.

Model Description

Model* 1: *Hellinga's Model. Hellinga (Hellinga, 2008) proposed an analytical model to decompose recorded partial link or route travel time into individual links considering the stopping probability and congestion probability. On the following, a brief introduction of this method is given. More detailed information about this method is given in Hellinga's original paper (Hellinga, 2008).

According to the definition proposed by Hellinga, link travel times in the urban road network can be decomposed into three parts:

- Free flow travel time;
- Stopping time caused by traffic control devices (deceleration and acceleration are included);
- Delay due to traffic congestion.

Therefore, the travel time between two consecutive time stamps of a probe vehicle can be expressed as:

$$t_{m,i+1} - t_{m,i} = \sum_{j=0}^{J(m,i)} \left\{ t_f(l_{m,i,j}) + t_s(l_{m,i,j}) + t_c(l_{m,i,j}) \right\} \tag{2}$$

where $t_{m,i}, t_{m,i+1}$ are consecutive time-stamps of probe vehicle m on link i and link $i+1$; $t_f(l_{m,i,j})$ is the free flow travel time on link j; $t_s(l_{m,i,j})$ is the stopping time on link j and $t_c(l_{m,i,j})$ is the congestion time on link j.

The free flow travel time on a link is calculated as the link length divided by the free flow speed:

$$t_f(l(n_i, n_j)) = \frac{|l(n_i, n_j)|}{v_f(n_i, n_j)} \tag{3}$$

where $|l(n_i, n_j)|$ is the length of the complete link or partial link $l(n_i, n_j)$, $v_f(n_i, n_j)$ is the free flow speed for the complete or partial link $l(n_i, n_j)$. However, in reality, free flow speeds vary with driving behavior, speed limit, weather conditions, etc. It is difficult to estimate free flow speeds. Instead, the maximum allowed speed is used to calculate the free flow travel time.

The congestion time and stopping time based on the probability function are calculated by Hellinga as:

$$t_c(l_{m,i,J}) = \int_0^{w_{\max}} \sigma_{m,i,J}\tau_c \frac{\sum_{j=0}^{J(m,i)} P_w(m,i,w)P_s(l_{m,i,J},w)}{Q_s(m,i)} dw \quad (4)$$

$$t_s(l_{m,i,J}) = \int_0^{w_{\max}} \tau_s \frac{P_w(m,i,w)P_s(l_{m,i,J},w)}{Q_s(m,i)} dw \quad (5)$$

where w is the degree of congestion which is the ratio of the congestion time on the route to the sum of the congestion time and the free flow travel time on the route; τ_c is the total congestion time; τ_s is the total stopping time; $P_w(m,i,w)$ is the congestion probability which is used to capture the likelihood of a certain degree of congestion experienced by a probe vehicle when traversing a given link m and $P_s(l_{m,i,J},w)$ is the stopping probability .

Model 2: Distance-proportion Model. The travel time between two time stamps is decomposed into individual links based on the distance as illustrated in Figure 1. For instance, the travel time of link 2 in Figure 1(a) based on distance proportion method can be derived as:

$$t_{2,decomposed} = (t_2 - t_1)\frac{L_2P_2}{L_1(1-P_1)+L_2P_2} \quad (8a)$$

$$t_{3,decomposed} = (t_4 - t_3)\frac{L_2(1-P_3)}{L_2(1-P_3)+L_3P_4} \quad (8b)$$

$$T_{L2} = t_{2,decomposed} + t_3 - t_2 + t_{3,decomposed} \quad (9)$$

where t_1, t_2, t_3, t_4 are GPS time stamps on the links; L_1, L_2, L_3 are link length; P_1, P_2, P_3 are fraction positions on link 1, 2 and 3, i.e. the position along link 1 at time t_1 is $P_1 \cdot L_1$.

This decomposition method is very simple and does not contain any other assumption than that travel times are allocated proportional to the length of the links that the probe vehicle has traveled between two registrations.

Model* 3: *Proposed ANN Model. Basically, the traffic data collected by probe vehicles include positions, time stamps and speeds on the route. Therefore, positions, time stamps and speeds can be used as the input data in the Artificial Neural Network model. Traffic flow and signal timings are considered optionally since on one hand, they are not always available on the urban road network and on the other hand, we are trying to develop a model to estimate travel time as accurate as possible with least information and make the model more generic. As discussed in (Hellinga 2008), the traffic condition the probe vehicle experiences during the most recent sampling interval is considered not substantially different from that on the route traversed by the same probe vehicle during the previous sampling interval. In our model, the probe vehicle information on previous sampling interval is incorporated with the information on the recent sampling interval. In section 3, an application of the model in terms of three types described in section 2.1 is discussed. Figure 2 shows the structure of the ANN model. Topology of an Artificial Neural Network for link travel time. The mathematical description of the model is as follows:

Input layer:

$$X(i)=\begin{bmatrix} x_1(i) \\ \vdots \\ x_n(i) \end{bmatrix}=\begin{bmatrix} p(i) \\ s(i) \\ t(i) \\ v(i) \end{bmatrix} \tag{10}$$

where $p(i)$ is the position vector of probe vehicle i on the upstream link, target link and downstream link; $s(i)$ is the link number vector indicating on which links the probe vehicle positions are; $t(i)$ is the time stamp vector which indicates the time instances when the probe vehicle sends the information; $v(i)$ is the speed vector.

$$p(i)=\begin{bmatrix} p_1(i) \\ . \\ . \\ . \\ p_j(i) \end{bmatrix},\ s(i)=\begin{bmatrix} s_1(i) \\ . \\ . \\ . \\ s_j(i) \end{bmatrix},\ t(i)=\begin{bmatrix} t_1(i) \\ . \\ . \\ . \\ t_j(i) \end{bmatrix},\ v(i)=\begin{bmatrix} v_1(i) \\ . \\ . \\ . \\ v_j(i) \end{bmatrix}$$

The number of input neurons in our model can be determined as:

$$N=n*m \tag{11}$$

where n is the number of information points taken into consideration for each probe vehicle; m is the categories of information (e.g. positions, link IDs, time stamps and speeds).

For the case in Figure 1(a), the information on the previous sampling interval is also taken into account, so the input neurons are 5*4 (5 positions+5 link IDs+5 time stamps +5 speeds) for each probe vehicle. For the case in Figure 1(b), 4*4 input neurons are used and 3*4 input neurons are needed for the case in Figure 1(c).

Hidden layer:

$$H(i)=\begin{bmatrix} h_1(i) \\ \vdots \\ h_m(i) \end{bmatrix}=\begin{bmatrix} \varphi\left(\sum_{j=1}^{n}\omega_{j,1}x_j(i)+b_1\right) \\ \vdots \\ \varphi\left(\sum_{j=1}^{n}\omega_{j,m}x_j(i)+b_m\right) \end{bmatrix} \tag{12}$$

where $h_m(i)$ denotes the value of the mth hidden neuron, $\omega_{j,m}$ denotes the weight connecting the jth input neuron and the mth hidden neuron, b_m denotes a bias with a fixed value for the mth hidden neuron; φ is the transfer function. Common forms of the transfer function are logistic sigmoid and hyperbolic tangent functions. In practice, the latter is found to give rise to faster convergence. Thus, we choose $\varphi(x)=\dfrac{1-e^{-2x}}{1+e^{-2x}}$.

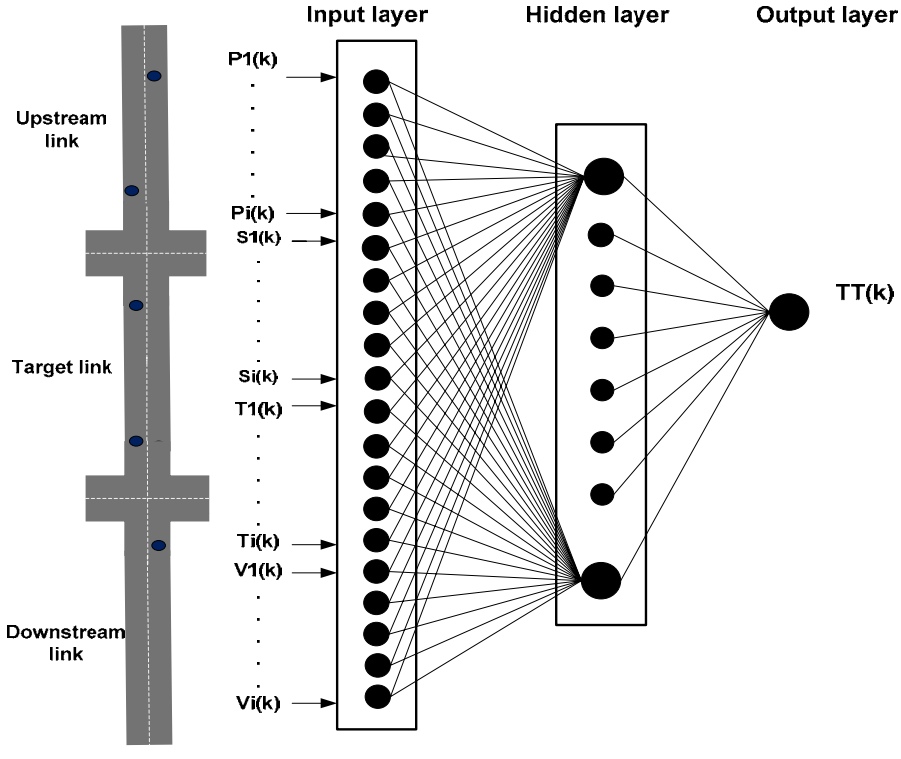


Figure 2. Topology of an Artificial Neural Network for link travel time

Output layer:

$$Y(i) = TT(i) = \phi\left(\sum_{k=1}^{m} \omega_k h_k(i) + b\right) \tag{13}$$

where $Y(i)$ denotes the estimated travel time of probe vehicle i; ω_k denotes the weight connecting the kth hidden neuron and the output neuron; b is the bias for the output; ϕ is the transfer function and a linear function is commonly used for the output units.

Model Application

Test Urban Road Network.As shown in Figure 3, an urban road called 'Kruithuisweg' in Delft city in the Netherlands was modeled using the Vissim simulation model. Kruithuisweg is a typical urban signalized road lying between two freeways, A4 and A13, in the Netherlands. In order to mimic the real traffic situation on this road, traffic was assigned using Dynamic Traffic Assignment (DTA) based the OD matrix. All the traffic signal controllers at the intersections are demand actuated. The free flow speed is set to be 80 km/h which is the speed limit on the real situation. The blue dots are the data collection points which record the information of vehicles every one second. The red arrow indicated in the figure is the target link for travel time estimation.



Figure 3. Vissim model of Kruithuisweg road in Delft, the Netherlands.

Data Preparation

Re-sampling Process. The network was simulated for a period of 65min for each simulation run. Data from the first 5min of simulation were considered to be the warm-up period and were not used in the analysis. Every second, positions, time stamps and speeds of vehicles were recorded by the data collection point. However, in real world, the sampling interval is much longer than 1s. Instead, the sampling interval of 30s or 60s is more often used in reality. Hence, a 'Re-sampling' procedure with 60s interval was taken to extract the data from the original simulation data set. One thing we should keep in mind is that the position of the probe vehicle can be anywhere on the link when sampling, which means if we just take one position on the link as the initial sampling position, the estimation results will be biased. The 'Re-sampling' strategy can be explained by Figure 4. E.g., if the initial sampling moment is i, when applying 60s sampling strategy, the next moment is $i+60$, and then $i+120$, $i+180$, etc. We can get different moment combinations to estimate the travel time of link 2.

Combination 1: $i, i+60, i+120, i+180, i+240, \ldots$

Combination 2: $i+1, i+61, i+121, i+181, i+241, \ldots$

Therefore, the average travel time of a probe vehicle k traversing link 2 can be calculated as:

$$TT_{k,L2} = \frac{1}{n}\sum_{i=1}^{n} TT_{k,i} \tag{12}$$

where n is the number of different initial sampling moments, $TT_{k,i}$ is the estimated link 2 travel time of probe vehicle k with the initial sampling moment i.

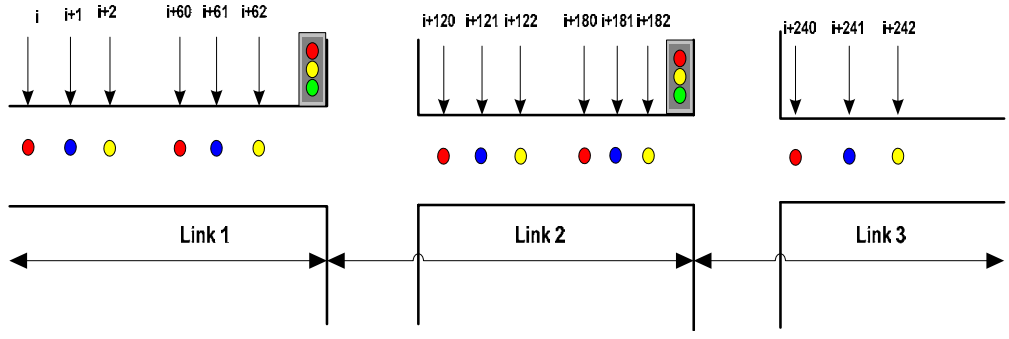


Figure 4. Different moments (positions) recorded by probe vehicles on different links

Table 1. Data sets for training and evaluation

	Training			Evaluation			
	training	validation	testing	Scenario1	Scenario2	Scenario3	Scenario4
Number of random seeds	18	6	6	10	10	10	10

The true link travel time of a probe vehicle traversing the target link is recorded by data collection points located at the beginning of the link and the end of the link.

Data for Training and Evaluation. After the 'Re-sampling' process, the extracted data were used for training the neural network and estimating the link travel time. Total 70 random seeds were simulated and probe vehicle data from 30 random seeds, in which 20 random seeds are in the undersaturated condition and 10 random seeds are in the highly oversaturated condition, were used for the training process. The other 40 random seeds were used for the performance evaluation. Four scenarios were chosen for evaluation. The subdivision of data sets for training and evaluation is indicated in

- **Scenario 1:** Original demand (undersaturated condition)
- **Scenario 2:** 20% demand increase
- **Scenario 3:** 50% demand increase
- **Scenario 4:** 100% demand increase (highly oversaturated condition)

Neural Network Training

A training process is needed before the ANN model can be applied to estimate link travel times. Three procedures including training, validation and testing were conducted in the whole training process. The total training data set was divided into three subsets which are 18 random seeds for training, 6 random seeds for validation and 6 random seeds for testing. During the training process, different hidden neurons, e.g. 10, 15, 20, 25, were chosen. The testing results show that the performance in terms of Mean Square Error(MSE) for the case of 10 and 15 neurons is not as good as that of 20 or 25 neurons. Therefore, 20 hidden neurons were used to build the network.

Levenberg-Marquardt algorithm (Ranganathan, 2004)was chosen so that the over fitting phenomenon could be avoided. Besides, the Levenberg-Marquardt algorithm can provide fast convergence even for large networks that contain a few hundred weights. The trained ANN model is applied to estimate link travel times both in undersaturated conditions and oversaturated conditions.

Evaluation

In order to evaluate how the ANN model and Hellinga's model perform, two performance indicators are used to quantify the performance, the Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(t_{PVD,i} - t_{true,i})^2} \tag{13}$$

$$MAPE = 100 * \frac{1}{n}\sum_{i=1}^{n}\left|\frac{t_{PVD,i} - t_{true,i}}{t_{true,i}}\right| \tag{14}$$

where, $t_{PVD,i}$ is the estimated travel time of the i^{th} probe vehicle; $t_{true,i}$ is the true link travel time of the *i*th probe vehicle recorded by the data collection points.

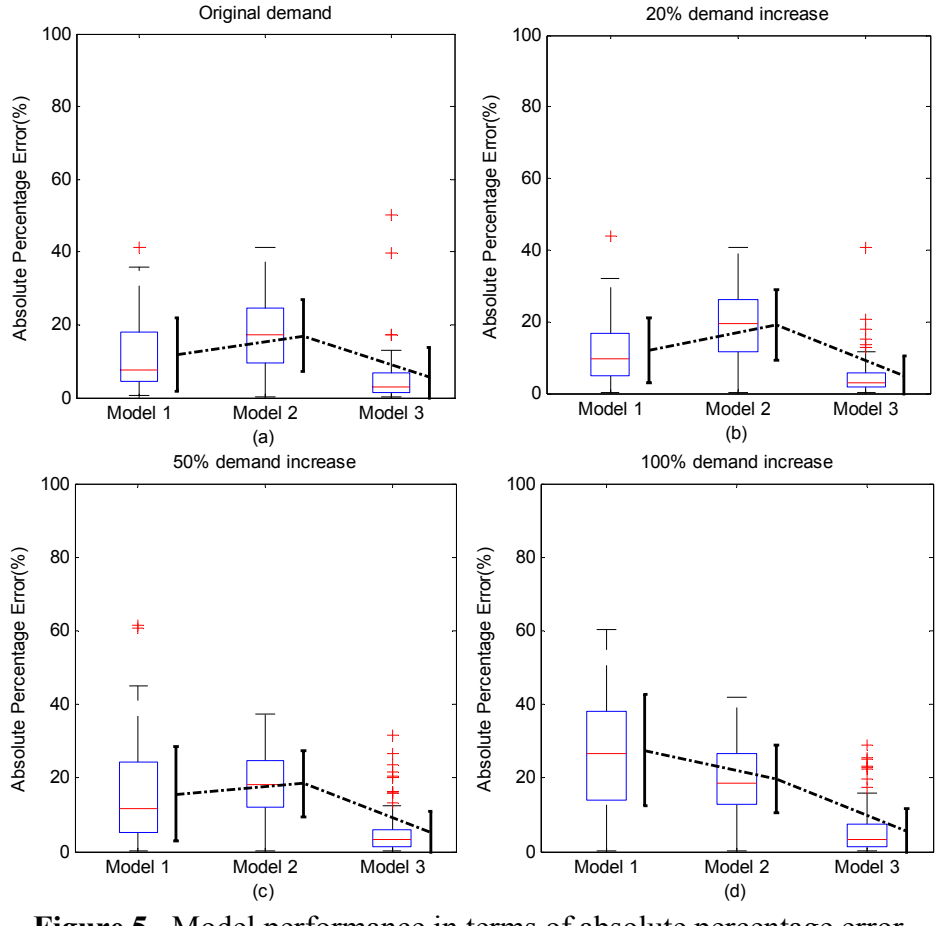


Figure 5. Model performance in terms of absolute percentage error.

Results

In order to be able to visualize both the median and spread of the Absolute Percentage Error (APE) in different models, the results are represented in box-plots as shown in Figure 5. The horizontal axis represents different models. The vertical axis represents the Absolute Percentage Error for each model. The horizontal lines in the boxes refer to the median value of APE, while the upper and lower boundaries of the boxes indicate respectively the upper quartile (75 percentile) and the lower quartile (25 percentile). The black lines on top and below the boxes show the maximum and minimal values considered, respectively. The extreme values are indicated as'+'. On the right side of each box is the error bar of the MAPE with one standard deviation. It is obvious that model 3 (ANN model) performs best among these three models. In low demand conditions as shown in Figure 5(a), (b), model 1 (Hellinga's model) performs better than mode 2 (Distance-proportion model) in terms of APE and MAPE. However, when in the highly oversaturated condition, model 1 performs even worse than model 2 as illustrated in Figure 5(d).

Discussion and conclusions

Link travel time estimation based on the travel times collected by probe vehicles is one important application of PVD. Up to now, there is not much research about travel time allocation using PVD. In this paper, a three-layer Artificial Neural Network model is proposed to estimate complete link travel times. The estimation results are compared with those from Hellinga's model and Distance-proportion model. As discussed in the previous section, the ANN model performs quite well under different traffic conditions. As for Hellinga's model, the performance decreases significantly when the traffic demand is increasing. In Hellinga's model, the link travel time is composed of free flow travel time, stopping time and congestion time. When congestion occurs, stopping time and congestion time are the main components of the estimated link travel time, which also suggests that stopping probability and congestion probability should be properly calibrated, especially when dealing with signalized intersections. The delay time can be caused by either traffic control or congestion. One thing worth mentioning is that the number of parameters in the ANN model is much more than those in Hellinga's model and Distance-proportion model. The higher performance of the ANN model is probably also due to the higher number of parameters.

The proposed ANN model in the paper deals with link travel time estimation. In future, it is recommended to apply this model to estimate the route travel time based on the PVD. Further research could be done to combine an analytical model with the ANN model to give better estimation of travel times on the urban road network.

References

Bhaskar, A., Chung, E., Dumont, A. G. (2009). "Travel time estimation on urban networks with mid-link sources and sinks," *TRB 88th annual meeting*. Washington D.C.

CHAN, K. S., TAM, M.L. (2008). "Using Spatial Travel Time Covariance Relationships for Real-time Estimation of Arterial Travel Times," *87th Annul meeting of Transportation Research Board.* Washington D.C., U.S.A.

Hellinga, B., Izadpanah, P., Takada,H., Fu,L.P. (2008). "Decomposing travel times measured by probe-based traffic monitoring systems to individual road segments", *Transportation Research Part C*: 15.

Li, Y., Mike McDonald (2002). "Link Travel time Estimation using single GPS equipped probe vehicle," *The IEEE 5th international conference on intelligent transportation systems*. Singapore.

Liu, H. (2009). "A virtual vehicle probe model for time-dependent travel time estimation on signalized arterials", *Transportation Research Part C* 17: 16.

Liu, H., H. Van Lint, H. Van Zuylen (2007). "Neural-Network-Based Traffic Flow Model for Urban Arterial Travel Time Prediction." *TRB 86th Annual Meeting*. Washington, D.C.

Liu, H., H.J. Van Zuylen, J.W.C Van Lint and M. Salamons (2006). "Urban Arterial Travel Time Prediction with State-space Neural Networks and Kalman Filters." *Transportation Research Record 1968*: 9.

Ranganathan, A. (2004). "The Kevenberg-MArquardt Algorithm." from http://www.sze.hu/~botzheim/LM/The%20Levenberg-Marquardt%20Algorithm.pdf.

Robinson, S., Polak, J.W. (2005). "Modeling Urban Link Travel Time with Inductive Loop Detector Data by Using the k-NN Method." *Transportation Research Record* 1935: 10.

Van Lint, J. W. C., S.P., Hoogendoorn and H.J., Van Zuylen (2005). " Accurate travel time prediction with State-Space Neural Networks under missing data." *Transportation Research Part C* ,13(5-6): 23.

Research Review about Moped Accidents and the Necessity of Time Dimensional Management in University Campus Traffic

Carlos Nava Fonseca[1] and Makoto Okumura[2]

[1] Doctoral Student, Graduate School of Engineering, Tohoku University, 41, Kawauchi, Aoba-ku, Sendai, Japan 9808576; Email: alberto@cneas.tohoku.ac.jp
[2] Prof., Center for Northeast Asian Studies, Tohoku University, 41, Kawauchi, Aoba-ku, Sendai, Japan 9808576; Email: mokmr@cneas.tohoku.ac.jp

Abstract

This paper discusses possible countermeasures for traffic accidents involving university students using mopeds, based on recent research findings. The first part of this paper is a review of research about traffic accidents involving mopeds. Second, we show the importance of decreasing the encounter of mopeds with large-sized vehicles in order to decrease the number of accidents, in particular the number of fatalities caused by these accidents, based on a macroscopic statistical analysis of Japanese traffic accidents classified by vehicle types. In the last part of our paper, we argue for the necessity of a time dimensional transport demand management on campus as one possible and effective way to reduce encounters between mopeds and heavy vehicles.
Key Words: moped, traffic accident, temporal traffic management, campus traffic

Introduction

Mopeds or motorized two wheeler vehicles are excellent for cost performance and user-friendly, which is why they are frequently used especially in developing countries. As a result, the increase of traffic accidents, in particular fatal accidents poses a serious problem in many countries (Salagberu, *et al*., 2006; Dandona, *et al*., 2006; Haque, *et al*., 2009). Based on empirical knowledge about problems such as a high mortality in accidents involving head injuries, frequent occurrence of accidents on right turns in intersections and frequent overlooking of mopeds by car drivers, several countermeasures have been taken in Japan, such as the compulsory use of helmets, two-step right turns in intersections, full-time lighting-up of the head light of mopeds (ITARDA, 2004). Recently, new research on the modeling of positioning

behavior of mopeds in lanes, and re-design of road sectional layouts to avoid the moped's passing-by other vehicles through a narrow clearance are executed and the restructuring of roads is now under consideration (Inagagi, *et al*., 2007). Some of these measures have resulted in the decrease of moped accidents in Japan, and many aspects of the Japanese case may be applicable to other countries.

A typical location of mopeds accidents is on the road to university campus, where new students and less experienced moped riders replace more experienced drivers upon graduation every year. In Japan, the driving license for mopeds can usually be acquired without actual field driving training or a practical driving test. Many riders are new license holders without any driving experience of larger motor vehicles. They include young riders who easily take offensive actions due to excessive trust in their own driving ability. These factors make safety education measures difficult and less effective. Considering the economic limits of students, they cannot change their travel modes to more expensive transportation means such as four-wheel automobiles. Therefore, we cannot solely rely on safety educational measures that are limited due to the migration cycle of students. However, hardware countermeasures, such as road section redesign, are also limited in their applicability in particular given spatial and budget constraints.

This paper discusses possible countermeasures for traffic accidents involving university students using mopeds, based on recent research findings. The first part of this paper is a review of research about traffic accidents involving mopeds. Second, based on a macroscopic statistical analysis of Japanese traffic accidents classified by vehicle types, we show the importance of reducing the encounters of mopeds with large-sized vehicles, in order to decrease the number of accidents and reduce the fatalities caused by these accidents. In the second part of this paper, we argue for the necessity of a time dimensional transport demand management on campus as one possible and effective way to reduce the encounters between mopeds and large-sized vehicles.

Review of Research about Moped Accidents

This section presents a comprehensive review of recent research papers about moped accidents, ranging from international journals such as Accident Analysis and Prevention, Injury, Injury Prevention, Journal of Safety Research, and so on. These papers are reviewed according to the following classifications;

- Overview of the occurrence trend of moped accidents,
- Research about the responsible persons in moped accidents,
- Research about the cause of the accident with four or more wheel vehicles,
- Research about the temporal and spatial distribution of moped accidents,
- Research about the influence of the confronting car types.

Occurrence trend of moped accidents. Due to being less expensive compared to four-wheel vehicles due to their flexibility even on narrow roads in downtown areas, the use of mopeds increases rapidly in the early stages of the economic development

of a country. At the same time, the number of moped accidents also increases rapidly (Salagberu, *et al*., 2006). Legislative countermeasures and control by police usually cannot be executed prior to the rise of accidents. Head injuries of moped riders who do not wear a helmet that result in fatal moped accidents have become one of the main causes for traffic-related deaths in developing countries (Dandona, *et al*., 2006).

Along with economic growth in later development stages of a country, the registered number of mopeds usually decreases. However, poorer people with less education continue to use mopeds. As a result, the occurrence rate of moped accidents does never decrease through economic development (Haque, *et al*., 2009). For example in Japan, the registered number of mopeds marked 18,000,000 at its maximum in 1985 and decreased to 11,000,000 in the following two decades. The number of victims of mopeds accidents decreased accordingly from the peak number of 2,575 in 1989 to half of that number in 2003, but the occurrence rate of the accidents did not decrease (ITARDA, 2004). It is remarkable in contrast with the decrease of fatalities and serious injury percentages of accidents involving four-wheel cars, where protective air-bag and seat belts on all seats have been widely installed in the past decades.

Responsible person of the moped accidents. In any country, the responsible person and the victim killed in moped accidents are concentrated among young male riders without major driving experience (Natalier, 2001). Whether such a concentration owes to "the low age" or more specifically to " insufficient driving experience" has been debated among researchers. In a mail survey sent to moped owners in the United Kingdom, the number of accidents proved to be dependent on low age along with a more relaxed attitude toward adhering to laws and traffic rules, but the number of accidents did not correlate with little driving experience (Rutter and Quine, 1996). A comparative study of drivers responsible for mopeds accidents and roadside surveys in New Zealand showed the statistical significance of low age for the accident occurrence ratio, but again there was no dependency on the number of years of experience. These studies proved that the large number of male drivers involved in accidents is simply due to the higher usage of mopeds of men compared to women (Mullin, et al., 2000). A study based on the Spanish ten years accident database revealed that young male moped riders and drivers tend to drive at higher speed higher than other older groups. When considering the difference of speed, the accident occurrence rate for elder drivers is higher (Lardelli-Claret, *et al*., 2005). We can conclude that young men's aggressive tendency toward driving at highspeed results in a higher occurrence of accidents and accidents involving bicycles and pedestrians. A consciousness investigation of driving license holders in Norway reveals that teenage drivers tend to consider the use of the motorcycle as a symbol of being free, independent and young. They insisted on the importance of sociological research on the formation process of such tendencies (Nja and Nesvag, 2007).

No clear conclusions have yet been drawn on whether safety education could change the consciousness and attitude of drivers. A comparative study on the participants of two-wheel riders in a lecture program in Ontario State in Canada from 1974 to 1977 an samples of drivers of equal age and gender showed a lower percentage of traffic

rule violations, but no significant difference of accident occurrence between the groups (Jonah, *et al*., 1982). In a similar research about the lecture program in 1979 in British Columbia State in Canada, attendance to the program did not have any significant effects either on rule violation or accident occurrence percentage (McDavid, *et al*., 1989). However, these comparative methods cannot statistically measure the effect of lecture because drivers who seek to improve safety tend to attend these lectures. An additional panel data analysis to verify a change before and after the education for the identical sample will be necessary.

Cause of the accidents with four or more wheelers. It is important to clarify the cause of collision accidents with other cars in order to review the avoidance plan of fatal moped accidents.

A research group in Southern California University investigated moped accidents in Los Angeles in 1976 and 1977 and showed that car's interference with the passage of the moped at the intersection caused two thirds of mopeds-car accidents (Hurt, *et al*., 1982). They suggested the constant use of moped's headlight and reflective color jacket wearing of riders as effective countermeasures to avoid being overlookeded by a car. This research also showed that 96 percent of the victims involved in fatal accidents are men, and that drinking prior to the moped ride was involved in half of the cases. A strict ban of drinking while driving and the wearing of helmets were proposed. However, another paper summarizing 21 experiment reports concluded that the always headlight lighting-up doesn't have any effect in an actual traffic environment (Wolf, *et al*., 1989). Wolf, et al., (1989) insisted that even if a moped can be seen by the car driver, if the car driver doesn't have any experience of encountering a moped before, the driver cannot recognize the motorbike as a dangerous object. Another report (ITARDA, 2008) showed that even if the existence of the moped is recognized, errors occur in case the speed and the distance of the moped are not estimated accurately. A comparative study in the EU on drivers involved in moped accidents and selected driver samples from 1999 to 2001 confirm that car drivers who have had operation experiences of a moped can understand the movement of a moped, and that they can sense the danger of a moped accident prior to the accident (Magazzu, *et al*., 2006).

Quite a few research results attribute the cause of moped-car accidents to the moped rider side. For example, research conducted in the United Kingdom based on a video indoor experiment and a street running experiment found a tendency of higher speed, more frequent passing and smaller distance ahead of the following car at motorbike driving comparing to for-wheel car driving, even for the identical person (Horswill and Helman, 2003). An analysis of video images taken from long-range trailer trucks that drove in southern parts of the United States and from delivery tracks in the State of Virginia attributed the cause of 80 percent of the dangerous situations to the two-wheel vehicle. Frequent entering and passing behavior with insufficient headway are detected from the video images (Hanowski, *et al*., 2007). However, the study also observed cases with causes on the side of the road; Incompetent operation of large-

sized cars, such as approaching with insufficient headway distance between cars, sudden backward movements, the swelling curve passage and slow braking operation.

Temporal and spatial distribution of moped accidents. When thinking of accident prevention measures, it is important to know the temporal and spatial distribution of accidents, but the relative occurrence ratio is difficult to calculate because we cannot comprehend real traffic distributions to be used as a denominator. It is also difficult to determine whether serious accidents occur at a higher rate during the night or whether a higher fatality rate correlates with a longer time span needed for rescue transportation at night, when we observe a higher frequency of fatal accidents during the night than during day time. It is desirable to analyze the occurrence of the dangerous phenomenon through continued probe video images (Hanowski, *et al*., 2007).

According to previous research, there is a remarkable difference between the places of the occurrence of moped accidents and the time between single moped accidents and moped accidents that involve cars. Single moped accidents frequently occur when the road surface is wet on roads with speed limit such as highways, whereas moped-car accidents tend to occur on lower ranked roads downtown, where moped traffic for daily business purposes during peak hours are often involved (Lapparent, 2006). During peak hours, the attention of car drivers tends to concentrate on larger vehicles and pedestrians just around the car, which is when a moped can be easily overlooked. Moped-car accidents at intersections are also observed frequently near the origin of short distance trips for daily purposes, and not related to the road structure or weather conditions (Hurt, *et al*., 1982). However, single moped accidents at night can easily turn into serious and fatal accidents due to decreased sight and higher speed (Lapparent, 2006). Research has been conducted on the installation effect of roadside ropes (Pieglowski, 2005) that prove to have a positive effect on the rope for the reduction of fatalities in severe accidents and single accidents during night hours; whereas no effect is expected for collisions that frequently occurred in the downtown area.

In some countries, strong seasonal differences in accident occurrence are observed. For example, in the northern Iranian region, fatal moped accidents tend to occur on summer weekends, and 15 percent of annual fatal accidents are registered respectively in June and in September due to the increase of traffic (Janmohammadi, *et al*., 2009). 82 percent of all fatal accidents occur during daytime, but at the same time, around 20 percent occur within one hour starting from 9 p.m., which is the peak hour in the evening. Another statistical study conducted in Indiana State in the United States (Savolainen and Mannering, 2007) reports a high occurrence of single moped accidents in April (9.3 percent) and July (16 percent). After the adjustment of the influence of alcohol intake, which is strong in July, they calculated that the seasonal elasticity in April of single moped accident fatalities is 111 percent higher than 98 percent for July. In the same geographical area, 11 percent of the moped-car accidents occur in April, but according to the smaller helmet-wearing ratio and the lower driving speed, Savolainen and Mannering, (2007) estimate that the fatality

propensity is as low as -64 percent in April. They describe that a higher number of new or inexperienced riders in April, who resume driving after the severe winter months, causes the high occurrence of single accidents in April. However, fatal collision with other vehicles seems to be less frequent in April.

Influence of the Confronting Car Type. The larger the confronting vehicle, the greater the energy of the crash becomes and the more severe the result of the accident in moped-car accidents A calculation of a model of dynamics shows that the degree of damage is proportional to the 2.5th power of the length ratio of the two vehicles (Wood and Simms, 2002). A Bayesian statistical analysis of the slight wounds, serious injuries, and fatal accidents compared to the number of no damage accidents in France shows a higher occurrence of fatal results when the confronting vehicle is a four-wheel car, while the fatal ratio is significantly smaller if the other party involved is a pedestrian, a bicycle, or a two wheeler (Lapparent, 2006). In the same paper, public transit vehicles (tram or bus) give mopeds with a smaller occurrence of accidents involving injuries, but a higher possibility of fatal results, however, these results are not statistically significant. The Indiana moped accident study referred to earlier, shows a higher occurrence of serious injuries in case of collision with a pickup truck, and a higher occurrence of fatal accidents in case of accidents with large-sized trailer trucks (Savolainen and Mannering, 2007). Furthermore, Harrop and Wilson, (1982) report a fatal accident involving a moped through a front-to-front crash with an approaching car on the reverse lane, after the moped driver's attempt to pass by a stopping bus.

Based on the studies above, we conclude that a situation in which mopeds and large-sized vehicles such as buses and trucks mingle at the same time is not desirable for traffic safety.

Table 1. Number of vehicle accidents in Tokyo in 2007

First party (responsible) \ Second party (affected)		Passenger car		Cargo/freight		Mopeds	y_i
		bus	compact & normal	heavy	compact & normal		
Passenger car	bus	1	91	2	18	58	170
	compact & normal	86	11054	66	2513	8456	22175
Cargo/freight	heavy	3	386	55	173	132	749
	compact & normal	26	3927	47	1444	2824	8268
Mopeds		13	1094	38	298	913	2356
z_j		129	16552	208	4446	12383	33718

Moped and Large Vehicle Mixture and Accident Occurrence

We confirm that the mixture of mopeds and large-sized cars poses a problem in terms of the occurrence percentage and the fatality of traffic accidents, based on statistical analysis of a Japanese accident statistic database. The data to be analyzed here is a traffic accident macro statistic database in the Tokyo Area in the year 2007 by the Institute for Traffic Accident Research Data Analysis (ITARDA).

Vehicle Type Combinations in Accident Occurrence. The first analysis was conducted to determine the number of accident occurrences categorized by vehicle type of the first party, the more responsible actor in the accident, and the second party, the more affected actor, based on the observed numbers of accidents as shown in Table 1.

At first, we set a null hypothesis that the occurrence of traffic accidents is dependent on the vehicle types of the two parties, but independent from the combinations of these two parties. Given that hypothesis, we can estimate the expected number of accidents occurring between the two parties as follows;

$$\bar{a}_{ij} = T\frac{y_i}{T}\frac{z_j}{T} = \frac{y_i z_j}{T}$$

where, i is a vehicle type of the first (responsible) party. j is a vehicle type of the second (affected) party. y_i is the number of accidents, in which the vehicle type of the first party involved is i. z_j is the number of accidents, in which the vehicle type of the second party involved is j. T is the total number of accidents. $\bar{a}_{ij}$ is the expected number of accidents for the combination of the types.

The chi square value for the test of independence is given as follows;

$$\chi_a^{\,2} = \sum_{ij}\left(\frac{a_{ij} - \bar{a}_{ij}}{\bar{a}_{ij}}\right)^2$$

where, a_{ij} is the observed number of accidents between the two parties i and j.Comparing this test value with the value of chi-square distribution, we determine that $\chi_a^{\,2} = 124 > \chi^2(16,0.05) = 26.3$. The result of this chi-statistic analysis shows a significant interdependence of accident occurrence based on the combinations of vehicle types of the two parties involved. According to the comparison of a_{ij} with $\bar{a}_{ij}$, we find that the numbers of accidents between mopeds (as first party) and large-sized vehicles such as buses and heavy trucks are higher than expected under the independent hypothesis. Similarly, when a small and medium-sized four-wheeler passenger car is the responsible actor, the occurrence rate with mopeds is high.

Vehicle Type Combinations in Serious and Fatal Accidents. The second analysis focuses on the seriousness of the accidents categorized by the vehicle type, based on the number of the fatal and serious injury accidents as shown in Table 2.

Table 2. Number of serious injury and fatal accidents in Tokyo in 2007

First party (responsible) \ Second party (affected)		Passenger car		Cargo/freight		Mopeds	Subtotal
		bus	compact & normal	heavy	compact & normal		
Passenger car	bus	0	0	0	0	1	1
	compact & normal	0	16	4	10	124	154
Cargo/freight	heavy	0	0	2	3	10	15
	compact & normal	0	5	4	9	56	74
Mopeds		1	35	14	26	13	89
Subtotal		1	56	24	48	204	333

We set a null hypothesis that the seriousness of the accident is independent from the combination of vehicle types. The expected number of serious accidents for each vehicle type combination is calculated as follows:

$$\bar{o}_{ij} = a_{ij} \frac{O}{T}$$

where, O is the total number of serious accidents. $\bar{o}_{ij}$ is the expected number of serious accidents for the combination of vehicle types.

The chi square value for the test of independence is given as follows;

$$\chi_o^2 = \sum_{ij} \left(\frac{o_{ij} - \bar{o}_{ij}}{\bar{o}_{ij}} \right)^2$$

where, o_{ij} is the observed number of serious accidents that occur between two bodies' vehicle types i and j. Comparing this test value with the value of chi-square distribution, we determine that $\chi_o^2 = 1579 > \chi^2(24, 0.05) = 36.4$.

The result shows the statistically significant effects of vehicle type combinations on the seriousness of accidents Specifically from the comparison of o_{ij} with $\bar{o}_{ij}$, we find that there are many seriously injured persons and fatalities caused by accidents involving buses and trucks, when the moped driver is the responsible actor. Also high numbers of fatalities can be observed in accidents involving mopeds and large-sized vehicle drivers where large-sized vehicle drivers are the responsible party.

Necessity of Time Dimensional Traffic Management on University Campus

According to the literature review and empirical analysis of accident data collected in

Japan, we have confirmed that the encounter of mopeds and large-sized vehicles have had a significant impact on the occurrence of traffic accidents and the seriousness of accidents. Especially on a university campus, where construction works are going on, a mixture of mopeds and large-sized construction vehicles pose serious challenges for traffic safety.

Considering the difficulty of modal shifts from mopeds to more expensive four-wheelers, we must investigate possible countermeasures in order to decrease the mixture of the two type vehicles in the traffic flow.

We conclude that temporal time dimensional traffic management is required on a university campus, where the moped traffic concentrates prior to the beginning time of classes. Large-sized construction vehicle schedules must be coordinated in avoidance of the peak hours of student commuters using mopeds as a means of transportation.

References

Dandona, R., Kumar G.A. and Dandona, L.(2006). "Risky behavior of drivers of motorized two wheeled vehicles in India." *Journal of Safety Research*, 37, 149-158.

Hanowski, R. J., Hickman, J. S., Wierwille, W. W. and Keisler, A. (2007). "A descriptive analysis of light vehicle – heavy vehicle interactions using in situ driving data." *Accident Analysis and Prevention*, 39, 169-179.

Haque M., Chin H. C. and Huang, H.(2009). "Modeling fault among motorcyclists involved in crashes." *Accident Analysis and Prevention*, 41, 327-335.

Harrop, S. N. and Wilson, R. Y. (1982). "Motorcycle fatalities in South west Cumbria." *Injury*, 13(5), 382-387.

Horswill, M. S. and Helman, S. (2003). "A behavioural comparison between motorcyclists and a matched group of non-motorcycling car drivers: factors influencing accident risk." *Accident Analysis and Prevention*, 35, 589-597.

Hurt, H. H. Jr., Ouellet, J. V. and Thom, D. R. (1982). *Motorcycle accident cause factors and identification of countermeasures Volume I*: *Technical Report*, Traffic Safety Center, University of Southern California.

Inagaki, T., Hino, Y., Uchida, T. and Yoshida, O. (2007). "Validation and application of a Quantitative Risk Model for Motorized Bicycles." *Conference of the Japan Society of Traffic Engineers*, 27, 25-28. (in Japanese)

Institute for Traffic Accident Research and Data Analysis (ITARDA) (2004). *The ITARDA Information*, No.52, http://www.itarda.or.jp/ info52/ info52.pdf (accessed March 14,2009) (in Japanese)

Institute for Traffic Accident Research and Data Analysis (ITARDA) (2008), *The ITARDA Information*, No.75, http://www.itarda.or.jp/ info75/ info75.pdf (accessed March 14,2009) (in Japanese)

Janmohammadi, N., Pouthossein, M. and Hasheml, S. R. (2009). "Pattern of motorcyclist's mortality in Mazandran Province, Northern Iran". *Injury Red Crescent Medical Journal*, 11(1), 81-84.

Jonah, B. A., Dawson, N. E. and Bragg, W. E. (1982). "Are formally trained motorcyclists safer?" *Accident Analysis and Prevention*, 14(4), 247-255.

Lapparent, M. (2006). "Empirical Bayesian analysis of accident severity for motorcyclists in large French urban areas." *Accident Analysis and Prevention*, 38, 260-268.

Lardelli-Claret, P., Jimenez-Moleon, J. J., Luna-del- Castillo, J. D., Garcia-Martin, M, Bueno-Cavanillas, A. and Galvez-Vargas, R. (2005). "Driver dependent factors and the risk of causing a collision for two wheeled motor vehicles." *Injury Prevention*, 11, 225-231.

Magazzu, D., Comelli, M. and Marioni, A. (2006). "Are car drivers holding a motorcycle license less responsible for motorcycle-car crash occurrence? A non-parametric approach." *Accident Analysis and Prevention*, 38, 365-370.

McDavid, J. C., Lohrmann, B. A. and Lohrmann, G. (1989). "Does motorcycle training reduce accidents? Evidence from a longitudinal quasi-experimental study." *Journal of Safety Research*, 20, 61-72.

Mullin, B., Jackson, R., Langley, J. and Norton, R. (2000). "Increasing age and experience: are both protective against motorcycle injury? A case-control study." *Injury Prevention*, 6, 32-35.

Natalier, K. (2001) "Motorcyclists' interpretations of risk and hazard." *Journal of Sociology*, 37(1), 65-80.

Nja, O. and Nesvag, S. M. (2007). "Traffic behavior among adolescents using mopeds and light motorcycles." *Journal of Safety Research*, 38, 481-492.

Pieglowski, T. (2005). *The influence of wire rope barriers on motorcyclists*, Master's Thesis, Lulea University of Technology, Sweden.

Rutter, D. R. and Quine, L. (1996) "Age and experience in motorcycling safety." *Accident Analysis and Prevention*, 28(1), 15-21.

Salagberu, B. A., Ofoegbu, C. K. P., Nasir, A. A., Ogundipe, O. K., Adekanye, A. O. and Abdur-Rahman, L. O.(2006). "Motorcycle injuries in a developing country and the vulnerability of riders, passengers, and pedestrians." *Injury Prevention*, 12, 266-268.

Savolainen, P. and Mannering, F. (2007). "Probabilistic models of motorcyclists' injury severities in single- and multi-vehicle crashes." *Accident Analysis and Prevention*, 39, 955-963.

Wolf, G., Hancock, P. A. and Rahimi, M. (1989). "Motorcycle conspicuity: an evaluation and synthesis of influential factors." *Journal of Safety Research*, 20, 153-176.

Wood, D. P. and Simms, C. K. (2002). "Car size and injury risk: a model for injury risk in frontal collisions." *Accident Analysis and Prevention*, 34, 93-99.

Future Forecast of Chinese Trade Amount and International Cargo Flow

Ryuichi Shibasaki[1] and Tomihiro Watanabe[2]

[1] Senior Researcher, National Institute of Land and Infrastructure Management, Ministry of Land, Infrastructure, Transport and Tourism, Japan, 3-1-1 Nagase, Yokosuka, Kanagawa, 2390826, Japan, Email: shibasaki-r92y2@ysk.nilim.go.jp
[2] Head of Port System Division, National Institute of Land and Infrastructure Management, Ministry of Land, Infrastructure, Transport and Tourism, Japan

Abstract

This paper forecasts the future trade amount and international cargo flow, especially focused in China, using by the stepwise trade amount and international logistics flow forecasting system in East Asia that the author had been developed. For the trade forecasting sub-model, the Global Trade Analysis Project (GTAP) model, the most famous model to predict world's trade amount applying Spatial Computable General Equilibrium (SCGE) model, is applied after a few modifications. A future scenario is prepared on the future economic direction based on interview survey to many experts, including advancement of economic cooperation such as FTA. For the international cargo flow forecasting sub-model, the original model developed by the authors in order to reproduce international logistics flow on multimode transport network in East Asia including road, land, and ferry transport as well as international maritime shipping is applied, where cargo transport demand (i.e. OD cargo volume) between regions are given. In the future simulation using this sub-model, the future investment plan of the transport infrastructure such as new container terminal construction in Chinese ports should be considered. Furthermore, by comparing the calculation result without any consideration of these policies, the effect of these policies can be quantitatively measured in terms of transport cost reduction.
Key Words: future forecast; trade; international logistics; SCGE model; Nash equilibrium model

Introduction

China still keeps high economic growth during the past decade. Chinese economy seems to already recover from the worldwide economic downturn after 2008. As

generally known, such rapid growth of Chinese economy mainly depends on the export of the goods. Therefore, the volume of international cargo shipping had also increased rapidly during the past decade.

Table 1 shows the container throughput ranking in 2008. As shown in the table, 6 Chinese ports (including Hong Kong) were ranked within top 10 container ports. In addition, in many Chinese ports, the increasing rates from the volume a year ago were relatively larger than those in other countries' ports.

However, do such high growth rate in the trade amount and the volume of international cargo shipping continue forever? Under the expectation that the economic growth rate in China will gradually decrease, the growth rate of the trade and the international cargo shipping are also expected to gradually decrease. Therefore, this paper aims to forecast the future trade amount, especially focused on China, according to a future scenario on the world economy based on the questionnaire survey to nearly 100 experts.

Another objective of this paper is to forecast the future flow of international cargo, reflecting the future trade amount estimated above and the future investment to port infrastructure in East Asian countries. Also, the impacts of the infrastructure investment are estimated, from the viewpoint of reduction of international shipping cost.

Table 1. Container throughput ranking in 2008.
(source: Containrisation International, March 2009)

rank 2008	rank 2007	port	country/region	throughput (million TEU)	increasing rate from previous year
1	1	Singapore	Singapore	2,992	7.1%
2	2	Shanghai	China	2,798	7.0%
3	3	Hong Kong	China	2,425	1.0%
4	4	Shenzhen	China	2,141	1.5%
5	5	Busan	Korea	1,343	1.3%
6	7	Dubai	UAE	1,183	11.1%
7	11	Ningbo/Zhoushan	China	1,123	20.0%
8	12	Guangzhou	China	1,100	19.6%
9	6	Rotterdam	Netherlands	1,080	0.1%
10	10	Qingdao	China	1,032	9.1%
11	9	Hamburg	Germany	970	−2.0%
12	8	Kaohsiung	Chinese Taipei	968	−5.7%
13	14	Antwerp	Belgium	866	5.9%
14	17	Tianjin	China	850	19.7%
15	16	Klang	Malaysia	797	11.9%
16	13	Los Angeles	USA	785	−6.1%
17	15	Long Beach	USA	649	−11.2%
18	18	Tanjung Pelepas	Malaysia	560	1.8%
19	20	Bremen/Bremenhaven	Germany	550	12.5%
20	19	New York/New Jersey	USA	524	−1.1%
21	21	Laem Chabang	Thailand	513	10.6%
22	22	Xiamen	China	503	8.6%
23	26	Dalian	China	450	−1.5%
24	25	Tokyo	Japan	427	3.6%
25	23	Tanjung Priok	Indonesia	418	1.0%
26	24	Jawaharlal Nehru	India	418	3.0%
27	29	Colombo	Sri Lanka	369	9.2%
28	34	Valencia	Spain	359	18.1%
29	28	Yokohama	Japan	349	1.7%
30	27	Gioia Tauro	Italy	347	0.6%

Literature Review

The Global Trade Analysis Project (GTAP) model, developed by Hartel et al. at Perdue University in the U.S., is a typical computable general equilibrium (CGE) model based on microeconomic theory, in which the behaviors of all economic bodies are simultaneously considered to reach equilibriums based on the balance of supply and demand. The model was often used not only in academic field, but also in intergovernmental cooperation efforts, such as by APEC (1997) to evaluate the effects of the Manila Action Plan and by GATT (1994) to evaluate the effects of the Uruguay Round. Until now, the authors (Shibasaki et al. 2005a, Shibasaki et al. 2008) examined the impact of international economic policies such as conclusion of Free Trade Agreement (FTA) on the amount of trade and international cargo shipping, by utilizing the GTAP model. However, those simulations were static as in year 2003; i.e. they did not forecast any future. This research tries to estimate the future trade amount, by incorporating the GTAP model continuously in five-year steps. The authors (Shibasaki et al. 2009a) also conducted the questionnaire survey on the future of the world economy. By inputting a future scenario based on these survey results into the model, the amount of future trade can be estimated.

There are several other researches on forecasting the future trade amount, especially by applying SCGE model. For example, Kunita et al. (2005) estimated the future trade amount especially focused in Japan, by using an original dynamic SCGE model. ERIA (2008) also estimated the future economy and international transport reflecting the future infrastructure investment in the ASEAN regions. In this paper, taking to apply to international cargo flow simulation into account, a simple application of the GTAP model is selected for forecasting the future Chinese economy.

For international cargo flow simulation model, although there are many papers to develop the models on international container cargo flow from one side of shippers (e.g. Malchow and Kanafani, 2001) or ocean-going carriers (e.g. Shintani, et al., 2007), in only few papers interaction of both actors' behaviour is considered such as Takebayashi et al. (2004). However, even in these few papers, the simulation targeting the entire East Asia including multimodal transport network including both maritime shipping and land transport is very difficult because the calculation based on neither strict solutions such as linear programming nor heuristic algorithms such as Genetic Algorithm is too complex to solve on such a huge network. Therefore, the author (Shibasaki et al. 2005b, etc.) developed the original model that meets the above condition by applying the network assignment methodology like FNEM that had developed by Friez et al. (1986) and extended it to include the land transport network of China (Shibasaki et al., 2007) and the ASEAN region (Shibasaki et al. 2009b). Using this extended model and inputting the estimated future amount of cargo transport demand on a regional basis calculated from the output of the trade model, the future simulation of international cargo flow is conducted.

Methodologies

The entire structure of the simulation system of future logistics developed by the authors is shown in Figure 1. The system mainly consists of two models: 1) the Trade Forecasting Model and 2) the International Cargo Flow Simulation Model. Into the first model—the Trade Forecasting Model, by inputting the future economic conditions of each country, such as the population growth rate, GDP, capital, etc., and the policies concerning international economics, including Free Trade Agreements (FTAs) and Economic Partnership Agreements (EPAs), the estimated amount of trade with each partner country is acquired as output. This output (future trade amount on a value basis) is transformed into international cargo demand (OD cargo volume) by region on a tonnage basis and a twenty-foot equivalent unit (TEU) basis, by multiplying the value per unit weight, modal share, containerization ratio, and weight per TEU in incremental steps.

In the next step, the international cargo demand calculated above is input into the second model—the International Cargo Flow Simulation Model, in order to simulate the cargo flow in the land and maritime transport network. By aggregating the output flow of all ports, the volume of exported, imported, and transshipped containers handled at each port are estimated. When simulating the effects of the infrastructure policies, such as construction and improvement of roads, railways, and ports, the data reflecting the improved network should be input into this model. By comparing the flow pattern and calculating the total transport cost in the case where these policies have been implemented and in the case where they have not been implemented, the effects of policies, such as economic benefits, can be estimated.

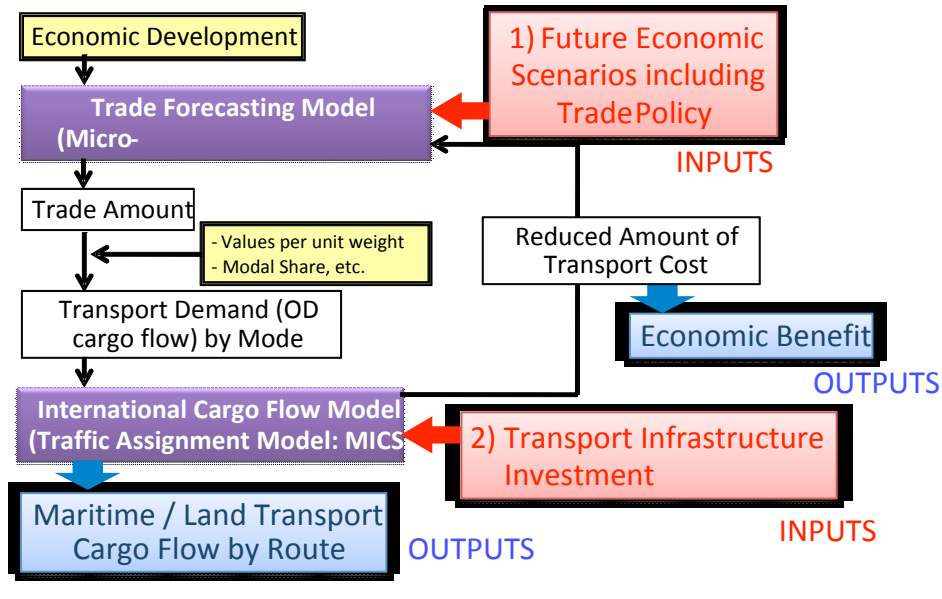


Figure 1. The entire structure of the simulation model.

Trade Forecasting Model. In this simulation, the GTAP model based on the GTAP 6.0 Database (year 2001 data) is used to forecast the trade amounts by five major commodities (agricultural, mining household consumption products, basic industrial materials, and processing and assembling), for the 35 countries and regions listed in Table 2, in five-year steps up to 2020. The input data on future economic development are set based on the results of the questionnaire survey the authors (Shibasaki et al. 2009a) conducted. In addition, not only the data on world's future economics, but also the data on trade policy for each country including conclusion of FTA and other tariff reduction are also set.

International Cargo Flow Simulation Model. The model for international cargo simulation (MICS) was originally developed by NILIM, Japan. The original model focuses on the international maritime container cargo flow between all of the Japanese container ports, the major ports of East Asia, and other areas, as well as on the land transport of container cargo within Japan, when the OD cargo flows (cargo demand) between regions are given. The structure of the model is shown in Figure 2. Shippers and carriers are assumed to independently optimize their different objectives (objective functions), and, as a result, reach a Nash equilibrium under which such objective functions of one party cannot be improved unless the other party changes its behavior. For the detail of the model including overview of both shipper and carrier submodel, convergence calculation, and validation of the model accuracy, please refer to Shibasaki et al. (2009b) due to limitation of space.

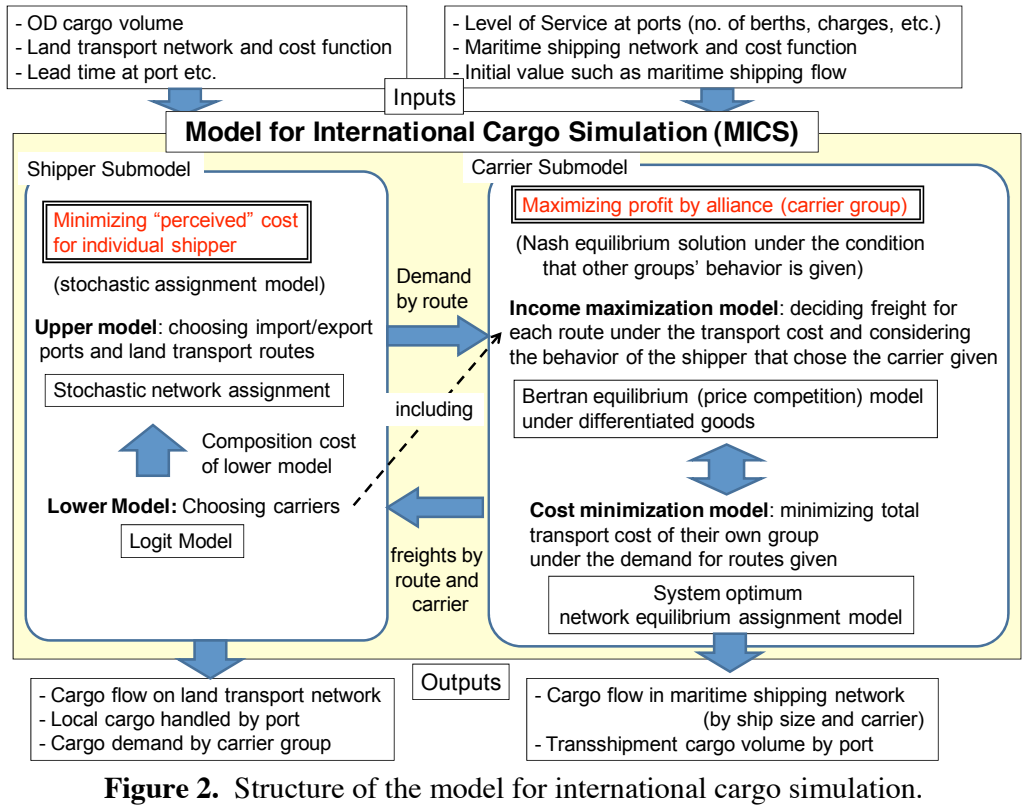


Figure 2. Structure of the model for international cargo simulation.

Estimation Results and Discussion

Estimation of the Future Trade Amount. The input data on future economic development are set, as shown in Table A1 in Appendix. The data set on trade policy including conclusion of FTA and other tariff reduction are also set, as shown in Table A2 in Appendix. When inputting FTA data, all of the tariff rates in the relevant combination of countries are assumed to be zero for all commodities. The estimated results of the trade amount as well as the GDP for each country and region are shown in Table 2. As shown in the table, in China, the predicted growth rate of the GDP until 2020 is the highest among the world as well as that of the trade amount is the second highest next to India. On the other hand, in Japan, the predicted growth rate of the GDP until 2020 is the lowest among the world; however, that of the trade amount is not lower, compared with other countries and regions of the world. That implies the trade in Japan will expand according to the growth of neighbour Asian countries including China, although the expected growth rate of its own economy is low.

Estimation of the Flow of International Cargo Shipping. In order to calculate the flow of international cargo shipping, many kinds of input data need to be prepared including 1) the transport demand of container cargo (OD cargo volume) by region; 2) transport network data such as the physical distance and operational costs of both international maritime shipping and hinterland transport; 3) the service level at each port such as the number of berths by depth, various fares and costs associated with berthing and sizes; and 4) initial input, such as link flow between ports by ship size and by carrier and total volume of containers handled by the ports. The original model was basically developed based on year-2003 data, for which the latest data are available. Most of input data in 2003 are set as well as Shibasaki et al. (2009b). The data set prepared by ports is shown in Table A3 in Appendix. For future simulation as in year 2020, input data are prepared as follows; 1) the transport demand of container cargo are estimated based on the results of the trade forecasting model as mentioned above; 2) transport network data are not changed in principle; 3) the service level at each port are set as shown in the right side of Table A4 in Appendix, extracted from many articles and reports (parameters that are not described in the table are not changed from 2003); and 4) the estimation results in 2003 calculation are used as initial input.

Predicted container cargo throughputs in 2020 for Asian major ports are shown in Figure 3. As shown in the figure, in all ports, the throughputs will increase. Also, predicted increasing rate of international land cargo flow by link from 2003 to 2020 in Northeast Asia are shown in Figure 4. As shown in the figure, in most of land links, the cargo flow will be also predicted to increase.

Table 2. Estimated results of the trade forecasting model.

No	Country / Region	GDP				Trade							
						Export				Import			
		Amount (billion US$)		total growth rate (%)	annual growth rate (%)	Amount (billion US$)		total growth rate (%)	annual growth rate (%)	Amount (billion US$)		total growth rate (%)	annual growth rate (%)
		2001 actual	2020 estimated			2001 actual	2020 estimated			2001 actual	2020 estimated		
1	jap	4,178	5,241	25%	1.2%	401	699	74%	3.0%	299	825	176%	5.5%
2	kor	428	1,179	176%	5.5%	155	565	264%	7.0%	124	469	278%	7.2%
3	prc	1,159	7,323	532%	10.2%	337	2,986	786%	12.2%	226	1,648	629%	11.0%
4	hkg	163	437	169%	5.3%	20	87	326%	7.9%	85	264	210%	6.1%
5	twn	281	682	142%	4.8%	121	323	167%	5.3%	96	271	182%	5.6%
6	xea	71	211	196%	5.9%	4	12	223%	6.4%	6	25	323%	7.9%
7	phl	33	131	301%	7.6%	35	129	274%	7.2%	37	137	267%	7.1%
8	vnm	115	374	226%	6.4%	13	58	347%	8.2%	14	76	434%	9.2%
9	tha	88	292	231%	6.5%	68	282	313%	7.7%	52	247	376%	8.6%
10	mys	85	253	199%	5.9%	101	385	281%	7.3%	60	246	310%	7.7%
11	sgp	145	534	267%	7.1%	87	275	215%	6.2%	102	311	205%	6.1%
12	idn	47	153	228%	6.5%	63	269	330%	8.0%	31	160	410%	9.0%
13	xse	477	2,260	373%	8.5%	7	23	210%	6.1%	5	21	290%	7.4%
14	bgd	16	44	180%	5.6%	7	25	265%	7.0%	9	29	238%	6.6%
15	ind	10,082	16,774	66%	2.7%	47	453	862%	12.7%	46	376	712%	11.7%
16	lka	715	1,289	80%	3.1%	5	17	225%	6.4%	5	18	249%	6.8%
17	xsa	618	1,376	123%	4.3%	11	43	283%	7.3%	12	52	350%	8.2%
18	xme	226	479	112%	4.0%	179	514	187%	5.7%	154	452	194%	5.8%
19	med	54	189	250%	6.8%	480	965	101%	3.7%	556	1,356	144%	4.8%
20	eur	66	177	166%	5.3%	1,834	2,635	44%	1.9%	1,758	2,837	61%	2.6%
21	rus	26	54	110%	4.0%	92	276	198%	5.9%	51	219	332%	8.0%
22	xsu	1,007	2,161	115%	4.1%	39	128	227%	6.4%	37	107	190%	5.8%
23	afr	2,412	5,384	123%	4.3%	90	263	193%	5.8%	80	278	249%	6.8%
24	usa	6,788	9,331	37%	1.7%	651	1,397	115%	4.1%	1,050	2,289	118%	4.2%
25	can	310	959	209%	6.1%	226	379	68%	2.8%	198	368	86%	3.3%
26	mex	322	908	182%	5.6%	149	324	118%	4.2%	127	294	131%	4.5%
27	xna	357	718	101%	3.7%	1	1	122%	4.3%	2	5	115%	4.1%
28	xcm	51	94	85%	3.3%	44	115	161%	5.2%	69	182	162%	5.2%
29	per	14	25	81%	3.2%	7	27	299%	7.6%	6	27	332%	8.0%
30	chl	29	83	183%	5.6%	19	56	200%	5.9%	15	48	220%	6.3%
31	xap	79	196	148%	4.9%	6	14	122%	4.3%	7	17	161%	5.2%
32	sae	84	261	209%	6.1%	123	303	147%	4.9%	107	344	223%	6.4%
33	aus	3	7	118%	4.2%	57	145	154%	5.0%	56	171	207%	6.1%
34	nzl	104	288	176%	5.5%	15	28	92%	3.5%	12	27	133%	4.5%
35	xoc	645	1,778	176%	5.5%	4	7	97%	3.6%	6	12	110%	4.0%
	Total	31,279	61,643	97%	3.6%	5,499	14,208	158%	5.1%	5,499	14,208	158%	5.1%

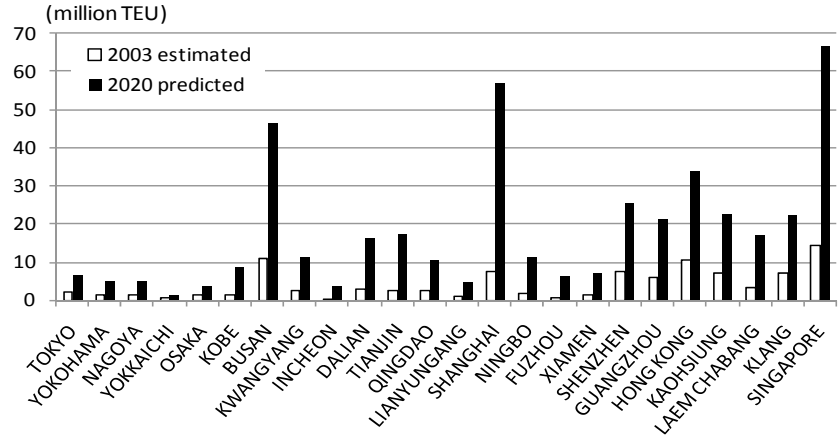


Figure 3. Predicted container cargo throughputs in 2020 by Asian major ports (excluding empty containers).

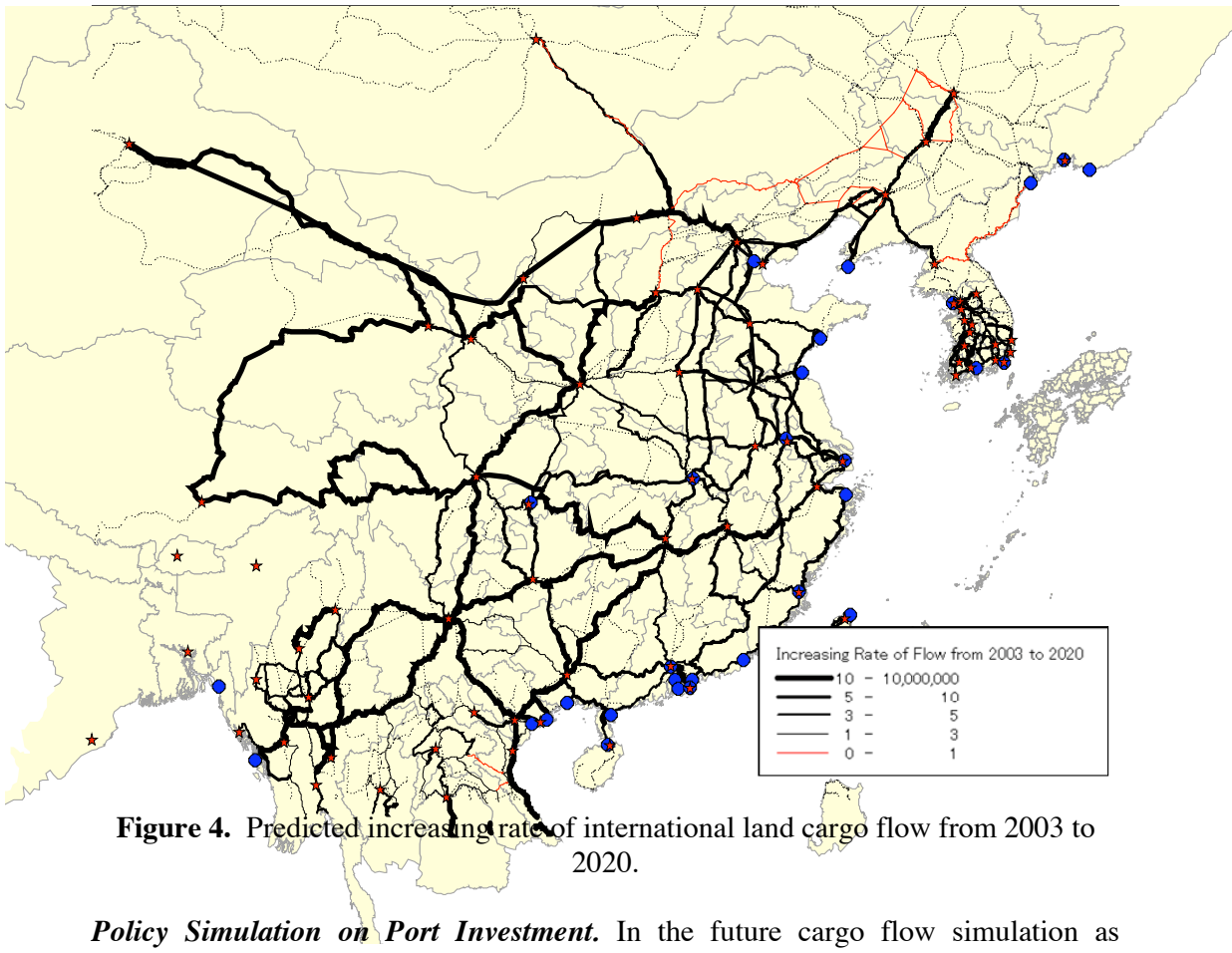


Figure 4. Predicted increasing rate of international land cargo flow from 2003 to 2020.

Policy Simulation on Port Investment. In the future cargo flow simulation as mentioned above, in all countries it is assumed that port investment to construct new berth and/or deepen the existing berths will be implemented as planned as shown in the right side of Table 6. On the other hand, the developed model can simulate under the hypothetical situation that any investments would not be implemented in some specific countries can be calculated. The authors prepare three hypothetical scenarios that in Japanese, Korean, and Chinese ports (except Hong Kong) any investments to container berths since 2008 be not implemented respectively. Settings of the parameters in each port are shown in the left side of Table 6 as described as the data in 2008. An example of the compared result in container cargo throughput by port between the "normal" future prediction (the case in 2020 with investments in all countries' ports) and the hypothetical prediction (the case in 2020 without investments in Chinese ports but with investments in other countries' ports) is shown in Figure 5. As shown in the figure, in most of Chinese ports, in the case assuming no more investments from now, the future throughput will be smaller, compared with the case assuming the complete implementation of the planned investments. Few exceptions are observed in port of Dalian, Qingdao, and Ningbo; in these ports, the container cargo throughput predicted in the case with investments in all ports including China is lower than those in the case without investments in Chinese ports. The reason is considered that, in the case with investments, much larger impacts of

the investment in the neighbor ports are expected and the impacts in these ports are overcome. On the other hand, in most of neighbor countries' ports including Japan and Korea, the throughput will be larger in the case without investments in Chinese ports. The same tendencies in the results are also observed in the simulations on Japanese and Korean port investments.

Meanwhile, the developed model can also calculate the total transport cost of international cargo shipping. Table 3 shows the estimated result on the reduced amount and rate of shipping cost by investment country, described in a longitudinal direction. These results can be calculated by comparing the total cost between the case in 2020 with investments and the case in 2020 without port investments of the country in question. Table 3 also shows the result by country which the cargo are originated from or destined into in a laterally direction. According to the results shown in the table, each country's investment is predicted to contribute in reducing the transport cost, not only of their own cargo, but also of other countries' cargo. For example, Chinese port investments from 2008 to 2020 would reduce the total shipping cost of Chinese cargo by 0.7% (27.3 billion US$); in addition, they can reduce the total cost of Japanese cargo by 0.6% (3.0 billion US$) and that of Korean cargo by 1.0% (5.3 billion US$).

Summarizing the above results, in terms of cargo throughput, each country's port investment has a negative effect to neighbor countries' ports as shown in Figure 5; in other words, they are in the competition. However, in terms of the saved amount of shipping cost, each country's port investment has a positive effect to neighbor countries' cargo as shown in Table 3; in other words, they are in the cooperative relation. This bilateral character is one of the features of international transport. When discussing the effect of international transport policies, it should be always taken into account.

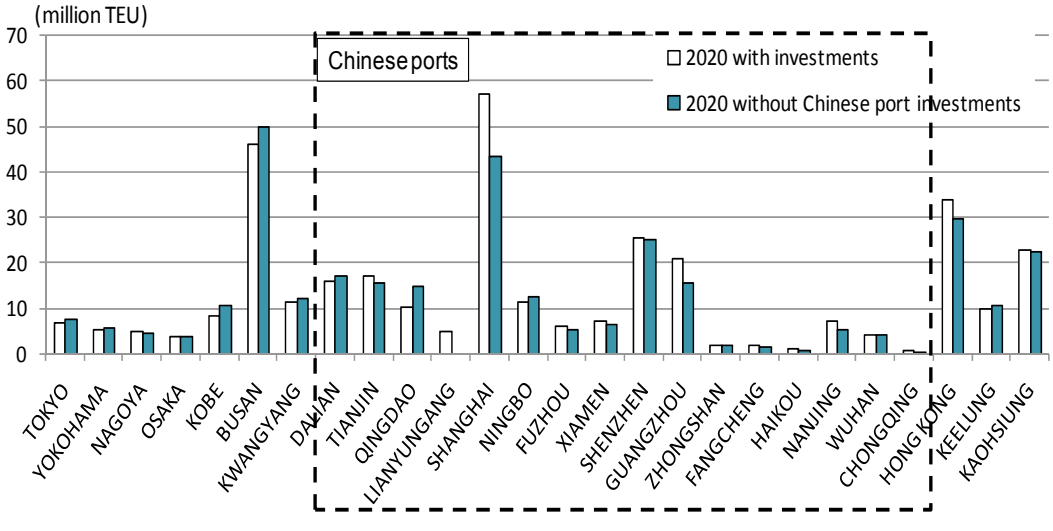


Figure 5. Policy simulation result on Chinese port investments: Comparison of container cargo throughput in Northeast Asian ports.

Table 3. Estimated reduced amount and rate of international cargo shipping cost by investment countries and by countries which cargo originated from or are destined into.

		Cargo from/to			
		Japan	China	Korea	World Total
port investment country	Japan	7.0 (1.5%)	14.8 (0.4%)	4.7 (0.9%)	17.5 (0.2%)
	China	3.0 (0.6%)	27.3 (0.7%)	5.3 (1.0%)	27.8 (0.4%)
	Korea	2.4 (0.5%)	13.2 (0.4%)	7.3 (1.4%)	22.8 (0.3%)

Conclusions

This paper attempted to forecast the future trade amount and international cargo flow, especially focused in China, using by the stepwise trade amount and international logistics flow forecasting system in East Asia that the author had been developed. In addition, by comparing the calculation result between the case with consideration and without any consideration of investments, the effect of these investments was estimated in terms of, not only the cargo throughput by port, but also the reduced cost of international shipping. As a result, although in terms of cargo throughput, each country's port is in the competition each other, in terms of the saved amount of shipping cost, each country's port is in the cooperative relation.

The authors will continue the effort to improve the model reliability, in order to contribute to the productive discussion for mutual development and cooperative investment on international transport infrastructure in the East Asian region.

Acknowledgements

This research was partly supported by Grant-in-Aid for Scientific Research (C) No. 21560565.

References

Economic Research Institute for ASEAN, and East Asia (2008). "Predicting Long-Term Effects of International Development Projects in Continental South East Asia: IDE Geographical Simulation Model." ERIA Discussion Paper Series, ERIA-DP-2008-02.

Friesz, T. L., Gottfried, J. A., and Morlok, E. K. (1986). "A sequential shipper-carrier network model for predicting freight flows." *Transportation Science*, 20(2), 80-91.

Hertel, T. W. (1997). "Global Trade Analysis Project." *Cambridge University Press*.

Kunita, A., Mizutani, M., Tsuchiya, K., Akiyoshi, S., Koike, A., Ishikawa, Y., and Ishiguro, K. (2005). "International Freight Transport Demand Forecasting by SCGE Model." *Proceedings of the 19th Pacific Regional Science Conference*.

Malchow, M., and Kanafani, A. (2001). "A disaggregate analysis of factors influencing port selection." *Maritime Policy and Management*, 28(3), 265-277.

Shibasaki, R., Ma, L., Kadono, T., Ishikura, T., and Ieda, H. (2005a). "An Estimation of International Maritime Container Volume Among Asian Countries by GTAP Model and Simulation on FTA and Transport Improvement Scenario." *8th Annual Conference on Global Economic Analysis*.

Shibasaki, R., Ieda, H., and Kadono, T. (2005b). "Model Improvement of International Maritime Container Cargo Flow and Policy Evaluation for International Logistics in Eastern Asia." *The 1st International Conference on Transportation Logistics*.

Shibasaki, R., Kannami, Y., Onodera, H., Li, J., Miao, L., and Watanabe, T. (2007). "Impact of Chinese Port Policy Using the Model for International Container Cargo Simulation." *Journal of the Eastern Asia Society for Transportation Studies*, 7, 1083-1098.

Shibasaki, R., Yonemoto, K., and Watanabe, T. (2008). "On the effects of trade liberalization policies on regional economies based on Transnational Interregional Input-Output Table between China and Japan." *11th Annual Conference on Global Economic Analysis*.

Shibasaki, R., Ishikura, T., Abe, M., Watanabe, T., Yamane T., Yoshida, T., Abe, K., Nemoto, T., Hanaoka, S., and Ono, K. (2009a). "Future Scenario of International Economics and Transport Based on Questionnaire Survey by the Delphi Method." *Technical Note of National Institute for Land and Infrastructure Management*, 495.

Shibasaki, R., Watanabe, T., and Araki, D. (2009b). "How Model Accuracy is Improved by Usage of Statistics? - An Example of International Freight Simulation Model in East Asia." *Proceedings of the Eastern Asia Society for Transportation Studies*,7.

Shintani, K., Imai, A., Nishimura, E., and Papadimitriou, S. (2007). "The container shipping network design problem with empty container repositioning." *Transportation Research Part E 43*, 39-59.

Takebayashi, M., Kuroda, K., Fujita, T., and Kanai, H. (2004). "Equilibrium Approach on International Maritime Transport Market concerning Global Carrier's Behabior on Trunk Lines." *Proceeding of World Conference on Transport Research*, Istanbul, Turkey.

Appendix

Table A1. Input data on future economic development into the trade forecasting model (%)

No.	Country / Region	population				unskilled labor			
		2001→2005	2005→2010	2010→2015	2015→2020	2001→2005	2005→2010	2010→2015	2015→2020
1	Japan	0.54	▲0.11	▲0.90	▲1.67	▲1.75	▲3.45	▲5.45	▲4.01
2	Korea	1.85	1.68	0.91	0.21	1.74	3.57	1.67	▲1.77
3	China	2.69	2.93	2.74	2.35	6.37	4.94	2.19	▲0.54
4	Hong Kong	4.69	5.13	4.38	3.82	3.99	5.79	3.04	▲1.43
5	Taiwan	1.70	1.91	1.54	1.04	1.70	1.91	1.54	1.04
6	Rest of East Asia	4.25	9.30	9.44	3.85	6.26	6.36	4.49	2.96
7	Philippines	8.58	9.97	8.70	7.58	10.51	12.19	10.76	8.84
8	Vietnam	5.91	6.84	6.19	5.38	10.50	12.02	7.44	4.54
9	Thailand	3.06	3.37	2.52	1.84	6.30	4.30	1.85	0.12
10	Malaysia	8.01	8.84	7.62	6.57	10.16	11.60	8.79	7.20
11	Singapore	6.08	6.12	4.71	3.24	5.92	16.96	3.70	▲1.78
12	Indonesia	5.36	5.99	4.99	4.09	7.01	7.62	6.63	5.49
13	Rest of Southeast Asia	4.25	9.30	9.44	3.85	6.26	6.36	4.49	2.96
14	Bangladesh	7.79	8.71	8.09	7.34	10.70	11.59	9.94	7.88
15	India	6.63	7.56	6.75	5.89	8.85	10.47	8.94	7.11
16	Sri Lanka	1.73	2.38	1.96	1.35	5.13	3.28	0.22	1.01
17	Rest of South Asia	4.25	9.30	9.44	3.85	6.26	6.36	4.49	2.96
18	Rest of Middle East	4.25	9.30	9.44	3.85	6.26	6.36	4.49	2.96
19	Mediterranean	3.27	3.23	2.49	1.90	4.28	4.59	2.78	1.72
20	Europe	1.09	0.94	0.67	0.50	1.44	1.09	▲0.97	▲1.56
21	Russian Federation	▲1.89	▲2.53	▲2.74	▲2.98	▲0.01	▲0.58	▲4.33	▲5.11
22	Rest of Former Soviet Union	4.25	9.30	9.44	3.85	6.26	6.36	4.49	2.96
23	Africa	8.70	10.28	9.94	9.61	9.64	11.95	12.25	12.62
24	United States	4.17	4.95	4.55	4.11	4.98	4.98	3.15	2.16
25	Canada	4.08	4.59	4.26	3.97	5.34	5.34	2.65	1.22
26	Mexico	3.60	5.78	4.95	4.15	7.11	8.00	6.71	4.64
27	Rest of North America	4.25	9.30	9.44	3.85	6.26	6.36	4.49	2.96
28	Central America	4.25	9.30	9.44	3.85	6.26	6.36	4.49	2.96
29	Peru	4.96	5.94	6.49	5.76	7.71	8.09	7.61	6.71
30	Chile	4.53	5.15	4.62	3.98	7.13	7.15	4.39	2.35
31	South America West Ciast	4.25	9.30	9.44	3.85	6.26	6.36	4.49	2.96
32	South America East Ciast	5.66	6.47	5.65	4.88	7.33	7.46	6.24	4.80
33	Australia	4.84	5.18	4.84	4.56	5.73	5.30	3.05	2.40
34	New Zealand	4.98	4.58	4.01	3.57	6.08	5.39	2.91	2.22
35	Rest of Oceania	4.25	9.30	9.44	3.85	6.26	6.36	4.49	2.96

No.	skilled labor				capital				natural resrouce			
	2001→2005	2005→2010	2010→2015	2015→2020	2001→2005	2005→2010	2010→2015	2015→2020	2001→2005	2005→2010	2010→2015	2015→2020
1	▲1.75	▲3.45	▲5.45	▲4.01	13.94	2.19	1.31	▲0.40	15.29	11.81	12.23	11.34
2	1.74	3.57	1.67	▲1.77	20.35	18.97	17.13	15.46	15.29	11.81	12.23	11.34
3	6.37	5.68	2.92	0.11	45.61	34.06	34.49	30.83	15.29	11.81	12.23	11.34
4	3.99	5.79	3.04	▲1.43	22.70	16.13	18.33	16.54	15.29	11.81	12.23	11.34
5	1.70	1.91	1.54	1.04	19.75	18.30	17.90	15.33	15.29	11.81	12.23	11.34
6	6.26	6.98	5.15	3.59	12.35	28.51	18.40	18.08	15.29	11.81	12.23	11.34
7	10.51	12.80	11.42	9.47	22.16	20.01	23.73	23.16	15.29	11.81	12.23	11.34
8	10.50	12.64	8.10	5.17	34.34	21.79	26.09	25.77	15.29	11.81	12.23	11.34
9	6.30	4.92	2.51	0.75	25.30	20.26	23.23	22.79	15.29	11.81	12.23	11.34
10	10.16	12.22	9.44	7.83	24.12	24.15	24.58	24.06	15.29	11.81	12.23	11.34
11	5.92	16.96	3.70	▲1.78	24.59	18.47	19.19	17.67	15.29	11.81	12.23	11.34
12	7.01	8.24	7.29	6.12	21.54	27.92	29.92	29.18	15.29	11.81	12.23	11.34
13	6.26	6.98	5.15	3.59	12.35	17.30	20.86	20.37	15.29	11.81	12.23	11.34
14	10.70	12.21	10.60	8.51	23.75	21.14	28.00	27.39	15.29	11.81	12.23	11.34
15	8.85	11.09	9.60	7.74	33.04	30.73	32.31	31.70	15.29	11.81	12.23	11.34
16	5.13	3.90	0.87	1.64	23.24	20.93	19.99	19.56	15.29	11.81	12.23	11.34
17	6.26	6.98	5.15	3.59	12.35	29.08	26.15	25.51	15.29	11.81	12.23	11.34
18	6.26	6.98	5.15	3.59	12.35	22.88	20.88	18.56	15.29	11.81	12.23	11.34
19	4.28	4.59	2.78	1.72	6.20	13.91	18.54	16.67	15.29	11.81	12.23	11.34
20	1.44	1.09	▲0.97	▲1.56	6.65	8.11	7.83	6.05	15.29	11.81	12.23	11.34
21	▲0.01	▲0.58	▲4.33	▲5.11	28.15	19.78	19.52	17.87	15.29	11.81	12.23	11.34
22	6.26	6.98	5.15	3.59	12.35	23.89	23.09	20.15	15.29	11.81	12.23	11.34
23	9.64	12.57	12.91	13.25	17.44	22.99	20.67	18.44	15.29	11.81	12.23	11.34
24	4.98	4.98	3.15	2.16	12.27	10.95	11.47	9.31	15.29	11.81	12.23	11.34
25	5.34	5.34	2.65	1.22	11.34	12.75	12.43	10.44	15.29	11.81	12.23	11.34
26	7.11	8.62	7.37	5.27	9.44	15.86	17.67	15.43	15.29	11.81	12.23	11.34
27	6.26	6.98	5.15	3.59	12.35	11.90	16.78	15.24	15.29	11.81	12.23	11.34
28	6.26	6.98	5.15	3.59	12.35	13.56	16.28	14.35	15.29	11.81	12.23	11.34
29	7.71	8.71	8.27	7.34	22.38	25.71	25.29	22.60	15.29	11.81	12.23	11.34
30	7.13	7.77	5.05	2.99	18.97	19.57	20.73	17.99	15.29	11.81	12.23	11.34
31	6.26	6.98	5.15	3.59	12.35	13.93	16.64	14.57	15.29	11.81	12.23	11.34
32	7.33	8.08	6.89	5.43	13.39	14.46	17.15	14.88	15.29	11.81	12.23	11.34
33	5.73	5.30	3.05	2.40	12.93	13.31	14.50	12.59	15.29	11.81	12.23	11.34
34	6.08	5.39	2.91	2.22	14.52	11.54	12.96	10.87	15.29	11.81	12.23	11.34
35	6.26	6.98	5.15	3.59	12.35	13.56	10.96	9.33	15.29	11.81	12.23	11.34

No.	manufacturing technology				tariff rate			
	2001→2005	2005→2010	2010→2015	2015→2020	2001→2005	2005→2010	2010→2015	2015→2020
1	2.79	5.97	9.58	7.86	-3.07	-1.50	-0.12	-0.50
2	6.96	15.49	19.75	22.94	-3.07	-1.50	-0.12	-0.50
3	20.48	36.44	45.56	49.01	-3.07	-1.50	-0.12	-0.50
4	7.83	10.77	19.50	23.08	-3.07	-1.50	-0.12	-0.50
5	10.03	13.54	18.54	19.80	-3.07	-1.50	-0.12	-0.50
6	4.04	29.47	19.96	21.59	-3.07	-1.50	-0.12	-0.50
7	4.05	11.30	17.95	19.54	-3.07	-1.50	-0.12	-0.50
8	7.41	17.58	27.87	30.11	-3.07	-1.50	-0.12	-0.50
9	5.21	15.39	21.91	23.23	-3.07	-1.50	-0.12	-0.50
10	5.43	16.08	20.32	21.88	-3.07	-1.50	-0.12	-0.50
11	8.05	9.58	21.75	26.33	-3.07	-1.50	-0.12	-0.50
12	5.56	18.65	23.86	25.17	-3.07	-1.50	-0.12	-0.50
13	2.31	11.01	17.65	18.94	-3.07	-1.50	-0.12	-0.50
14	7.52	13.02	25.84	28.48	-3.07	-1.50	-0.12	-0.50
15	12.59	24.80	31.52	33.81	-3.07	-1.50	-0.12	-0.50
16	10.75	19.39	22.63	22.85	-3.07	-1.50	-0.12	-0.50
17	3.49	24.72	25.13	26.92	-3.07	-1.50	-0.12	-0.50
18	2.04	17.44	19.10	20.42	-3.07	-1.50	-0.12	-0.50
19	0.82	8.95	20.52	22.23	-3.07	-1.50	-0.12	-0.50
20	2.81	3.79	6.49	7.17	-3.07	-1.50	-0.12	-0.50
21	14.45	18.96	28.46	30.53	-3.07	-1.50	-0.12	-0.50
22	3.65	17.47	25.03	28.21	-3.07	-1.50	-0.12	-0.50
23	3.97	16.12	15.99	16.44	-3.07	-1.50	-0.12	-0.50
24	4.25	4.26	8.49	9.69	-3.07	-1.50	-0.12	-0.50
25	3.27	6.77	11.04	12.66	-3.07	-1.50	-0.12	-0.50
26	0.31	8.85	13.27	14.40	-3.07	-1.50	-0.12	-0.50
27	3.15	7.75	17.80	19.13	-3.07	-1.50	-0.12	-0.50
28	2.91	8.35	15.14	16.53	-3.07	-1.50	-0.12	-0.50
29	4.66	18.26	21.49	22.56	-3.07	-1.50	-0.12	-0.50
30	4.82	12.47	17.87	19.40	-3.07	-1.50	-0.12	-0.50
31	3.65	8.43	15.65	17.18	-3.07	-1.50	-0.12	-0.50
32	3.14	7.76	14.13	15.45	-3.07	-1.50	-0.12	-0.50
33	3.94	7.95	14.82	16.08	-3.07	-1.50	-0.12	-0.50
34	3.97	5.02	10.94	11.93	-3.07	-1.50	-0.12	-0.50
35	2.81	8.97	7.96	9.05	-3.07	-1.50	-0.12	-0.50

Table A2. Input data on future trade policy to the trade forecasting model

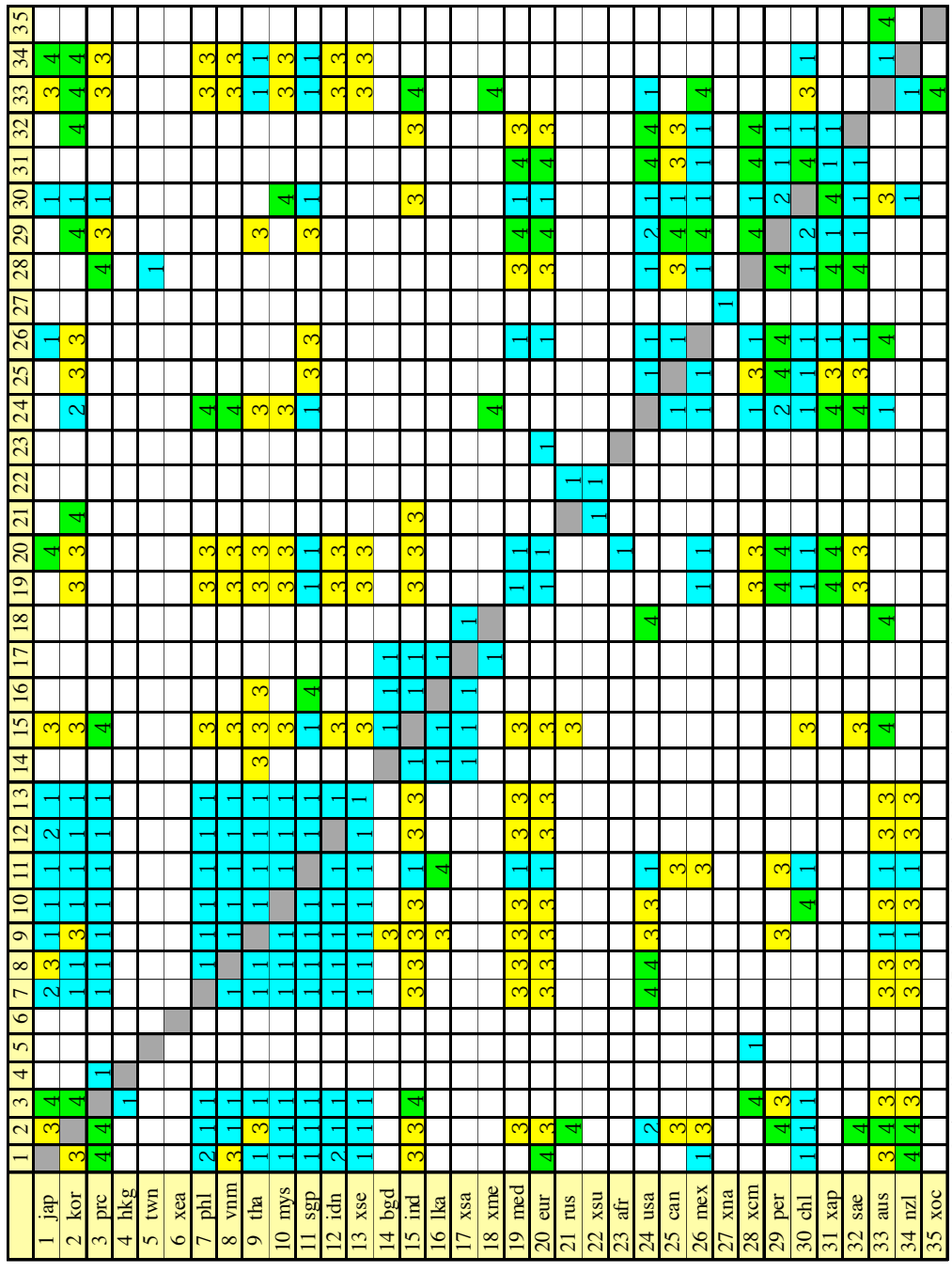


Table A3. Input data on ports into the international cargo flow simulation model

No.	Port Name		Handling Charge (US$/TEU)*	Port Charge* (thousand US$/vessel)				Terminal Charge (million US$/year/berth)*	Handling Time** (hours)			Number of berth***			
				-1000 TEU	1000-2500 TEU	2500-4000 TEU	4000 TEU-		Export Cargo	Import Cargo	Transshipped Cargo	>-11m	-11--13m	-13--14m	<-14m
1	Japan	Tomakomai	150	17.6	20.5	24.9	33.7	10.8	72	72	48	0	0	0	1
2	Japan	Sendai-Shiogama	150	17.6	20.5	24.9	33.7	10.8	72	72	48	0	1	0	1
3	Japan	Tokyo	150	17.6	20.5	24.9	33.7	10.8	72	72	48	3	2	0	10
4	Japan	Yokohama	150	17.6	20.5	24.9	33.7	10.8	72	72	48	0	7	10	6
5	Japan	Niigata	150	17.6	20.5	24.9	33.7	10.8	72	72	48	4	1	0	0
6	Japan	Shimizu	150	17.6	20.5	24.9	33.7	10.8	72	72	48	0	5	0	1
7	Japan	Nagoya	150	17.6	20.5	24.9	33.7	10.8	72	72	48	3	6	0	2
8	Japan	Yokkaichi	150	17.6	20.5	24.9	33.7	10.8	72	72	48	0	2	0	0
9	Japan	Osaka	150	17.6	20.5	24.9	33.7	10.8	72	72	48	3	4	5	3
10	Japan	Kobe	150	17.6	20.5	24.9	33.7	10.8	72	72	48	5	8	5	10
11	Japan	Mizushima	150	17.6	20.5	24.9	33.7	10.8	72	72	48	5	0	0	0
12	Japan	Hiroshima	150	17.6	20.5	24.9	33.7	10.8	72	72	48	7	0	0	1
13	Japan	Tokuyama-Kudamatsu	150	17.6	20.5	24.9	33.7	10.8	72	72	48	0	1	0	0
14	Japan	Hakata	150	17.6	20.5	24.9	33.7	10.8	72	72	48	0	1	2	1
15	Japan	Kita-Kyushu	150	17.6	20.5	24.9	33.7	10.8	72	72	48	3	4	0	0
16	Japan	Shibushi	150	17.6	20.5	24.9	33.7	10.8	72	72	48	4	0	0	0
17	Japan	Naha	150	17.6	20.5	24.9	33.7	10.8	72	72	48	0	0	1	0
18	South Korea	Busan	220	6.5	9.4	13.8	22.6	6.9	48	24	24	0	7	2	11
19	South Korea	Kwangyang	220	6.5	9.4	13.8	22.6	6.9	48	24	24	0	0	0	6
20	South Korea	Incheon	220	6.5	9.4	13.8	22.6	6.9	72	48	60	0	3	0	2
21	North Korea	Najin	590	6.5	9.4	13.8	22.5	5.0	312	432	276	1	0	0	0
22	Far East Russia	Vostochniy	260	7.5	10.4	14.8	23.5	5.0	72	48	60	0	2	2	0
23	Far East Russia	Vladivostok	260	7.5	10.4	14.8	23.5	5.0	72	48	60	0	2	0	0
24	China	Dalian	80	6.8	9.7	14.1	22.8	5.0	72	72	168	0	3	4	0
25	China	Tianjin	80	6.8	9.7	14.1	22.8	5.0	72	72	168	0	1	4	3
26	China	Qingdao	80	6.8	9.7	14.1	22.8	5.0	72	72	168	0	2	0	6
27	China	Lianyungang	80	6.8	9.7	14.1	22.8	5.0	72	72	168	0	2	0	0
28	China	Shanghai	80	6.8	9.7	14.1	22.8	5.0	72	72	168	7	6	5	4
29	China	Ningbo	80	6.8	9.7	14.1	22.8	5.0	72	72	168	0	0	3	3
30	China	Fuzhou	80	6.8	9.7	14.1	22.8	5.0	72	72	168	2	3	0	1
31	China	Xiamen	80	6.8	9.7	14.1	22.8	5.0	72	72	168	0	2	3	0
32	China	Shenzhen	80	6.8	9.7	14.1	22.8	5.0	72	72	168	0	2	0	10
33	China	Guangzhou	80	6.8	9.7	14.1	22.8	5.0	72	72	168	0	6	0	0
34	China	Zhongshan	80	6.8	9.7	14.1	22.8	5.0	72	72	168	4	0	0	0
35	China	Fangcheng	80	6.8	9.7	14.1	22.8	5.0	72	72	168	0	2	0	0
36	China	Haikou	80	6.8	9.7	14.1	22.8	5.0	72	72	168	2	0	0	0
37	China	Nanjing	80	6.8	9.7	14.1	22.8	5.0	72	72	168	0	2	0	0
38	China	Wuhan	80	6.8	9.7	14.1	22.8	5.0	72	72	168	6	0	0	0
39	China	Chongqing	80	6.8	9.7	14.1	22.8	5.0	72	72	168	1	0	0	0
40	Hong Kong	Hong Kong	250	15.8	18.7	23.1	31.8	45.3	24	24	12	0	0	0	23
41	Chinese Taipei	Keelung	180	12.4	15.3	19.7	28.4	8.7	48	48	48	0	14	0	0
42	Chinese Taipei	Taichung	180	12.4	15.3	19.7	28.4	8.7	48	48	48	0	0	0	6
43	Chinese Taipei	Kaohsiung	180	12.4	15.3	19.7	28.4	8.7	48	48	48	4	5	0	10
44	Philippines	Subic Bay	230	6.5	9.5	13.9	22.6	5.0	72	96	84	0	3	0	0
45	Philippines	Manila	230	6.5	9.5	13.9	22.6	5.0	72	96	84	19	7	2	1
46	Philippines	Batangas	230	6.5	9.5	13.9	22.6	5.0	72	96	84	2	0	0	0
47	Philippines	Cebu	230	6.5	9.5	13.9	22.6	5.0	72	96	84	4	0	0	0
48	Philippines	Davao	230	6.5	9.5	13.9	22.6	5.0	72	96	84	4	0	0	0
49	Vietnam	Cai Lan	-	-	-	-	-	-	-	-	-	-	-	-	-
50	Vietnam	Haiphong	450	6.5	9.4	13.8	22.5	1.5	72	96	84	3	0	0	0
51	Vietnam	Da Nang	450	6.5	9.4	13.8	22.5	1.5	72	96	84	5	4	0	0
52	Vietnam	Van Phong	-	-	-	-	-	-	-	-	-	-	-	-	-
53	Vietnam	Ho Chi Minh City	450	6.5	9.4	13.8	22.5	1.5	72	96	84	10	6	2	0
54	Cambodia	Phnom Penh	150	8.7	13.0	13.8	22.5	2.5	72	120	96	3	0	0	0
55	Cambodia	Sihanoukville	150	5.8	9.4	13.8	22.5	2.5	72	120	96	1	0	0	0
56	Thailand	Laem Chabang	160	7.3	10.2	14.6	23.4	1.6	72	48	60	0	0	0	7
57	Thailand	Bangkok	160	7.3	10.2	14.6	23.4	1.6	72	48	60	21	0	0	0
58	Thailand	Songkhla	160	7.3	10.2	14.6	23.4	1.6	72	48	60	3	0	0	0
59	Malaysia	Kuantan	140	9.7	12.6	17.0	25.7	5.0	72	48	60	0	3	0	0
60	Malaysia	Pasir Gudang	140	9.7	12.6	17.0	25.7	5.0	72	48	60	0	0	0	3

No.	Port Name		Handling Charge (US$/TEU)*	Port Charge* (thousand US$/vessel)				Terminal Charge (million US$/year/berth)*	Handling Time** (hours)			Number of berth***			
				-1000 TEU	1000-2500 TEU	2500-4000 TEU	4000 TEU-		Export Cargo	Import Cargo	Trans-shipped Cargo	>-11m	-11--13m	-13--14m	<-14m
61	Malaysia	Tanjung Pelepas	140	9.7	12.6	17.0	25.7	5.0	72	48	48	0	0	0	6
62	Malaysia	Port Klang	140	9.7	12.6	17.0	25.7	5.0	72	48	60	0	0	0	19
63	Malaysia	Penang	140	9.7	12.6	17.0	25.7	5.0	72	48	60	2	3	0	0
64	Malaysia	Kuching	140	9.7	12.6	17.0	25.7	5.0	72	48	60	6	6	0	0
65	Malaysia	Kota Kinabalu	140	9.7	12.6	17.0	25.7	5.0	72	48	60	3	0	0	0
66	Singapore	Singapore	200	10.1	13.0	17.4	26.2	6.2	24	24	12	6	15	4	18
67	Myanmar	Thilawa	450	6.5	9.4	13.8	22.5	5.0	72	120	96	2	3	0	0
68	Myanmar	Kyaukphyu	-	-	-	-	-	-	-	-	-	-	-	-	-
69	Indonesia	Surabaya (Tanjung Perak)	160	7.7	10.7	15.0	23.8	5.0	48	144	96	11	0	0	0
70	Indonesia	Semarang	160	7.7	10.7	15.0	23.8	5.0	48	144	96	2	0	0	0
71	Indonesia	Jakarta (Tanjung Priok)	160	7.7	10.7	15.0	23.8	5.0	48	144	96	2	4	0	6
72	Indonesia	Bojonegara	-	-	-	-	-	-	-	-	-	-	-	-	-
73	Indonesia	Belawan	160	7.7	10.7	15.0	23.8	5.0	48	144	96	5	0	0	0
74	Indonesia	Pontianak	160	7.7	10.7	15.0	23.8	5.0	48	144	96	4	0	0	0
75	Indonesia	Balikpapan	160	7.7	10.7	15.0	23.8	5.0	48	144	96	0	2	0	0
76	Indonesia	Ujung Pandang	160	7.7	10.7	15.0	23.8	5.0	48	144	96	0	2	0	0
77	Brunei Darussalam	Muara	280	6.5	9.4	13.8	22.5	5.0	48	24	36	0	2	0	0
78	Bangladesh	Chittagong	590	6.5	9.4	13.8	22.5	5.0	120	96	108	2	0	0	0
79	India	Jawaharlal Nehr	220	7.5	10.4	14.8	23.5	3.5	72	144	108	0	5	0	0
80	Sri Lanka	Colombo	150	7.5	10.4	14.8	23.5	5.0	72	72	24	4	1	2	4
81	Pakistan	Karachi	150	7.5	10.4	14.8	23.5	5.0	96	72	84	5	3	0	0
82	Russia (European Area)		130	13.1	16.0	20.4	29.2	5.0	72	48	60	0	0	0	100#
83	PNW		460	16.5	19.4	23.8	32.6	5.0	48	24	36	0	0	0	100#
84	PSW		460	16.5	19.4	23.8	32.6	5.0	48	24	36	0	0	0	100#
85	East Coast of North America		460	16.5	19.4	23.8	32.6	5.0	48	24	36	0	0	0	100#
86	West Coast of South America		20	7.5	10.4	14.8	23.5	5.0	96	96	96	0	0	0	100#
87	East Coast of South America		140	7.5	10.4	14.8	23.5	5.0	72	96	84	0	0	0	100#
88	Middle East		110	7.5	10.4	14.8	23.5	5.0	48	24	36	0	0	0	100#
89	Mediterranean		350	7.5	10.4	14.8	23.5	5.0	72	72	72	0	0	0	100#
90	Europe		270	13.1	16.0	20.4	29.2	5.0	24	24	24	0	0	0	100#
91	Africa		290	13.1	16.0	20.4	29.2	5.0	216	336	276	0	0	0	100#
92	Oceania		310	6.5	9.4	13.8	22.5	5.0	24	72	48	0	0	0	100#

Source: * authors' estimation. ** World Bank: Doing Business. *** Containerisation International Yearbook etc.
Notice: # hypothetical assumption that there are no port congestions at all in non-Asian ports.

Table A4. Data set on future conditions of each port into the model calculation

No.	Port Name		2008							2020						
			Handling Time (hours)			Number of berths				Handling Time (hours)			Number of berths			
			Export	Import	Trans-shipped	>-11m	-11--13m	-13--14m	<-14m	Export	Import	Trans-shipped	>-11m	-11--13m	-13--14m	<-14m
1	Japan	Tomakomai	72	72	48	0	0	0	1	72	72	48	0	0	0	2
2	Japan	Sendai-Shiogama	72	72	48	0	1	0	1	72	72	48	0	1	0	1
3	Japan	Tokyo	72	72	48	2	2	0	10	48	48	24	2	3	0	13
4	Japan	Yokohama	72	72	48	0	4	9	7	48	48	24	0	4	9	9
5	Japan	Niigata	72	72	48	1	1	0	0	72	72	48	1	0	0	1
6	Japan	Shimizu	72	72	48	0	3	0	1	72	72	48	0	2	0	2
7	Japan	Nagoya	72	72	48	3	5	0	6	48	48	24	3	5	0	6
8	Japan	Yokkaichi	72	72	48	0	0	1	1	48	48	24	0	0	1	1
9	Japan	Osaka	72	72	48	2	4	5	3	48	48	24	1	4	5	4
10	Japan	Kobe	72	72	48	2	4	2	10	48	48	24	2	4	2	10
11	Japan	Mizushima	72	72	48	4	0	0	0	72	72	48	4	1	0	0
12	Japan	Hiroshima	72	72	48	3	0	0	1	72	72	48	3	0	0	2
13	Japan	Tokuyama-Kudamatsu	72	72	48	0	1	0	0	72	72	48	0	2	0	0
14	Japan	Hakata	72	72	48	0	1	2	2	72	72	48	0	1	2	2
15	Japan	Kita-Kyushu	72	72	48	3	2	0	2	72	72	48	3	2	0	2
16	Japan	Shibushi	72	72	48	2	0	0	0	72	72	48	2	0	0	2
17	Japan	Naha	72	72	48	0	0	2	0	72	72	48	0	0	2	0
18	South Korea	Busan	48	24	24	0	2	2	23	24	24	24	0	2	2	46
19	South Korea	Kwangyang	48	24	24	0	2	0	14	24	24	24	0	2	0	14
20	South Korea	Incheon	72	48	60	0	3	0	3	72	48	60	0	4	0	8

No.	Port Name		2008							2020						
			Handling Time (hours)			Number of berths				Handling Time (hours)			Number of berths			
			Export	Import	Trans-shipped	> -11m	-11- -13m	-13- -14m	< -14m	Export	Import	Trans-shipped	> -11m	-11- -13m	-13- -14m	< -14m
21	North Korea	Najin								312	432	276	1	0	0	0
22	Far East Russia	Vostochniy								72	48	60	0	2	10	0
23	Far East Russia	Vladivostok								72	48	60	0	2	10	0
24	China	Dalian	72	72	168	0	2	3	12	72	72	72	0	2	3	17
25	China	Tianjin	72	72	168	0	1	0	15	72	72	72	0	1	0	38
26	China	Qingdao	72	72	168	0	0	1	11	72	72	72	0	0	1	26
27	China	Lianyungang	72	72	168	0	2	0	0	72	72	168	0	2	0	5
28	China	Shanghai	72	72	168	9	12	6	13	72	72	72	9	12	11	20
29	China	Ningbo	72	72	168	0	0	6	8	72	72	72	0	0	6	29
30	China	Fuzhou	72	72	168	2	3	0	3	72	72	168	2	3	0	5
31	China	Xiamen	72	72	168	0	8	7	3	72	72	72	0	8	7	3
32	China	Shenzhen	72	72	168	0	2	0	30	72	72	72	0	2	0	52
33	China	Guangzhou	72	72	168	0	9	0	10	72	72	72	0	9	0	16
34	China	Zhongshan	72	72	168	4	0	0	0	72	72	168	5	2	0	0
35	China	Fangcheng	72	72	168	0	2	0	0	72	72	168	0	2	0	0
36	China	Haikou	72	72	168	2	0	0	0	72	72	168	2	2	0	0
37	China	Nanjing	72	72	168	0	5	0	0	72	72	168	0	11	0	0
38	China	Wuhan	72	72	168	8	0	0	0	72	72	168	12	0	0	0
39	China	Chongqing	72	72	168	1	0	0	0	72	72	168	9	0	0	0
40	Hong Kong	Hong Kong								24	24	12	0	4	0	20
41	Chinese Taipei	Keelung								48	48	48	2	6	5	8
42	Chinese Taipei	Taichung								48	48	48	0	0	0	6
43	Chinese Taipei	Kaohsiung								48	48	48	4	4	1	22
44	Philippines	Subic Bay								48	48	48	0	3	2	0
45	Philippines	Manila								48	48	48	29	7	2	1
46	Philippines	Batangas								48	48	48	2	2	0	0
47	Philippines	Cebu								72	96	84	4	0	0	0
48	Philippines	Davao								72	96	84	4	0	0	0
49	Vietnam	CaiLan								48	48	48	7	0	0	0
50	Vietnam	Haiphong								72	96	84	3	0	0	2
51	Vietnam	Da Nang								48	48	48	5	4	2	0
52	Vietnam	Van Phong								48	48	48	0	0	0	2
53	Vietnam	Ho Chi Minh City								48	48	48	10	6	2	8
54	Cambodia	Phnom Penh								48	48	48	3	0	0	0
55	Cambodia	Sihanoukville								72	120	96	1	0	0	0
56	Thailand	Laem Chabang								48	48	48	0	0	0	11
57	Thailand	Bangkok								72	48	60	21	0	0	0
58	Thailand	Songkhla								72	48	60	3	0	0	0
59	Malaysia	Kuantan								72	48	60	0	3	0	0
60	Malaysia	Pasir Gudang								72	48	60	0	0	0	3
61	Malaysia	Tanjung Pelepas								72	48	48	0	0	0	6
62	Malaysia	Port Klang								72	48	60	0	0	0	20
63	Malaysia	Penang								72	48	60	2	3	0	0
64	Malaysia	Kuching								72	48	60	6	6	0	0
65	Malaysia	Kota Kinabalu								72	48	60	3	0	0	0
66	Singapore	Singapore								24	24	12	6	15	4	37
67	Myanmar	Thilawa								72	120	96	2	3	0	0
68	Myanmar	Kyaukphyu								72	120	96	2	0	0	0
69	Indonesia	Surabaya (Tanjung Perak)								48	48	48	11	0	0	4
70	Indonesia	Semarang								48	144	96	2	0	0	0
71	Indonesia	Jakarta (Tanjung Priok)								48	48	48	0	4	0	6
72	Indonesia	Bojonegara								48	48	48	0	2	0	6
73	Indonesia	Belawan								48	144	96	5	0	0	0
74	Indonesia	Pontianak								48	144	96	4	0	0	0
75	Indonesia	Balikpapan								48	144	96	0	2	0	0
76	Indonesia	Ujung Pandang								48	144	96	0	2	0	0
77	Brunei Darussalam	Muara								24	24	24	0	2	0	0
78	Bangladesh	Chittagong								120	96	108	2	0	0	0
79	India	Jawaharlal Nehr								72	72	72	30	30	30	0
80	Sri Lanka	Colombo								72	72	24	4	2	1	5
81	Pakistan	Karachi								96	72	84	10	10	10	4
82	Europe Rosia									72	48	60	0	0	0	100
83	PNW									48	24	36	0	0	0	100
84	PSW									48	24	36	0	0	0	100
85	East Coast of North America									48	24	36	0	0	0	100
86	West Coast of South America									96	96	96	0	0	0	100
87	East Coast of South America									72	96	84	0	0	0	100
88	Middle East									48	24	36	0	0	0	100
89	Mediterranean									72	72	72	0	0	0	100
90	Europe									24	24	24	0	0	0	100
91	Africa									216	336	276	0	0	0	100
92	Oceania									24	72	48	0	0	0	100

Influencing Trip Departure Decisions: The Role of Institutions

Anthony T. H. Chin[1]

[1] Associate Professor, Department of Economics, National University of Singapore, Blk AS2 Level 6, 1 Arts Link, Singapore 117570; PH: +65-9069-4769; Email: anthonychin@nus.edu.sg

Abstract

Rising levels of affluence, living standards and diversity of lifestyles result in rising expectations on the performance of the road transport sector. The Vehicle Quota Scheme and Electronic Road Pricing remain the cornerstones of Singapore's transport policy towards private transport. Few drivers however, are drawn to public transport. This study adopts the behavioural approach to explore possible solutions to smoothing the peak-hour travel by focusing on time-based differential pricing. We propose the introduction of staggered concessions for peak hours as commuters have different values of time. This implies that some are willing to pay more without changing travelling habits while others would take advantage of a concession and travel off-pea. However, the concession scheme would have an even greater impact if flexible working hours is introduced. This is drawn from the observation that a higher concession rate would tend to have a bigger variance in deviation. A lower concession rate is less attractive and few people would be enticed. A concession rate 5 times as large would be deemed as being significant and more than five times the amount of people would be enticed to change trip departure decisions.

Introduction

Changing departure time was recognised as an important response to both increasing congestion and transport policy, second only to routing changes (SACTRA 1994). In spite of Singapore's 710.2 square kilometre, traffic speeds fall below optimum speeds in particular during the morning and evening commute. This is in spite of a Vehicle Quota Scheme (VQS) where Certificates of Entitlement (COE) are obtained from a bidding process for the purchase of a new automobile. Usage is subject to Electronic Road Pricing (ERP) along roads within the CBD, selected expressways and arterial roads. In spite of a comprehensive public transport system, patronage on

public transport has fallen. Public transport modal share during the morning peak period fell from 67% in 1997 to 63% in 2004 (Lim 2009). Trips by public transport increased by 16% while trips by automobile increased by 31%. This led to an overall fall in public transport mode share from 63% in 2004 to 59% in 2008. The Land Transport Masterplan (LTMP) introduced in 2008, set a target of 70% by 2020.

Increasing the number of buses may just marginally increase commute times and could contribute to traffic congestion. The Mass Rapid Transit (MRT) train headways are operating at optimal. Adding more rail lines will improve travel times and crowding but limiting automobile ownership through the VQS would result in corner solutions.

The main objective of this paper is to explore the opportunity of 'tweaking' commute through smoothing peak travel by changing trip departure times, routes and modes based on time-based differential road pricing. The feasibility and effectiveness of time-sensitive pricing strategies will be investigated with the aim of altering travel behaviour. This involves the introduction of staggered concessions for travels away from peak travel. Achieving general deviation from the original peak hours might be the solution: bringing the crowd away from peak hour, as opposed to bringing the crowd away from congested routes. Since peak-hour congestion is a time-related, it is most fitting that a solution targeted at deviating departure times be considered.

In addition commuters have different values of time. There are several categories of commuters, a) The majority of commuters who would be willing to pay more than change travel habits; b) Those would respond to maximum concessions and travel off-peak, and c) Those somewhere in-between. To smooth out travel is to design the appropriate stimulus to nudge commuters to make optimal choices. This may include introducing more premium bus services to ply dedicated routes at specific times.

Economic Fundamentals of Mobility

There have been several attempts in looking at influencing trip departure times. Holyoak (2007) looked the development of a peak period departure timing model for car users and the additional abilities of a model to represent the influence peak spreading policy options. While many studies looked at spreading the peak from various perspectives (Chin 1990, 2000, Koster 2007, Van Wren et al. 2007), few have looked at 'tweaking' trip departure by public transport commuters from a behavioral approach. The attractiveness of public transport depends on the quality of service in particular reducing waiting and journey times and overcrowding during the peak. Both the productivity of the services and the utility function of consumers need close evaluation. One postulation is that the productivity of the public transport sector has been exhausted. If so what is the role of technology, change in the market structure and innovation in increasing productivity.

Most public transport commuters are in 'corner situations' as do not own an automobile due to the high cost associated with owning one and the VQS. When

choices are limited or made unachievable by design, consumption is sub-optimal and welfare loss occurs. Figure 1 shows consumption of public transport trips at point B before improvements in public transport. Improvements in quality of service which results in time savings $PT_{T1\ to}\ PT_{T2}$, shifts the 'generalized budget constraint' from AB to AC, pushing the where the trips increased from B to C. When service strategies in public transport take into account price-travel time consideration based on the different preferences of different consumers, more of the market is captured and at the same time increased welfare. There is some indicators that reductions in transit or transfer time between public transfer modes may attract automobile users if packaged with other policies.

Public transport users are heterogeneous and utility function $U_2(PT_{C1},PT_{T2})$ represents a user which values time such as a automobile owner who needs a lot of compensating effect to give up his automobile. The high cost of ownership has led to an endowment effect which affects 'stickiness' to car usage. This study focuses on public transport commuters and the impact of differential pricing between peak and non-peak hours. A lower fare during the off peak will mean a comfortable ride with fewer commuters. Elasticities or the measure of responsiveness of commuters between time departures would be crucial measure to account for change in behavior and a measure of the effectiveness of the different schemes. Across schemes, effectiveness can be measured in instances such as, testing elasticity of peak hour travel (with surcharge) vis-à-vis elasticity of a "uni-fare" system (where there is no time-price differential). Various elasticities from each scheme allows comparison of the welfare loss that could arise given price differentials in each scheme. This enables the authorities to minimize welfare loss. Intuitively, high elasticities mean lower welfare losses.

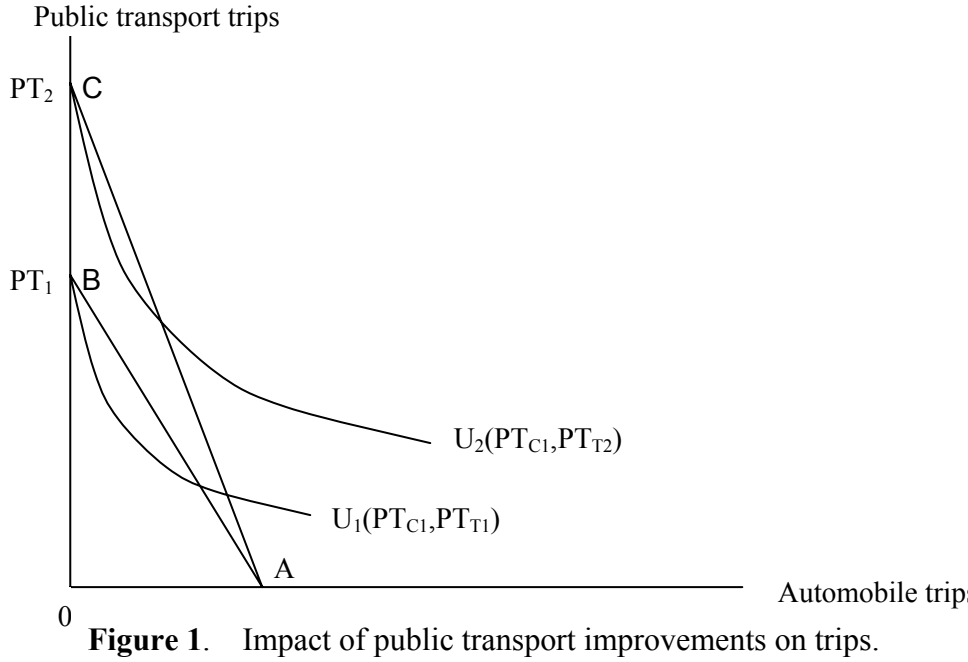


Figure 1. Impact of public transport improvements on trips.

Limitations to Frequency Increase

Surveys were conducted during both morning and evening peaks at the vicinity of MRT stations and bus stops within the CBD as 80% of all commute trips are made there. In addition the experimental survey consists of two sections. Section one focus on socio-economic variables and section two on builds upon current trip characteristics and creates various scenarios on concessions and surcharges at various times of travel. Regression analysis was to analyze the response to staggered concession scheme. Two different models are used to test the significance of concession.

The scenarios also explored the feasibility of other possibilities such as increasing bus frequency as a measure to ease peak hour crowd. The survey design was based on 10 bus stops 10 bus stops and 10 MRT stations across the CBD and its vicinity. Information on service operations for both buses and MRT such as journey times and frequency of service were obtained through the website[1]. Additional data such as average actual number of buses per hour, average duration of loading/unloading and average waiting time during peak hours were gathered through on-site surveys. This gives some indication on whether adding more buses might improve service quality such as arriving within 60 minutes leaving their homes. Observations were made on the number of buses from the particular service arriving at the bus stop within an hour. For example, "Service number "857" has a frequency of 8-12 minutes during morning peak hour. Using the median duration between each bus, which is 10 min, there will be 6 bus "857" arriving at the Raffles Hotel Bus stop within the hour.

Based on random samples of the mean waiting time at bus bays on-site is 1 min and 6 s. The total time spent spend in bays (in an hour) is given by the product of average number of buses at the bus stop in an hour, which is the sum of the expected total number of buses across all services and the mean waiting time of 1.1 min.

The proposed solution of increasing bus frequencies is examined through calculating the percentage of time that the bus stops are utilized. Currently, the practice of bus drivers during peak hour is to allow passengers to board and alight even when there is already a bus in the bus bay. Hence, the assumption is that each bus stop allows passengers from 2 buses to board and alight concurrently. This assumption can be supported by the large number of buses at the bus stop in an hour within peak period which means that buses typically arrive in pairs if not more. With that assumption, the percentage of time that each bus stop is utilized is calculated using this formula:

$$\text{Percentage of time bus stop is fully utilized during peak hour} = \frac{(\text{Total duration that buses spend in bays}/2)}{60} \text{ X } 100\%$$

[1] SBST, http://www.sbstransit.com.sg/transport/trpt_bus_timetable.aspx; SMRT,http://www.smrtbuses.com.sg/ebusguide/busguide.asp

For example, it was noted that bus stop *opposite AIA Towers* is utilized 85.25% of the time during peak hour. The average utilization rate for the whole network is 75.63% in a typical work day. This high utilization rates of bus stops, indicates to some extent the limit to the number of buses that *can* be added. This seemingly diminishing marginal returns is backed by the 2008 LTMP which states that "even with an increase in bus and train frequencies, the time taken to travel on public transport can only be reduced from 1.7 to 1.5 (as compared to car travel) (LTA 2008).

A possible solution to reduce this utilization rate would be to split out bus services to operate on alternate bus stops. An example would be the bus stops at Tangs Plaza and Lucky Plaza, which runs along Orchard Road. Although they are along the same road, the buses servicing this stretch would stop at only either stop. This allows for additional bays for loading/unloading passengers and helps in reducing waiting time. This is shown through utilization rates that are below that of the average calculated above, for both the bus stops.

However, we also need to take into consideration the constraints of the road capacity. With only one bus lane, increasing buses servicing in the area leads to a problem. For example, along Orchard Road, the bus stops are located on the bus lanes without a separate, dedicated bus bay. Hence, even with more bus stops, the increase in frequency would lead to more buses getting stuck along the bus lanes waiting for other buses at bus stops to move off. This results in a decrease in optimal speeds of the buses such that even with an increase in frequency, the total travel time taken may end up being longer. However, increasing bus frequencies is a short-term solution and has limited impact in improving the quality of the journey given the constraints above.

Trains in the MRT network face space constraints given that platforms in current stations can only take a maximum of six coaches. Given the maximum wait time for trains is 3 minutes for morning peak hours and the safety distance, there is a limit to increases in frequencies. There is a need to take a different approach – one that does not focus on *contributing* to the already overcrowded roads and space, but rather one that is aimed at smoothing out commute traffic through travel at alternative *times*.

Changing Trip Departure Times

Surveys were conducted during morning and evening peak hours targeted at working adults in the CBD and tertiary students (including polytechnic and university students). Different questions were asked to ensure relevancy of the survey, and also to prevent a bias in the findings because these two groups of commuters actually pay different travel fares, and have different spending power. They make up the main bulk of the peak hour travel in the city area. Several hypotheses were tested. These are, (1) An individual from a higher household per capita income would not be affected by the differential fares and less likely to deviate his/her travel timings, or by a smaller magnitude, as compared to another individual who has a lower household per capita income; (2) A commuter who is already travelling very early in

the morning might be less motivated to travel even earlier in the morning, assuming transport is made available; (3) A commuter who arrives at a later time might be more easily moved to travel earlier. This tests if the value of time is indeed different when one travels early or later; (4) Evening peak travel is more responsive to higher fares between 6.30–7.30 pm and lower fares for later departure times; (5) Travellers are responsive to travelling outside peak periods though incentives such as concessionary *rates*. Different rate of concession in each case would provide for a rough estimation of commuter's value of time; (6) Introduction of a flat rate surcharge for travel between 8:00–9:00 am and 6:30–7:30 pm, would alter trip departure; (7) Students are more responsive incentives, surcharges and quality changes.

Staggered Concessions Schemes[2]

The objective here is to ascertain if the commuter is willing to leave 10 minutes earlier or later if offered concessionary discounts during morning or evening peaks. The final model considered incorporated household per capita income (HCI). High HCI would be less affected by the concession. Intuitively, transport expenditure would make up a small proportion of the disposable income. HCI is introduced as an interaction term[3] (HCI and concessions). The other variable in the equation is a binary term "EARLY". This depicts originally arrival time at work before 0830 or after i.e., 0831[4]. It is expected to provide an insight as to whether commuter's value of time or willingness to switch departure time is significantly dependent on how early they originally travel.

From the Table 1, the equation can be formed:

$$\underset{\substack{(-0.129698)\\ (1.41218)}}{\text{DEPTIME} = -0.1269} + \underset{(8.516751)}{1.414\ \text{Concession}} \underset{(-2.959583)}{-0.1699\ \text{HCI x CONC}} + \underset{(1.41218)}{0.0395\ \text{EARLY x CONC}}$$

Table 1 suggest that concession rates remain a significant variable in influencing trip departure times. HCI has a significant interaction with concession rates. A commuter with a higher HCI is less likely to travel earlier. It is also interesting to note that the interaction term of "EARLY X Concession" yields a statistically insignificant effect on the magnitude of deviation. This suggests that the commuter's value of time is marginal in influencing trip departure time. This is perhaps due to inflexible work start times.

---

[2] Transport for London introduced various discount schemes aimed at shifting peak-hour traffic targeted at different categories of travelers in 2009. Singapore's Mass Rapid Transit (SMRT), has launched a scheme in 2008 that discounted fares by 10 cents if passengers board trains outside the city and alight in the city area before 0730. However commuters who traveling at around 0830 would not find this attractive in trading to travel an hour earlier for a 10 cent discount!

[3] An interaction term shows the effect of HCI on concession.

[4] 0830 is the mean arrival time in the city from the survey.

Table 1. Impact of concessions on morning peak departure time.

Dependent Variable: DEPTIME				
Method: Least Squares				
Morning Data				
Included observations: 450				
	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.126984	0.979078	-0.129698	0.8969
Concession (per 10 mins)	1.41435	0.166067	8.516751	0
HCI x CONC	-0.169938	0.05742	-2.959583	0.0032
EARLY x CONC	0.039563	0.095984	1.412182	0.4204
R-squared	0.189967	Mean dependent var		8.622222
Adjusted R-squared	0.184519	S.D. dependent var		9.071085
S.E. of regression	8.191553	Akaike info criterion		7.052933
Sum squared resid	29927.29	Schwarz criterion		7.08946

Table 2. Impact of concessions on evening peak departure time

Dependent Variable: DEPTIME				
Method: Least Squares				
Evening Data				
Included observations: 450				
	Coefficient	Std. Error	t-Statistic	Prob.
C	1.015873	1.100191	0.92336	0.3563
Concession (per 10 mins)	1.674654	0.169168	9.899328	0
HCI x CONC	-0.360827	0.062773	-5.748114	0
R-squared	0.18014	Mean dependent var		9.177778
Adjusted R-squared	0.176472	S.D. dependent var		10.14327
S.E. of regression	9.204861	Akaike info criterion		7.283985
Sum squared resid	37874.07	Schwarz criterion		7.31138
Log likelihood	-1635.897	Hannan-Quinn criter.		7.294782
F-statistic	49.10749	Durbin-Watson stat		1.980616
Prob(F-statistic)	0			

DEPTIME = 1.0158 + 1.6746 Concession - 0.3608 HCI x CONC
(0.92336) (9.899328) (-5.748114)

The results suggest that an increase in HCI results in a marginal effect on departure time, at 1% level of significance. The exact magnitude of the time changes is not a main concern when introducing the staggered concessions but rather does it smooth out commuter-congestions given that have different income levels and respond to different levels of concessions. Concessions do affect departure times. High income commuters with high value of time and value comfort would probably not optimise their time by not changing while those with lower values of time would choose to later to take advantage of concessions. Here, a win-win outcome results as peak travel *within* public transport is smooth.

Peak Travel Surcharge

The response to the introduction peak surcharge was investigated to ascertain elasticity. A 10% surcharge on current public transport fares between 0800 to 0900 would lead to a fall in passenger trips by 22.6%. Most responded that they would shift travel to before 0800 or after 0900. There are welfare implications in addition to creating peaks before and after the hour. The responsiveness increases to 41.7% in response to a 10% surcharge imposed on fares for travel between 1800 and 1900. The significant difference in elasticity may be explained by a greater flexibility in travel times after work given that there is no compulsory arrival time to the home.

Both of these elasticities are significantly higher than the public transport elasticity during peak hour, which is estimated to be between –0.15 and –0.3. This is because these elasticities were calculated with respect to a change in public transport fare for a given distance and commuters largely did have alternatives other than automobile. On the other hand, the elasticities of *peak hour travel* with respect to surcharge are naturally higher because of the availability of close substitutes i.e. non-peak travels with the assumption of flexible work hours.

Tertiary students in Singapore are entitled to purchase monthly concession pass on their EZ-link cards for a flat fee and enjoy unlimited trips on buses and 4 trips per day for trains. However, not all students choose to purchase this monthly pass and hence it would be sensible to first get a gauge of the proportion of students actually using this concession scheme, and then try to estimate the impact of limiting student concessions to non-peak hour travels.

The below calculations were weighted according to each student's preference so that a student who makes few trips during the peak a month would have a smaller effect as opposed to a student who make many peak hour trips.

Results in Table 3 show that limiting concession travels to non-peak hour travels would result in an estimated drop in student ridership by 11% in the morning and 20% in the evening. Once again, the difference in proportionate deviation could be due to greater time flexibility in the evenings when students can more easily avoid travelling between 6:30 pm and 7:30 pm, as compared to in the morning between 8am and 9am when there might be morning lessons.

Table 3. Travel patterns of students.

Morning travel patterns	
Total morning peak trips	480
Total peak trip when concession not applicable during peak	423
Change in number of trips	–57
Weighted proportional change in number of peak trips	–0.11875
Evening travel patterns	
Total evening peak hr trips	618
Total peak trip when concession not applicable during peak	490
Change in number of trips	–128
Weighted proportional change in number of peak trips	–0.207119741

Limiting students' concession is therefore a feasible and effective manner to further reduce peak hour congestion, complementing any other schemes that would be put in place for working adults. Whether this is a socially beneficial is debatable as this depends on issues such as equity, change in school start times and access to education versus congestion.

Travel Behavior Can be Influenced

A staggered concession scheme that charges a lower fare before the peak hour in the morning and after the peak hour in the evening seems ideal in incentivizing commuters to change trip departure times. The concession is in the form of a fixed rate of discount for every additional time period (10 minute block) earlier from one's current departure time. A concession rate of 8-10% per 10 minutes is recommended would 'nudge' commuters to change trip departure of 10–14 minutes for the average commuter. This is because each commuter has a different value of time. The peak of the peak hour period (from 0800 to 0900) occurs at 0830. The initial distribution of commute trips during the peak period fell after the implementation of the concession (see Figure 2). While the shift appears marginal, commuters arrive before 0830. Those who leave after 0830 less incentive to change travel habits, due to perhaps work start time constraints, higher value of home or travel time. This would cause the new curve to skew towards the left. This new curve is lower and wider than the original curve, with an increased ridership in the earlier hours and a decreased ridership at timings nearer to 0830.

Perhaps a more comprehensive scheme which begins at 0800 and 1900 with discounts from between 10% to 30% may be effective in changing trip departure as the majority of commuters are not willing to change more than 30 to 40 minutes from their current departure times. Some indicated that flexible work hours and innovative work methods may influence change.

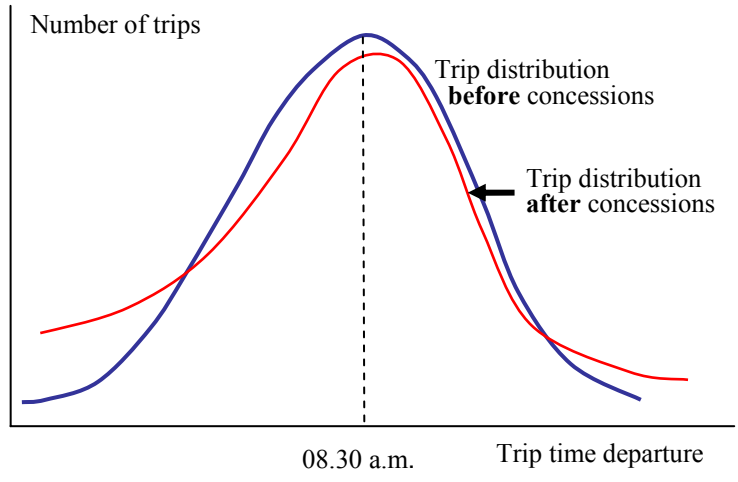


Figure 2. Smoothing trip departure time through concessions.

The alternative to incentives is imposition of a peak surcharge which seems effective and efficient. However, it suffers from two drawbacks. First, with a non-staggered system, a commuter traveling before and after the surcharge takes effect would be paying a significantly different fare even if one is a minute late. It is not unreasonable to expect reluctance to travel just before the peak period at the margin. Travel demand would 'pent up' until 0901 or 1931. Secondly, "secondary peaks" could occur at the fringes of the surcharged peak.

A simple scheme that can be introduced without much administrative issue is student concessions that limit student travel to non-peak hour travels. Students are generally more flexible with departure times. It is estimated that limiting student concessions to off-peak travels would reduce up to 20 percent of student ridership from 0630 to 0730. This however, would require school start times to change.

Reporting time at work influences a commuter's departure time. Concession schemes would have a greater impact if flexible working hours are introduced especially for non-key and non-front desk employees. A higher concession rate would lead to a larger variance in departure times. A lower concession rate is less attractive and few would switch. However in surveying random commuters at MRT stations, selection bias may have occurred. This is a problem because the commuters who stopped to complete our surveys are likely to have a lower value of time. Because of this lower value of time, the surcharge that is required for commuters on the whole, to change their travel time may be underestimated. However, this is a very minor concern and the over-estimation may be notably insignificant.

References

Chin, A. T. H. (1990). "Influences on commuter trip departure time decisions in Singapore", *Transportation Research Part A: General*, 24(5), 321-333.

Chin, A. T. H. (2000). "Impact of the road pricing scheme on commuter trip departure time." *Traffic and Transportation Studies, Proceedings of ICTTS 2000*, Beijing, People's Republic of China, Reston, VA:ASCE, 0-7844-0503-4, 2000, 998.

Holyoak, N. M. (2007). "Modelling the trip departure timing decision and peak spreading policies", *European Transport Conference Proceedings,* Association for European Transport, Leeuwenhorst Conference Centre, The Netherlands, 17-19.

Koster, P. (2007). "The costs of uncertain travel times for car travelers", *European Transport Conference Proceedings,* Association for European Transport, Leeuwenhorst Conference Centre, The Netherlands, 17-19.

Land Transport Authority, (2008), *Land Transport Master Plan*, Government of Singapore, Singapore.

Lim, R. (2009). "Speech delivered at the 2nd World Roads Conference", *SUNTEC CITY CONVENTION CENTRE*, Singapore, 26th October 2009.

SACTRA 1994, "The Standing Advisory Committee on Trunk Road Assessment", Department for Transport, London, U.K.

Transport for London, 2009, http://www.tfl.gov.uk/assets/downloads/bus-and-tram-discount-photocard-leaflet-april-2009.pdf , assessed 9 February 2009.

Van Wren T., Carmichael S., Polak J., Hyman G., and Cross, S. (2007), "Modelling peak spreading in continuous time", *European Transport Conference Proceedings,* Association for European Transport, Leeuwenhorst Conference Centre, The Netherlands, 17-19 October 2007.

Traffic Modal Splitting Model Based on Game Theory with Incomplete Information

Chunjiao Dong[1], Chunfu Shao[1], Yuewen Gao[2], Yinhong Liu[1] and Dan Zhao[1]

[1] MOE Key Laboratory for Transportation Complex Systems Theory and Technology, School of Traffic & Transportation, Beijing Jiaotong University, P.O. Box 100044, Beijing, Haidian District Shang Yuan Cun 3#, China; PH: (86) 10-51683665; Email: dongchj2006@163.com, cfshao@center.njtu.edu.cn

[2] Transportation Planning & Design Department, Tianjin Bohai Urban Planning & Design Institute, P.O. Box 300450, Tianjin, Tanggu District Zhongxin North Road, Sunny Zone 2-1-02, P.R.China

Abstract

Traffic modal splitting plays an important role in the traditional four-stages-forecast model. Usually in a stable urban traffic network, traffic supply can meet the demand basically, and a variety of traffic modal will bear a fixed proportion of passengers. Compared with other trip mode, public transportation is often influenced by some uncertain factors, e.g. transferring and waiting, which will cause different marginal cost and finally come to a new balance or imbalance between demand and supply. In this paper, traffic modal was supposed as a rational gamer, and generalized trip cost was considered as the cost of gamer. Then, traffic mode share can finally be got via comparing supply of all kinds of traffic means. A case study is conducted and compared with disaggregate modal, which shows the game theory-based model is more effective and the outcome is closer to the real traffic modal share.

Key words: traffic modal spilt; game theory with incomplete information; function of generalized trip cost; marginal cost

Introduction

As the third step of the traditional four stages forecast modal, traffic modal spilt plays an important role in the traffic planning, for which there have been a great of methods, such as curve method of traffic model split, tripendmodel, trip conversion model, disaggregate model and so on (Shao 2004; Zhu 2008; Gao 2007). However, there is still some weakness in existing models, if some unpredictable factors are concerned, e.g. individual customs, hobbies. There are two modeling approaches of traffic modal spilt model. One is to research the change of traffic demand on condition that everything goes on as what it used to be, and the other one is to build more traffic facilities or make some policy to influence people's choice in traffic mode, in which way mode share can be got (Zhang 2005; Zhang 2008; Wu 2005; Yi 2009). The modeling follows the first approach in this paper, which is based on game theory with incomplete information on condition that traffic supply and demand can get a balance in the whole urban transport network.

In the traditional four-stages forecast modal, traffic modal splitting is often got via multiplying mode share by trip summation. In this paper, traffic modal was supposed as a rational gamer, and generalized trip cost was considered as the cost of gamer. Then, traffic modal spilt can finally be got via comparing supply of all kinds of traffic means.

Trip Mode Split

Modal split is the division of travel into the various transportation modes, which builds model to research people's choice of trip modes via forecasting change of traffic demand and supply, e.g. infrastructure rehabilitation and other outer environment change (Zhou 2005; Liu 2008; Liu2008).

Urban transportation modes can be divided into three kinds: the first kind is being uncompetitive with other modes, represented by walking; the second is being competitive with other modes, represented by private car and bicycle; the last kind is competitive cooperation, represented by access bicycle and walk **(**Hu 2003**)**. The main characteristics of above 3 are shown in Table 1.

Table 1. Characteristics of urban traffic modal.

Type	Characteristics
Uncompetitive mode	People will choose any traffic mode freely if they can afford
Competitive mode	People will choose traffic mode with convenience completely
Competitive & cooperative mode	People will choose two or even more modes with convenience

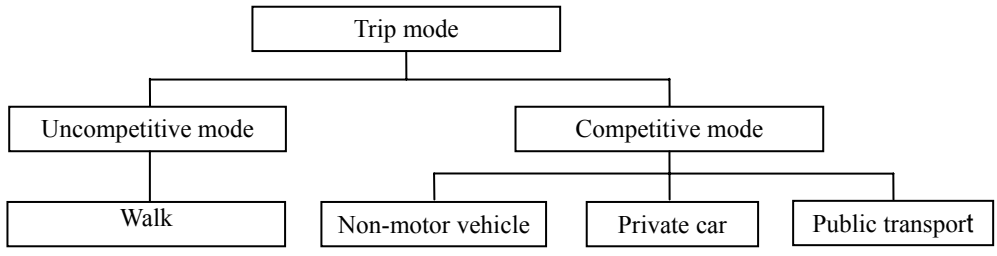


Figure 1. Trip mode.

Table 2. Suggested walk-sharing rate in different scale city of China.

Scale of city	Characteristics	Suggested rate (%)
Big & mega	Developed economic, superbly appointed traffic facilities, high service quality public transport	20–30
	Less developed economic, general traffic facilities, low service quality public transport	25–35
	Developed economic, land resources abundant, superbly appointed traffic facilities	25–35
Small & medium-sized	Less developed economic, land resources lack, general traffic facilities	30–40

It's difficult to order the mode and path selection when various trip modes are competitive and cooperative each other, because both of which are always carried out in the same step. Usually, competitive & cooperative mode is ruled out when planning urban traffic, and then the trip mode can be simplified into two kinds in Figure1: uncompetitive mode and competitive mode, which will be used in this paper.

Walk is the only uncompetitive mode. Therefore, the volume of people choosing competitive modes can be got by full mode trip volume diminishing traffic volume by walking, which can be got by walking sharing rate (Table 2, Yong-zheng Kang 2006) multiplying predicted value of full mode trip volume.

Generalized Trip Cost Function

When traffic modes gaming, cost function always determines the quantity of traffic demand. The analysis shows that generalized trip cost is in connection with traffic impedance, which means generalized cost will increase along with the increase of trip time and cost while traffic demand will decrease. Generalized trip cost function of various trip modes, which are public transport, non-motor (mainly bicycle) and personal car, will be researched separately as follows.

Generalized Public Transport Trip Cost Function. Public transport trip cost function should consider trip time and cost factors generally so as to reflect the characteristic of public transport trip in truth.

People often consider trip time and fare firstly when choosing public transport. The trip time is determined by trip distance and transfer times, and public transport fare contains single ticket and metered fare. Therefore, high marginal cost means high transfer rate and metered fare system in public transport trip, while low marginal cost means no transfer and single ticket system. The generalized trip cost function can be represented as Eq. (1):

- for high marginal cost case:

$$c_H = \alpha R_m T_H + \beta L_H P_H / 100 \tag{1}$$

where c_H is generalized public transport trip cost (RMB), R_m is time value, which can calculate as GNI per capita (RMB/h), T_H is traffic impedance in trip with transfer (h), L_H is trip distance (km), P_H is public transport fare (cent/km), α, β, is undetermined constant, which can be determined via measured data fitting. T_H in Eq. (1) should reflect service level of the whole public transport network while not be limited into a particular public transport network, which can be represented as Eq. (2):

$$T_H = \frac{L_{ij} \cdot f_m}{\overline{V}_{ij}} + \frac{(T_d + T_b)}{60} + n \times T_h \tag{2}$$

where L_{ij} is shortest path between traffic zone i and j, f_m is nonlinear coefficient of public transport network between traffic zone i and j, $\overline{V}_{ij}$ is mean speed of public transport network between traffic zone i and j (km/h), T_d is mean waiting time of public transport network between traffic zone i and j (min), T_b is mean duration of walking (from origin to stop/station and from stop/station to destination) of public transport network between traffic zone i and j (min), T_h is mean transferring time of public transport network between traffic zone i and j (h), n is mean transfer times of public transport network between traffic zone i and j.

- for low marginal cost case:

The single ticket both decreases and increases influence of trip distance when choosing traffic path. For low marginal cost case, generalized cost of passenger by public transport can be represented as Eq. (3):

$$c_L = \alpha R_m T_L + \beta P_L \tag{3}$$

where P_L is public transport fare with single ticket system (RMB), T_L is traffic impedance in trip without transfer (h), $T_L = \frac{L_{ij} \cdot f_m}{\overline{V}_{ij}} + \frac{(T_d + T_b)}{60}$, other symbols ibid.

Generalized Non-motor (Bicycle) Trip Cost Function. The cost function can be represented as Eq. (4):

$$c_1 = \alpha R_m \times (\frac{D}{\overline{V}_z} + T_{zq}) + \beta W_z \times D \tag{4}$$

where c_1 is generalized bicycle trip cost (RMB), D is length reduction value of the shortest path, $D = L \cdot z$, where z is reduction coefficient, general admission is 0.6 to 0.8, $\overline{V}_z$ is mean speed of bicycle between traffic zone i and j, suggesting valuing 8–10 km/h, T_{zq} is mean duration of walking between traffic zone i and j when collecting and parking bicycle (min), W_z is depreciation cost of bicycle (RMB/km), other symbols ibid.

Generalized Personal-car Trip Cost Function. Personal motor vehicle contains car, motorbike, minibus and so on. Because of too many vehicles on the road and crossing signals factors, most of vehicles run at a similar speed. Therefore, the following will mainly research personal car, and the cost function can be represented as Eq. (5):

$$c_2 = \alpha R_m \times (\frac{L_{ij}}{\overline{V}_c} + T_{cq}) + \beta[(P_{cy} + W_c) \times D + P_{ct}] \tag{5}$$

where c_2 is generalized personal-car trip cost (RMB); $\overline{V}_c$ is mean speed of personal car between traffic zone i and j (km/h); T_{cq} is mean duration of walking between traffic zone i and j when collecting and parking car (min); P_{cy} is running cost of personal car between traffic zone i and j, which is mainly oil wear cost (RMB/km); P_{ct} is mean parking fee between traffic zone i and j (RMB); W_c is depreciation cost of personal car (RMB/km), other symbols ibid.

Traffic Modal Splitting Model

A game consists of a set of players, a set of moves (or strategies) available to those players, and a specification of payoffs for each combination of strategies (Zhang 2004). Game theory attempts to mathematically capture behavior in strategic situations, in which an individual's success in making choices depends on the choices of others (Caplin 1997). According to information classification, gaming consists of full information game and incomplete information game (Hausken 2000). According to order of making strategies, gaming consists of static game and dynamic game. In the paper, traffic mode are supposed as gamers, who will make static gaming with each other under incomplete information, and where generalized trip cost and trip demand are separately input and output of gamers.

Put the case that, non-motor (bicycle) is mode 1, personal car is mode 2, and public transport is mode 3. If C_{rs} is inverse demand function, it can be represented as Eq. (6):

$$C_{rs} = D_{rs}^{-1}(Q_{rs}) = \begin{cases} a - Q_{rs} & Q < a \\ 0 & Q \geq a \end{cases} \quad (6)$$

In order to research further, some rules are supposed as following:

firstly, there's no fixed cost in trip by mode 1 or 2, and marginal cost $c_i < a$ ($i = 1, 2$), all of which will be shared by three traffic modes. Therefore, the cost of mode 1 can be represented as:

$$C_1(q_1) = c_1 q_1 \quad (7)$$

where q_1 is quantity of traffic supply from mode 2; the cost of mode 2 can be represented as:

$$C_2(q_2) = c_2 q_2 \quad (8)$$

where q_2 is quantity of traffic supply from mode 2.

What's more, cost function of mode 3 is confidential to mode 1 and mode 2. The marginal cost of mode 3 is c_H for the probability P and c_L for the probability 1-P, which is open to mode 1 and mode 2. This assumption is based on 3 characteristics of public transport:

- randomicity of waiting and transferring time, which are affected by departure interval, traffic delay and so on;
- volatility of walking time from origin (stop/station) to stop/station (destination), which is different for every traveler;
- difficulty of estimating time value when waiting, which is usually higher than time value on bus.

Therefore, the information of three traffic modes is anisomerous, and mode 3 is superior in possession of information when making decision compared to mode 2 and mode 3.

Traffic modes game should be transformed into normal form in order to get Nash equilibrium. The normal game is usually represented by an expression which shows the players, strategies, and payoffs for every combined strategy. There're three players in traffic mode splitting model: mode 1, mode2 and mode3. The strategy of each player is traffic supply quantity. Supposing traffic supply is continuous and divisible, the space of each traffic mode can be represented as:

$$S_1 = S_2 = S_3 = [0, +\infty) \quad (9)$$

where s_i is traffic supply, $q_i \geq 0$.

Generalized trip cost varies inversely with traffic demand. As the assumption above, mode 1 and mode 2 can forecast mode 3 will change its traffic supply according to the change of generalized trip cost. For high trip cost case, the optimum traffic supply of mode 3 ($q_3^*(c_H)$) can be represented as:

$$q_3^*(c_H) = \max_{q_3}[a-(q_1^*+q_2^*+q_3)-c_H]q_3 \tag{10}$$

where q_1^* is the optimum traffic supply of mode 1, q_2^* is the optimum traffic supply of mode 2. For low trip cost case, the optimum traffic supply of mode 3 $(q_3^*(c_L))$ can be represented as:

$$q_3^*(c_L) = \max_{q_3}[a-(q_1^*+q_2^*+q_3)-c_L]q_3 \tag{11}$$

It is known that the marginal cost of mode 3 is c_H for the probability p and c_L for the probability 1-p. The optimum traffic supply of mode 1 and mode 2 can be represented as following:

$$q_1^* = \max_{q_1}\left\{p[a-(q_1+q_2^*+q_3^*(c_H))-c_1]q_1+(1-p)[a-(q_1+q_2^*+q_3^*(c_L))-c_1]q_1\right\}$$

$$q_2^* = \max_{q_2}\left\{p[a-(q_1^*+q_2+q_3^*(c_H))-c_2]q_2+(1-p)[a-(q_1^*+q_2+q_3^*(c_L))-c_2]q_2\right\}$$

Therefore, traffic modal splitting model based on game theory with incomplete information can be constructed as following:

$$\begin{aligned}
&\max_{q_3}[a-(q_1^*+q_2^*+q_3)-c_H]q_3\\
&\max_{q_3}[a-(q_1^*+q_2^*+q_3)-c_L]q_3\\
&\max_{q_1}\left\{p[a-(q_1+q_2^*+q_3^*(c_H))-c_1]q_1+(1-p)[a-(q_1+q_2^*+q_3^*(c_L))-c_1]q_1\right\}\\
&\max_{q_2}\left\{p[a-(q_1^*+q_2+q_3^*(c_H))-c_2]q_2+(1-p)[a-(q_1^*+q_2+q_3^*(c_L))-c_2]q_2\right\}\\
&s.t.\begin{cases} q_i \ge 0 & (i=1,2,3)\\ c_i \ge 0 & (i=1,2)\\ c_H \ge 0 & \\ c_L \ge 0 & \end{cases}
\end{aligned} \tag{12}$$

First-order conditions of optimalization problems can be separately represented as following:

$$q_3(c_H) = \frac{a-(q_1^*+q_2^*)-c_H}{2},\quad q_3(c_L) = \frac{a-(q_1^*+q_2^*)-c_L}{2}$$

$$q_1^* = \frac{p\{a-[q_2^*+q_3^*(c_H)]-c_1\}+(1-p)\{a-[q_2^*+q_3^*(c_L)]-c_1\}}{2}$$

$$q_2^* = \frac{p\{a-[q_1^*+q_3^*(c_H)]-c_2\}+(1-p)\{a-[q_{1'}^*+q_3^*(c_L)]-c_2\}}{2}$$

Solving the equation set, which consists of the three first-order conditions, the outcome of traffic mode splitting can get as equation set (13):

$$\begin{cases} q_3^*(c_H) = \dfrac{a + c_1 + c_2 - (2+p)c_H - (1-p)c_L}{4} \\ q_3^*(c_L) = \dfrac{a + c_1 + c_2 - (3-p)c_L - pc_H}{4} \\ q_1^* = \dfrac{a - 3c_1 + c_2 + pc_H + (1-p)c_L}{4} \\ q_2^* = \dfrac{a - 3c_2 + c_1 + pc_H + (1-p)c_L}{4} \end{cases} \tag{13}$$

According to the outcome above, it can be concluded:

$$q_3^*(c_L) - q_3^*(c_H) = \frac{c_H - c_L}{2} > 0$$

The formula proves that $q_3^*(c_H)$ is smaller than $q_3^*(c_L)$, which is in line with reality. Then we can calculate traffic volume shared by various traffic modes by Eq. (14):

$$Q_i = k_i \cdot Q = \frac{q_i^*}{\sum_i q_i^*} \cdot Q \tag{14}$$

where Q_i is traffic volume shared by mode i, k_i is splitting proportion of mode i.

Empirical Study

Take the case of the city of Changchun, the model can be proved effective. Table 3 is the trip mode splitting proportion of people in Changchun. Via data collection, we can get that mean speed of personal car is 25.4 km/h, and mean speed of public bus is 11.89km/h, and other parameters of model are shown in Table 4.

Table 3. Trip mode splitting proportion of people in Changchun (%).

Mode type	Public traffic	Non-motor vehicle	Personal-car	Walking	Sum
Full traffic mode	23.30	32.10	17.60	27.00	100
Competitive traffic mode	31.92	43.97	24.11	-	100

Table 4. Parameters of model.

Time value	α	β
4.81	0.76	0.04

Table 5. Generalized trip cost of various traffic modes (Unit: RMB).

	c_1	c_2	c_L	c_H
Generalized trip cost	2.98	5.19	3.36	4.13

Imputing the actual data to model, we can get generalized trip cost of various traffic modes as Table 5. Then, imputing the cost into Eq. (13), we can get predicted value of traffic mode splitting.

In order to proving rationality of new model, we compared result of model based on theory with incomplete information with disaggregated model, as Figure 2. The splitting proportion of new model is closer to actual value than disaggregated model.

Error analysis of new model result is shown in Table 6. Mean absolute error of new model is less than 3%, and mean relative error is less than 10%, which can meet define required. The value of traffic modal splitting forecasting of personal-car meets the reality to a great extent, absolute error of which is the least among various modes. Although many factors affect the cost of public transport, relative error of public transport is the least thanks to the division of high and low marginal cost in model based on game theory with incomplete information.

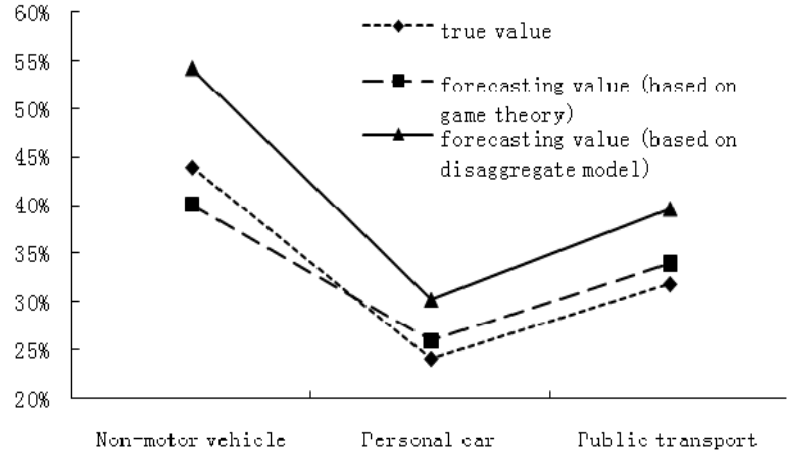


Figure 2. Comparison diagram of traffic modal splitting forecasting

Table 6. Error of predicted value in game theory based mode with incomplete information (%)

	Non-motor vehicle	Personal car	Public transport	Mean value
Absolute error	3.92	1.85	2.06	2.61
Relative error	8.91	7.69	6.46	7.69

Conclusions

In this paper a new traffic modal splitting model is constructed, which is based on game theory with incomplete information. In new model, generalized trip cost of various traffic modes determines traffic demand was considered. Using mode based on game theory with incomplete information can split traffic modes both inner zone and inter-zones. What's more, the example of Changchun proved that new model is superior to the disaggregated mode.

However, as rail transport carries more passengers, further studies such as traffic modal splitting of rail transport need to be considered when using the model. Meanwhile, inverse demand function under different traffic condition should be researched.

References

Shao, C. F. (2004). *Traffic Planning*, Beijing: China Railway Publishing House, 151-169.

Zhang, R., and Wu, H. Y. (2005). "Study on reasonable pattern of traffic mode in Beijing." *Journal of Beijing Institute of Civil Engineering and Architecture*, 21(1), 24-27.

Zhu, S. Y., Li, J., Li, A. X. and Wang, H. (2008). "The disaggregate logit model for mode split of rapid rai1way in the urban agglomeration." *Journal of Wuhan University of Technology* (*Transportation Science & Engineering*), 32(6), 1063-1066.

Gao, T., and Xu Y. (2007). "Implementation of nested logit model in transCAD." *Computer and Communication*, 25(4), 82-88.

Zhang, H. R. Ren,G., and Wang, W. (2008). "Application of discrete choice model in trip mode structure forecasting." *Journal of Transportation Systems Engineering and Information Technology*, 8(5), 44-49.

Wu, H. B., and Chen, Y. H. (2005). "A combinatorial model of modal split and traffic assignment for mixed urban traffic networks." *Systems Engineering*, 23(7), 77-80.

Yi, F. J. Deng, W. Zhou, Z. P. (2009). "Traffic modal splitting method for economic circle based on nested-Logit model." *Journal of Transportation Systems Engineering and Information Technology*, 9(4), 110-115.

Zhou,J. Wang, S. S. and Sui, S. X. (2005). "Study on nested logit model for transport mode split and its solution." *Journal of Shandong Jiaotong University*, 13(4), 28-31.

Liu, Z. M., Deng, W., and Guo, T. Y. (2008). "Application of disaggregate model based on rp/sp survey to transportation planning." *Journal of Transportation Engineering and Information*, 6(3), 59-64.

Liu., S., Zhou, T., and Hu, S. H. "Study on traffic mode-split based on neural network". *Shanxi Architecture*, 34(25), 21-22.

Hu, G. (2003). *Research for Optimization Practical Technology of Urban Public Transit Network and Stop*, Nanjing: Southeast University.

Kang, Y. Z., Niu X. Q., and Huang, S. G. (2006). "Study on the model of city bus route optimization." *Communications Standardization*, 145-147.

Zhang, C. W. (2004). "Cournot model and game theory." *Journal of Shandong Institute of Building Materials*, 18(2), 155-157.

Caplin, A., and Nalebuff, B. (1997). "Competition among Institutions." *Journal of economic theory*, 306-342.

Hausken, K. (2000). "Cooperation and between-group competition." *Journal of Economic Behavior & Organization*, 42, 417–425.

Inductive-based Machine Learning and the Application in Assessment of Ship Handling

Xinjie Han, Xianyao Meng and Song Meng

Information Science and Technology College, Dalian Maritime University, Dalian 116026, China; Email: xinyu2000h@sohu.com, mengxiany@163.com

Abstract

It is a matter of utmost importance to predict dangers accurately in advance and to prevent accidents from occurring for a ship, which is navigating on the sea. Based on analyzing the cases of a shipwreck which have occurred so far, we discovered that more than 80% of the accidents were caused by wrong instructions issued by the captain of a ship. Especially in severe weather or in an emergency, inaccurate instructions have caused fatal damages. Based on machine learning, a method for analyzing and assessing the operation instructions sent from the bridge is discussed in this paper. ID3 algorithm is applied to build up an assessing decision tree, which can be improved and modified in the course of practical application. Operation instructions can be analyzed and accessed using this decision tree. The objective is to prevent the accidents caused by false instructions, and hence the safety and security of a ship navigating on the sea will be improved.

Introduction

With the development of the world economy, the volume of marine trade has proliferated in recent years. At the same time, the number of ships has been increasing year by year, and cargo tonnage and speed have increased as well. Consequently, navigation accidents began to occur frequently. Although ships have been equipped with many kinds of modern facilities, such as ARPA radar, AIS, satellite navigator, electronic charts, etc., shipwrecks still occurred. According to a report from IMO, 80% of the accidents were caused by human factors (Han and Meng, 2009). So decreasing the accidents caused by human factors is the key for ensuring the safety of navigation.

In the course of navigation, the officer of a ship has to concentrate his mind on

watching the ambient sea area and observing a mass of data, such as the data of navigation environment, attributes of the ship, working condition of equipment, etc. Once the data changed, after analyzing all the data synthetically, the officer should make an accurate decision in time on the basis of learning the navigation regulations by heart and comprehending them sufficiently. And hence, the final result of the decision is entirely affected by the quality of an officer and the condition of his body and mind.

Many researchers have applied themselves to study how to decrease the accidents caused by human beings. Most of the studies focus on the field of enhancing the quality of crew and ameliorating navigation regulations. Although these research findings have been implemented in practice, false decisions still cannot be avoided entirely.

In order to solve this problem, the theory of machine learning is applied in this paper. The instructions of handling a ship issued from the officer on a ship will be assessed by a decision tree, then the accurate instructions will be sent to actuators, and the false instructions will be held back, at the same time an alarm signal will be sounded to remind officers.

Inductive-based Machine Learning

In the course of navigation, once a ship encounters a storm or emergency, the officer should make a decision as soon as possible according to considerable data and regulations. It is boring work and consequently, mistakes are made.

As computer technology progresses, it is no longer hard work for a computer to retrieve, analyze, and manage a mass of multifarious data. In recent years, improvement in the machine learning theory provides a fresh approach for this kind of problem.

Machine learning is a learning way simulating humanity. In 1983, Simon gave the definition of machine learning. “Learning can make a system perform well than it did last time for the same tasks or different tasks of the same quantity (Feng , 2006).” Afterward, scholars generalized the definition further “Machine learning is an intelligent process of acquiring knowledge aiming at certain object, in the inner of a system it exhibits the foundation of fresh knowledge structure and the improvement of it, in the outer of a system it exhibits the improvement and enhancement of the performance (Feng, 2006).”

Induction learning is one of the widespread methods in machine learning because it is close to the learning process of humans. Induction learning is a learning method which derives general rules from some specific examples (Michalski et al., 1986). It is an important means for humans to acquire new knowledge. The inductive-based machine learning focuses on discovering rules from the past experiences, and even further acquiring some related facts to prove the rules for a purpose. All these

experiences are accumulated gradually in the course of practice for many years. They are comprised of working recorders of a system itself and the working cases of experts (Luo and Wang, 1994).

ID3 algorithm is extensively used in the field of inductive-based machine learning. ID3 algorithm is introduced by Quinlan in 1986. It is a decision tree classifying algorithm based on information entropy (Quinlan, 1995). The principle of ID3 algorithm is to determine which factor is easy to make decision and to classify the data into several subsets according to this factor. Next, choose a factor that is easy to make decision in each subset and make that classification. Until only one kind of data is contained in all subsets, classification can be stopped. In this paper, the ID3 algorithm is adopted to assess the instruction of handling a ship.

The Principle of the ID3 Algorithm

The core theoretics of the ID3 algorithm is based on the principle of information entropy, choosing an attribute whose information entropy is the minimum as an attribute for classifying, then extending the branches of a decision tree recursively until the decision tree is created. The entropy of an attribute is defined as the sum of the weighted entropy of a single attribute value.

Suppose that $E = F_1 \times F_2 \times \cdots \times F_n$ is n-dimensional finite vector space, in which F_i, $i = 1, 2, \cdots, n$ are sets of finite discrete symbols. The elements $e = (v_1, v_2, \cdots, v_n)$ in E are called samples, where $v_i \in F_i, i = 1, 2, \cdots, n$.

On the assumption that in the vector space E, the size of positive examples set is p and the size of counterexamples set is n, ID3 algorithm is based on the following two hypotheses: (1) An accurate decision tree in the vector space, E, has a consistent classification probability for any samples with the probability of positive examples and counterexamples in E. (2) The amount of information which is needed by a decision tree to make an accurate classification judgment for a sample is:

$$I(p,n) = -\frac{p}{p+n}\log_2\frac{p}{p+n} - \frac{n}{p+n}\log_2\frac{n}{p+n} \tag{1}$$

If the attribute, A is as the root, which has m of values $\{u_1, u_2, \cdots, u_m\}$, it can divide E into m of subsets $\{E_1, E_2, \cdots, E_m\}$. Suppose that there are p_i of positive examples and n_i of counterexamples in E_i, the information entropy of subset E_i is $I(p_i, n_i)$, and the information entropy of attribute A is:

$$E(A) = \sum_{i=1}^{m} \frac{p_i + n_i}{p+n} I(p_i + n_i) \tag{2}$$

The information gain of root A is:

$$gain(A) = I(p,n) - E(A) \tag{3}$$

ID3 algorithm chooses the attribute A^* which makes the $gain(A)$ maximum as the root node. Then invoke the above procedure recursively for m of subsets in E_i that correspond to different values in A^*, and create sub-nodes $B_1, B_2, \cdots, B_m$ of A^*. Since $l(p, n)$ is a definite value for each node, the entropy $E(A)$ of attribute A can be selected as a comparison standard among nodes.

ID3 algorithm is a greedy algorithm. It creates a decision tree by means of recursion which is from top to bottom, dividing and conquering. It is a famous decision tree induction algorithm.

Application of ID3 Algorithm in Assessment of Ship Handling

Ship handling refers to the fact that seafarers control a ship by maneuvering the navigating control equipment on a ship (Hong, 2008). The work contains:

(1) Keep the course: for the sake of navigating safely and economically, seafarers have to keep a ship navigating on a scheduled course.

(2) Change the course: when a ship meets an obstacle or other ships, seafarers should change the course of their ship in time to prevent crashing.

(3) Change the speed: while berthing or anchoring a ship in a port, seafarers should change the speed of their ship in time.

When a ship is navigating in open water (no obstacles or other ships), in good weather, seafarers should only keep their ship navigating at an economical speed and on a scheduled course. Otherwise, in emergencies (such as an encounter with other ships or storms), seafarers have to change the course or speed of their ship depending upon the actual situation. In this paper we mainly do research in the safety of ship handling in severe weather (in a storm).

The Factors Influencing the Safety of a Ship Navigating in a Storm. In a storm, the factors which influence the safety of a ship mainly include: environmental factors, the attribute factors of a ship itself, and the factors of navigating condition. Factors of environment are comprised of wind direction, wind speed, wave height, wave period, etc. The attribute factors of a ship contain length, beam, draught, rolling period, pitching period, heaving period, etc. The factors of navigating condition involve speed, course, working condition of equipments in cabin, etc. Thereby, attribute sets are built as below:
Environmental attributes: {wind direction, wind speed, wave height, wave length, wave period}.
Ship attributes: {length, beam, draught, rolling period, pitching period, heaving period}.
Navigation attributes: {speed, course}.

Training Samples. Based on the opinions of specialists, rules of ship handling, and the international regulations of preventing a collision, we can sum up some training samples. These samples are the input, and the output is the assessment result that is true or false. Six representative training samples are illustrated in Table 1. It is obvious that six samples are not enough for building up a believable decision tree. In fact, an abundance of samples is needed for creating a perfect decision tree. So we gathered 40 samples (omit) for training.

Creating a decision tree . Among the training samples, the number of positive examples and counterexamples are 20 respectively. According to the ID3 algorithm, $P(p) = P(n) = \frac{20}{40}$, and hence information entropy of the assessment result is:

$$I(p,n) = -\frac{20}{40}\log_2\frac{20}{40} - \frac{20}{40}\log_2\frac{20}{40} = 1 \tag{4}$$

Table 1. Data of training samples.

Sample	1*	2	3	4	5	6
Wind power	8	10	7	9	5	6
Wind direction	Northwest	Southeast	Northwest	North	Northeast	Southwest
Wave height (m)	5	6	4	5	2	3
Length (M)	126	71	127.8	118.6	277	186.2
Beam (M)	20	11	20.5	18.2	32.3	26.7
Draught(M)	4.81	2.65	8.04	8	11	8.1
Speed (Kn)	12	5	12	12	10	8
Course (°)	58	213	37	86	39	26
Instruction	Speed down (10 kn), starboard the helm, course 220°	Speed up (10 kn), starboard the helm, course 300°	Speed down (10 kn), Keeping the course	Speed down (10 kn), Port the helm, Course 10°	Keep the course and speed	Speed down (4 kn), Port the helm, Course 230°
Assessment result	False	True	False	True	True	False

*Zhang, 2004.

In this paper, the attribute of wind power is taken as the root of a decision tree. The values of wind power are 5, 6, 7, 8, 9, and 10. And the number of samples for each of them is 5, 6, 8, 7, 8, and 6 respectively. Among the five samples for the wind power of 5, three samples are positive example, and two samples are counterexample. Among the six samples for the wind power of 6, three are positive example, and three are counterexample. Among the eight samples for the wind power of 7, two are positive example, and six are counterexample. Among the seven samples for the wind power of 8, four are positive example, and three are counterexample. Among the eight samples for the wind power of 9, five are positive example, and three are counterexample. Among the six samples for the wind power of 10, three are positive example, and three are counterexample. According to Eq. (2), the entropy of wind power is worked out:

$$E(\text{wind power}) = \sum_{i=5}^{10} \frac{p_i + n_i}{p+n} I(p_i + n_i) = 1.8937 \tag{5}$$

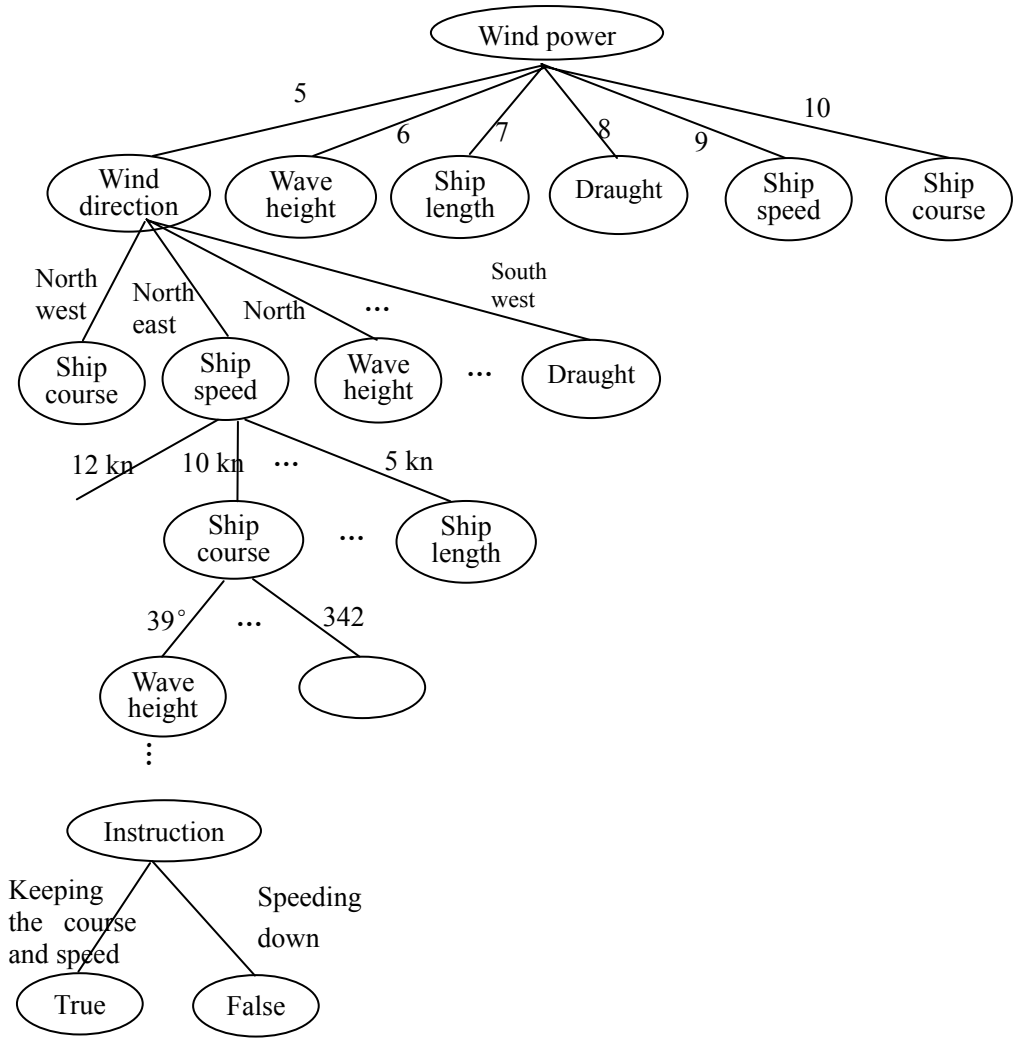


Figure 1. Decision tree of instruction assessment in ship handling.

Similarly, the entropies of other attributes can be figured out, and then the information gain of every attribute will be worked out. The attribute whose gain is the maximum will be chosen as the classification attribute, and a decision tree will be built up recursively. The decision tree of ship handling instruction assessment, which applies ID3 algorithm, is illustrated in Figure 1.

Conclusions

For the sake of decreasing severe ship accidents caused by human factors, ensuring the safety and security of navigation, the factors which have affected the operation instructions are summed up in this paper. A modern induction algorithm in machine learning - ID3 algorithm is applied to create a ship handling assessment decision tree. The instructions sent by an officer will be assessed by this decision tree, so that wrong instructions will be distinguished and an alarm signal will sound as soon as possible so that a severe accident will be avoided. This decision tree will be modified and improved in practice. Only basic studies have been conducted in this paper, and in-depth research will continue in the future.

References

Feng, D. (2006). *Neural Network Expert System*, Science Press.

Han, X.J., and Meng, X.Y. (2009). *Study on Ships Safety Control System*, Marine Navigation and Safety of Sea Transportation.

Hong, B.G. (2008). S*hip Handling*, Dalian Maritime University Press.

Luo, Y. and Wang, K.H. (1992). "An implementation study on inductive-based machines learning method." *Computer Science*, 12(2).

Michalski, R. S., Mosetic, I., and Hong, J. R. (1986). "The multi-purpose incremental learning system AQ15 and its testing application to three medical domains." *Proc. AAA,* I86.

Quinlan, J. R. (1993). "C4.5: programs for machine learning." *Morgan Kauffman*, 2(3), 89-94.

Zhang, J. P. (2004). *Navigation Practice Manipulation Assessment*, Dalian Maritime University Press.

Travel Mode Choice Modeling Based on Improved Probabilistic Neural Network

Dan Zhao, Chunfu Shao, Juan Li, Chunjiao Dong and Yinhong Liu

[1]MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University; P.O. Box 100044, Beijing, China; PH(86) 10-51683665, Email: zhaodans@yahoo.cn, cfshao@center.bjtu.edu.cn

Abstract

Travel mode choice can be regarded essentially as a problem of pattern recognition. In the past, numerous researchers applied artificial neural network (ANN) to travel mode choice modelling. Feed-forward back propagation neural network (BPNN) is widely used, and most of the studies show that BPNN presents better prediction accuracy and generalization capacity. Probabilistic neural network (PNN) is also known as a general solution to pattern classification problems by following an approach developed in statistics. Comparing to BPNN, its shorter training time and better stability increase the reliability of simulation results, even so PNN is barely used in travel behaviour analysis. Therefore the paper tries to apply PNN for travel mode choice modelling. First, network structure is established based on the data obtained from resident trip survey, and then the K-means cluster algorithm is applied to optimize the hidden node number so that they can be modified dynamically. Moreover, the data centers and extended constants of gaussian radial basis function are changed adaptively during the learning progress, and the improved PNN is called KPNN. To reveal the superiority of KPNN, BPNN is also introduced to make a comparison. For BPNN, the optimal number of hidden nodes is recognized, and then the well trained network is applied. By proposing two performance measures, the classification and predictive capability of KPNN and BPNN is evaluated, it is found that KPNN outperforms BPNN on both the simplicity and prediction accuracy. Anyway, the simulation results show that KPNN can deal with the problem of travel choice modelling as well as, if not better than BPNN.

Key Words: travel mode choice; artificial neural network; k-means cluster; probabilistic neural network; BP neural network

Introduction

Discrete choice models based on random utility maximization rules are conventional approaches used to deal with travel behaviour analysis, especially disaggregate choices, such as travel mode, destination, and route choice, and it has received much attention before artificial intelligent algorithm is widely used in the field of transportation. Artificial neural network (ANN) is the important component of intelligent algorithm which is famous for its powerful information processing ability. Since ANN is imported as an alternative to analyze travel related behaviour, the advantages in pattern recognition and prediction become more and more obvious. Essentially, travel mode choice modeling is a pattern recognition problem in which multiple human behaviour patterns reflected from explanatory variables determine the choices between alternatives and classes. It tallies with the claim that ANN can tackle the problems of travel demand modeling and forecasting as well as, if not better than, the discrete choice approach (Dougherty 1995), and numerous literatures have also manifested that neural network shows high flexibility and adaptability as well as a superior prediction performance (Xie and Parkany 2003; Sayed and Razavi 2000; Hensher and Ton 2000). Referring to the computing principles, neural network modelling does not require an algorithm to compute a specific output for a preset functional form (Faghri and Hua1992), they are composed of a number of neurons that act as processing elements, and data processing procedures in the network structure are unknown. In terms of network structure and activation function, neural network can be classified into several types, such as follows: feed-forward back-propagation neural network (BPNN), radial basis function neural network (RBFNN), feedback neural network (FNN), and self-organizing competitive network (SOMN). Each of them has its own characteristics and applicability. For different sources of data, the optimal network structure is distinct, but first of all, a proper type of network should be identified. In most of the cited papers, BPNN has been widely and maturely used for travel mode choice modelling, but for some drawbacks of BPNN, other kinds of neural network should be employed to broaden the applicability of neural network, so in this paper probabilistic neural network (PNN) is introduced as an alternative to try to solve the same problem. PNN based on the theory of statistics is a transformation of RBFNN, which is equivalent or similar to the optimal classifier-Bayesian classifier in function. Its completely forward algorithm becomes the most important difference from BPNN, thus PNN consumes shorter training time and avoids converging easily to local minima. Therefore, PNN can be regarded as a proper approach to settle pattern recognition problems.

The article consists of six sections. In section 2, an overview of the applications of neural network to travel mode choice analysis is summarized. Section3 presents the empirical data of the resident trip survey followed by a description of common variables and data sets selected for training and testing neural network. In Section 4, concepts of PNN are first specified, and the K-means algorithm is applied to optimize neural network structure and performance (the improved PNN is called KPNN), finally the simulation results are presented, and the optimal parameters (including weight, threshold) are identified through the data illustrated in the second

section. In order to reveal the superior of KPNN, BPNN is established and calibrated to make a comparison in section 5. In section 6, to evaluate the classification and predictive capability of KPNN and BPNN, two performance measures are proposed and calculated. Section 7 lists the predictive performance of KPNN and concludes with comments on the merits of neural networks.

Literature Review

Compared to the neural network, discrete choice models based on the random utility maximization rule is not flexible enough, such as explanatory variables can not completely describe the factors that influence travel mode choice, and the assumption of independent and identical distributed errors in the model is not always true. Neural network provides the capability without the complication of having to formalize all the complex causal variables and relationships which other models require (Shmueli et al. 1996). Adding neural networks as a new modelling methodology stems from its apparent relevance to problems requiring large scale, highly dimensional, data analysis, such as travel-related behaviour. Paliwal and Kumar (2009) carry out a comprehensive review of articles that involve a comparative study of feed forward neural networks and statistical techniques used for prediction and classification problems in various areas of applications. Statistical model clearly stands out as it allows interpretation of coefficients of the individual variables, while neural networks can automatically approximate any nonlinear mathematical function that is particularly useful when the relationship between the variables is unknown or is complex, and hence it is difficult to handle statistically.

The previous relevant studies that applied neural network to travel mode choice modeling can be divided into three classes. The first one is applying feed-forward network models to travel mode choice modeling and making a contrast with the closed-form discrete choice models based on random utility theory aiming to illustrate the superiority of the neural network. Hensher and Ton (2000) employed three datasets obtained from a stated preference survey to calibrate both BPNN and a nested-logit (NL) model to analyze travel mode choice behaviour of commuters. The comparison results were reported that BPNN performs better. Xie and Parkany (2003) applied data mining to work travel mode choice modeling, and the prediction results show that the two data mining models, including BPNN, offer comparable but slightly better performance than multinomial logit (MNL), and BPNN model gives a superior prediction performance in most cases. Cantarella and Luca (2005) have also applied multilayer feed-forward networks (MLFFNs) to two real case studies focusing on medium distance intercity journeys and make a comparison with three discrete choice models-MNL, NL and CNL (cross-nested logit). The obtained results show that MLFFNs may outperform discrete choice models when the values of mode shares are quite similar and can be a feasible tool for travel demand analysis. The second one is method combination of various theories and neural network. Besaw (2010) developed a recurrent artificial neural network to replicate nonlinear spatio-temporal discrete choice patterns produced by a spatially-explicit agent-based model with social influence to describe consumer travel mode choice. Vythoulkas

Table 1. Mode split proportions in the datasets.

Mode	Total datasets		Training sets		Testing sets	
	Quantity	Percent(%)	Quantity	Percent(%)	Quantity	Percent(%)
Railway	133	13.75	11	14.18	3	15.00
Bus	706	73.01	58	72.13	14	70.00
Car	128	13.24	11	13.69	3	15.00
Sum	967	100.00	80	100.00	20	100.00

and Koutsopoulos (2003) introduced fuzzy sets and fuzzy control to simulate the related vague perceptions of the various attributes to the travel mode choice preferences towards the available alternatives of travellers This approach is extended by incorporating rule weights, which capture the importance of a particular rule in the decision process and presents an approach for calibrating the weights using concepts from neural networks (called neural-fuzzy framework).The result indicates that the neural-fuzzy framework has the potential to model general discrete choice problems. Xianyu and Juan (2008) construct a feed-forward network based on rules from a decision tree generated by C4.5 method, called knowledge-based neural network (KBNN), the simulation results show that decision tree and neural network models offer slightly better performance and higher precision than discrete choice model and standard BPNN. The third one is improving network structure or training (learning) algorithm so as to enhance network performance. Celikoglu (2006) employs a radial basis function and generalized regression neural network to propose a calibration process for travel mode choice analysis, but instead of application directly calibrates models, and both of them are used as a sub-process for alternative non-linear model specifications on utility function.

Empirical Data

The data used for implementing neural network models comes from resident trip survey along a certain suburban railway line in Beijing (2009). From nine traffic zones, a total 1100 questionnaires were sent out, and 967 valid ones were retrieved. The transportation mode structure in the investigative transport corridor is comprised of a suburban railway, bus, and car, which determines the number of output nodes is three, among them railway and car capture relatively less market share, about 13% each. The content of investigation covers respondents' travel information, which describes the characteristics of both traffic mode and traveller. After correlation analysis, four of them are chosen as the input variables of neural network, which is, respectively, cost, time, car ownership, and income, and it can be confirmed that neural network has four input nodes.

Because of numerous duplications in the total datasets, 100 samples are selected randomly without duplications in order to avoid over fitting. Furthermore, the 100 samples are split randomly into two datasets, and 3/4 of them are partitioned into training sets, the others are allocated to test sets. The three travel modes share approximate proportions in training and testing sets with the total datasets, which is shown in Table 1.

Probabilistic Neural Network

***The Architecture of PNN*.** PNN is an important transformation of RBFNN. It is also known as belief networks, Bayesian networks, and knowledge maps. When it is applied to deal with classification problems, it has similar functions as the optimal Bayesian classifier.

Generally, PNN is established with three-layer forward networks. And the whole processing can be divided into three modules corresponding to each layer. The first layer-input layer offers input vectors to the hidden layer; the second layer-hidden layer is actually RBF layer, it first reckons the distance from input vectors to weighted vectors, which is called distance vector and expressed as $\|dist\|$, whose elements indicate how close the input is to a weighted input. Then it uses a gaussian function as a transfer function to deliver the product of distance vector and threshold to the next layer; the third layer-output layer is also called a competitive layer. It calculates the probability of the occurrence of each pattern to form a vector of probabilities, and then the maximum of these probabilities are picked out and produces 1 for one category and 0 for the rest. Figure 1 described the architecture of PNN.

In Figure 1, R, Q, and K represent the number of input node, hidden nodes, and output nodes, respectively. As mentioned above, $\|dist\|$ indicates the distance from input vectors P to weighted vectors w_1, and n_1 is the product of $\|dist\|$ and threshold b_1, also represents the input vectors of the hidden layer. (·*) is an arithmetic unit, and the calculation formula is expressed as Eq.1:

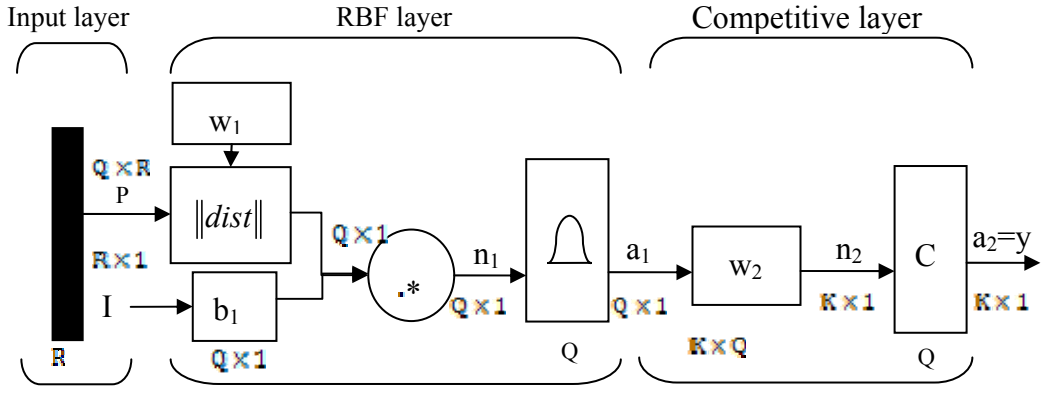


Figure 1. The architecture of PNN.

$$n_1 = \|dist\| \times b_1 = \sqrt{\sum (w_1 - p)} \times b_1 \tag{1}$$

After transfer by a gaussian function, the output of hidden layer a_1 is acquired, just as Eq. 2 shows. It can be deduced that as the distance from input vectors to weighted vectors decreases, the output value increases.

$$a_1 = Radbas(n_1) = \exp[-{(n_1)^2}/{d}] \tag{2}$$

When a_1 is presented to the competitive layer, the other weighted vectors w_2 are set to compute vector n_2 with a linear weighted summation method (as expressed in Eq. 3), n_2 can be regarded as the estimated probability density function of three travel modes. Each component of n_2 corresponds with a pattern category, and it sums the elements of a_1 due to each of the K input classes.

$$n_2 = \sum a_1 w_2 \tag{3}$$

Module C is a competitive transfer function, which functions as a classifier to find the maximum of n_2 and produces a_2 as 1 for the corresponding category and 0 s elsewhere, so target vector has K elements and composed of 1 and 0, and each input vector is associated with one of K classes. The operation rule is shown in Eq. 4.

$$a_{2j} = \begin{cases} 1 & n_{2j} = \max(n_{2_{1j}}, n_{2_{2j}}, \cdots n_{2_{kj}}) \\ 0 & n_{2j} \neq \max(n_{2_{1j}}, n_{2_{2j}}, \cdots n_{2_{kj}}) \end{cases} \quad i = 1, 2, \cdots k \tag{4}$$

Thus, the network classifies the input vector into a specific K class because that class has the maximum probability of being correct (Araghi et al. 2009).

Improved algorithm. The hidden nodes number of ordinary PNN is the same as the quantity of training samples, as sample size grows larger, the architecture of PNN will become more complicated, resulting in a longer training time, so simplifying training sets is necessary for network optimization. The paper adopts K-means cluster algorithm to optimize the hidden node number of PNN, and the data centers and extended constants (width vector) of gaussian radial basis function are changed adaptively during the learning process (Guo et al. 2009), and the improved PNN is called KPNN. The K-means cluster algorithm assigns each point to the cluster whose center is nearest. The center is the average of all the points in the cluster, that is, its coordinates are the arithmetic mean for each dimension separately over all the points in the cluster. It must be noticed that class centers correspond to the weight vector w_1 and the extended constants equivalent to vector d in Eq. 2, which also named as spread, it updates dynamicly as $\|dist\|$ changed when the class centers w_1 are modified. The K-means cluster algorithm is clarified through five steps:

(1).Initial the k class centers $w_{1,0}$ of training sample sets;

(2).Calculate the distance vector $\|dist\|$ between training samples and class centers to afresh the former cluster;

(3).Calibrate the mean value of the new k classes and set it as the new centre $w_{1,1}$, then calculate the distance again;

(4).Judge whether $w_{1,1}$ equals to $w_{1,0}$; and

(5).Repeat step (2) to (4) until $w_{1,k}$ equals to $w_{1,k-1}$.

K-means cluster algorithm functions only in the input and hidden layers. By integrating the competitive module of ordinary PNN, the calculation flow chart of KPNN is described in Figure 2.

Simulation results. The architecture used in this study is comprised of three layers with four input nodes and three output nodes, and the number of hidden nodes is the number of clusters, that is three. The input vectors contain four explanatory variables, each input vectors represents an observation. And the output layer offers numerical mode choice results, and each output vector fits into one of the three mode choice, $[1,0,0]^T$, $[0,1,0]^{T,}$ and $[0,0,1]^T$ represents bus, car, and suburban-railway, respectively. After network is established, the training process is also completed at the same time, which is the most prominent advantage of PNN.

The first advantages of optimization algorithm is the reduction from 100 to 3 of hidden nodes number, the second advantage is calculation accuracy improvement, the errors of hidden nodes output decrease, and maximum probabilities of the three classes obtained from competitive layer can be distinguished easily. For PNN, when the extended constants (spread) is 0.4 the best classification is achieved, while in KPNN, the spread is 0.533, and the larger spread means a more accurate neural network.

BP Neural Network

In order to reveal the superior of PNN, BPNN is also introduced to make a comparison. The principle of BPNN is well known to us so that it will not be discussed in this paper. But the crucial procedure that influences the performance of network has to be clarified here, which is the identification of the number of hidden nodes and the training algorithm. In this chapter, a three-layer network with the same input and output nodes with PNN is established, the optimal number of hidden nodes is recognized as shown in Table 2 and Figure 2, thus it can be learned that the number 12 is what we want since the training error and mean square error is minimum. Besides, 'tansig' and 'logsig' are used as transfer function of hidden layer and output layer, respectively.

The training process of BPNN lasted for 604 epochs, and each epoch is presented with the input vectors in random order, then converging to the target error 0.005. Adaptive learning rate with Momentum (traingdx) is adopted for faster training. After training, network is tested.

Table 2. Performance of BPNN with different hidden nodes.

Number of hidden node	7	8	9	10	11	12
Training error	0.1669	0.1166	0.1455	0.982	0.1141	0.0742
Mse（$*10^{-4}$）	5.21	2.88	4.92	178	3.27	1.08
convergence epochs	834	625	521	529	523	604

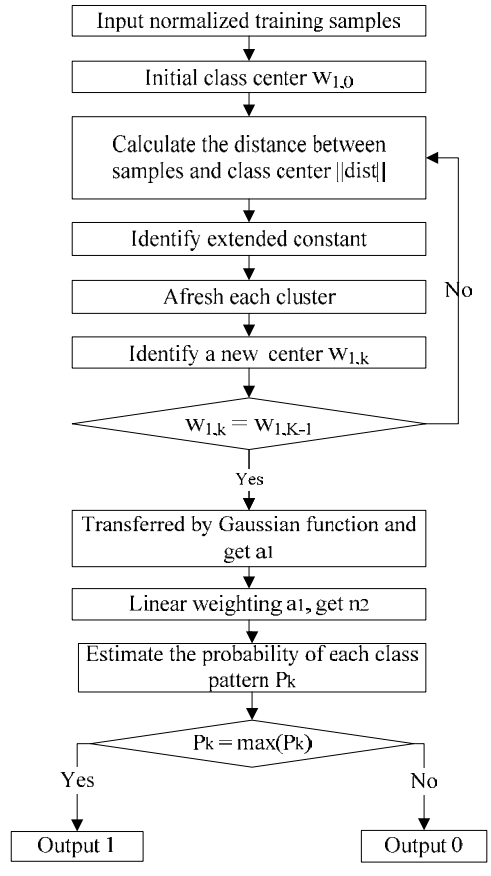


Figure 2. The calculation flow chart of KPNN.

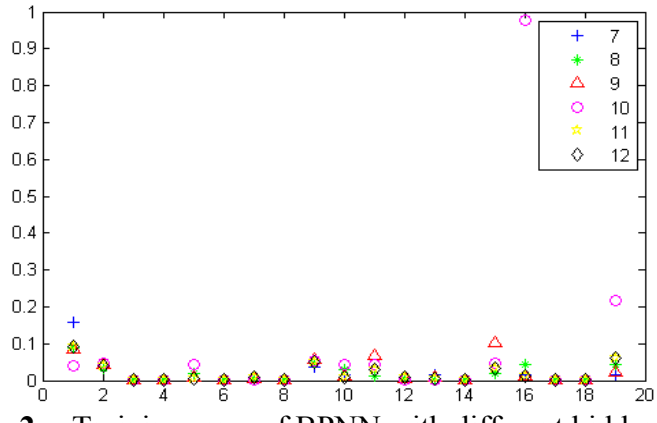


Figure 2. Training error of BPNN with different hidden nodes.

Evaluation

In the paper, two performance measures are defined to evaluate the prediction accuracy of PNN and BPNN. They reflect the prediction accuracy of the two networks from individual and group, which are named as individual hit ratio (H_{ij}) and group hit ratio (H_{gj}), respectively, the computing equation are expressed as follows:

$$H_{ij} = N_{pj} / N_{aj} \qquad (j = 1,2,3) \tag{5}$$

$$H_{ij} = N_{kj} / N_{aj} \qquad (j = 1,2,3) \tag{6}$$

N_{pj} represents the number of observations for mode j, which is predicted accurately, and N_{aj} is the number of observations, which actually choose mode j. N_{kj} indicates the number of observations whose prediction results are mode j.
H_{ij} and H_{gj}, respectively, reflect the micro and macro prediction accuracy of the network, the former is always less than 1 while the latter is indefinite. The calculation results are shown in Table 3.

It can be seen from Table 3 that KPNN presents a greater precision than BPNN, and BPNN fails to distinguish bus and railway exactly. The explanation for this maybe the travellers who choose railway and bus share some common characteristics.

Conclusions

In this paper, KPNN proves to be valid in structure simplification and excellent in training time reduction as well as prediction accuracy. This conclusion paves the way for its widely use for travel mode choice modelling in the future. First, KPNN optimizes the hidden nodes number and simplifies the structure, which leads to a shorter prediction time. Second, by integrating K-means cluster algorithm to ordinary PNN, the weight vector of hidden layer (class centers) and extended constants (spread) can be identified in a dynamic process. Comparing to PNN and BPNN, prediction accuracy is improved, the errors of both hidden layer and competitive layer outputs decrease, which can be seen from the calculation results of the two measures.

The shortage of the paper lies in the lack of generalized testing, but the simulation results show that KPNN can deal with the problem of travel choice modelling as good, if not better than BPNN. What is more, ANN is still unable to substitute discrete choice models, because the latter can analyze and quantize influence factor of travel mode choice by explanatory variables easily, but the calculation process of neural network is a “black box” actually, what can be obtained is prediction results, whereas neural network operate directly on data, thus avoiding a prior assumption of a discrete choice model with many consequential constraints. Its prediction results have proved to be more accurate.

Table 3. Comparison of prediction accuracy based on KPNN and BPNN.

	KPNN		BPNN	
	H_{ij} (%)	H_{gj} (%)	H_{ij} (%)	H_{gj} (%)
Bus	100	100	81.25	107.14
Car	100	100	100	100
Railway	100	100	66.67	66.7

Acknowledgments

This research is funded by National Basic Research Program of China (2006CB705505), National Natural Science Foundation of China (50778015), and National Science and Technology Supporting Project (2006BCG01A01).The author appreciates the assistance of Professor Shao in the modifications and suggestions and of this paper.

References

Araghi, F. L., Khaloozade, H., and Arvan, M. R. (2009). "Ship identification using probabilistic neural networks." *Proceedings of the International MultiConference of Engineers and Computer Scientists*, Hongkong, 1291-1294.

Besaw, L. E., Rizzo, D. M., Eppstein, M. J., Grover, D. K., and Marshall, J. S. (2010). "Up-scaling agent-based discrete choice transportation models using artificial neural networks." *Transportation Research Board*, CD-ROM.

Cantarella, G. E., and Luca, S. D. (2005). "Multilayer feed forward networks for transportation mode choice analysis: An analysis and a comparison with random utility models." *Transportation Research Part C*, 121-155.

Celikoglu, H. B. (2006). "Application of radial basis function and generalized regression neural networks in non-linear utility function specification for travel mode choice modeling." *Mathematical and Computer Modelling*, 640-658.

Dougherty, M. (1995). "A review of neural networks applied to transport." *Transportation Research Part C*, 247-260.

Faghri, A., and Hua, J. (1992). "Evaluation of artificial neural network applications in transportation engineering." *Transportation Research Record*, 71-80.

Guo, G. Q., Cao C. T., and Xu, J. M. (2009). "Traffic condition recognition based on probability neural network." *Computer Engineering and Applications*, 214-216.

Hensher, D. A., and Ton, T. T. (2000). "A comparison of the predictive potential of artificial neural networks and nested logit models for commuter mode choice." *Transportation Research Part E*, 155-172.

Ji, D. C., and Song, B. F. (2009). "Equipment fault diagnosis based on probabilistic neural networks and simulation analysis." *Fire Control & Command Control*, 82-85.

Jia, J. G., and Yang, D. Y. (1999). "Modelling traffic modal split with probabilistic neural network." *J. Univcrsity of Shanghai for Science and Technology*, 9-11.

Paliwal, M., and Kumar, U. A. (2009). "Neural networks and statistical techniques: A

review of applications." *Expert Systems with Applications*, 2-17.

Sayed, T., and Razavi, A. (2000). "Comparison of neural and conventional approaches to mode choice analysis." *Journal of Computing in Civil Engineering*, 23-30.

Shmueli, D., Salomon, I., and Shefer, D. (1996). "Neural network analysis of travel behavior: evaluating tools for prediction." *Transportation Research Part C*, 151-166.

Vythoulkas, P. C., and Koutsopoulos, H. N. (2003). "Modeling discrete choice behavior using concepts from fuzzy set theory, approximate reasoning and neural networks." *Transportation Research Part C*, 51-73.

Xianyu J. C., and Juan Z. C. (2008). "Knowledge-based neural networks and its application in discrete choice analysis." *IEEE International Conference on Networked Computing and Advanced Information Management*, Tianjin, 491-496.

Xie, C., Lu, J., and Parkany, E. (2003). "Work travel mode choice modeling using data mining: decision trees and neural networks." *Transportation Research Record*, 50-61.

Complex Networks Model for Urban Passenger Transport Systems: A Case Study of Harbin Taxi

Jian Wang [1], Xiaowei Hu and Chong Wei

[1] Prof., School of Transportation Science and Engineering, Harbin Institute of Technology, No. 73, Huanghe River Road, Nangang District, Harbin, Heilongjiang Province, PRC, 150090; Email: wang_jian@hit.edu.cn.

Abstract

The urban passenger transport system plays a key role in providing services for passengers. By combining a science system with complex network theory, we analyze the relationship among various actors of the urban passenger transport system, which is a network model consisting of managers, companies, passengers, and their interactions. The structural and dynamic property of the urban passenger transport system is described. Through analyzing the network's nodes and structural edges, the network dynamic evolution model can be established. The Harbin taxi passenger service market system was studied as an example to simulate and validate a proposed model, which reflects the objective laws of the urban passenger transport system and provides a new way to investigate the urban passenger transport system.
Key Words: complex networks; urban passenger transport system; structural properties; evolving networks

Introduction

Complex network theory starts with the most basic 'link' of the complex systems, abstracts the complex system into a network of nodes, and explores the systems' nature from the network topology and network dynamics. Recently, complex network's research scope has extended from the social networks, Internet, and genetic networks to the transportation field (Wu et al., 2007; Boccaletti et al., 2006).

As a complex and open social system, the urban passenger transport system has complex features, such as multi-disciplinary, cross fields, and large scale. Complex network theory is a new system research tool used to explain this complex system. The structure of this paper is: section 1 reviews the related literature on the complex science of transportation and taxi's study, section 2 analyzes the structure of the urban

passenger transport system, section 3 establishes the network dynamic evolution model, section 4 simulates the Harbin taxi passenger service market system, and the last section 5 concludes this paper and examines possible future research.

Literature Review

The research of complex science on the urban passenger transport network has achieved more outcomes. Su et al. (2007) adopted a game theory to simulate the urban public transport network evolution and found an existing game behavior among the passengers, bus companies, and government traffic management agencies. Ying (2007) studied the complex characteristics of the transport network optimization pricing principle and calculation method. Wang et al. (2006) adopted the parallel systems in the complexity of scientific research to propose the urban passenger transportation system status management methods and established a system state evaluation mode.

Taxis are a significant form of urban passenger transportation, which plays a key role in providing personalized travel services and bridges the gap between private and other public transport (i.e., linking to the bus or subway). Cairns and Liston-Heyes (1996) thought that the taxi price regulation was necessary to produce a balanced system; the second-best pricing only can be achieved under the fees and taxi use under controlled conditions. Schaller and Consulting (2005) utilized a regression model to analyze the number of U.S. cities that have a taxicab issue, involving variables such as population, employment, vehicle ownership, transit use, airport passenger volumes, and taxi fares. Gwilliam (2008) found that taxi fares are more inclined to increase in the taxicab market liberalization process. As illustrated by Morrison (1997) in Wellington's (New Zealand) findings, taxicab market liberalization will lead to the company's merger and smaller operators' closure.

While the research of complex science on urban passenger transport network and the taxi passenger transport system has evolved, the present study encounters the following problems, which can not effectively describe the structure and dynamic evolution characteristics of urban passenger transportation system. This article adopts a system analysis method to abstract the urban passenger transport system into management, passenger services companies (public and private), and passengers composed system. Through analysis the relationship among those parts to achieve the urban road passenger transport system's complex network model, and then identifies the static topological structure characteristics and dynamic evolution features, at last through the simulation of Harbin taxi passenger service market system to reflect the urban passenger transport system's laws.

Structural Analysis of the Urban Passenger Transport System

Urban passenger transport system is a complex huge system that covers a large number of participants and requires abstracting the main unit and elements from the system and reflecting the structural relationship between the various elements. Van de

Velde (1999) divided the city public transport's planning and control into a strategic layer, tactical layer, and the execution layer. Ongkittikul and Geerlings (2006) analyzed the business relationship between the public and private transport sector. Roumboutsos and Kapros (2008) divided the urban passenger transport market structure into administrative departments, and public and private companies, by establishing policy and providing concrete service to satisfy the passengers. Therefore, this article abstracts the urban passenger transport system into three main parts: management, passenger services companies (public and private), and passengers.

Assumption 1: management department provides services for the passenger services companies and passengers. This assumption accords with the current status of China's passenger service market. At present the bureau of transportation (in some cities it is the Transportation Committee) manages the city bus, taxi, and subway to constrain the various companies, and passengers choose the transport modes freely from the varieties of public passenger transport mode.

Assumption 2: the impact factors of the urban passenger transport price only consider the oil price, transport supply, and demand and income, regardless of price levels, land use, and the foreign impact.

The management department influences both of the urban passenger system's supply and demand by establishing of policies and measures; passenger services companies (public and private) are the supply side. Its funding source is mainly on the state-owned, private, and financing. Passengers are the demand side, and their mode choices are primarily based on their own economic situation, the understanding of time value, comfort, and so on (Mokhtarian and Chen 2004; Scheiner 2009). Each main part of the urban passenger transport system is affected by the macro-economic environment and the activities in the urban passenger transport price's feedback. Figure 1 shows the details of the various elements of the urban passenger transport system and its relationship with the host-environment interactions.

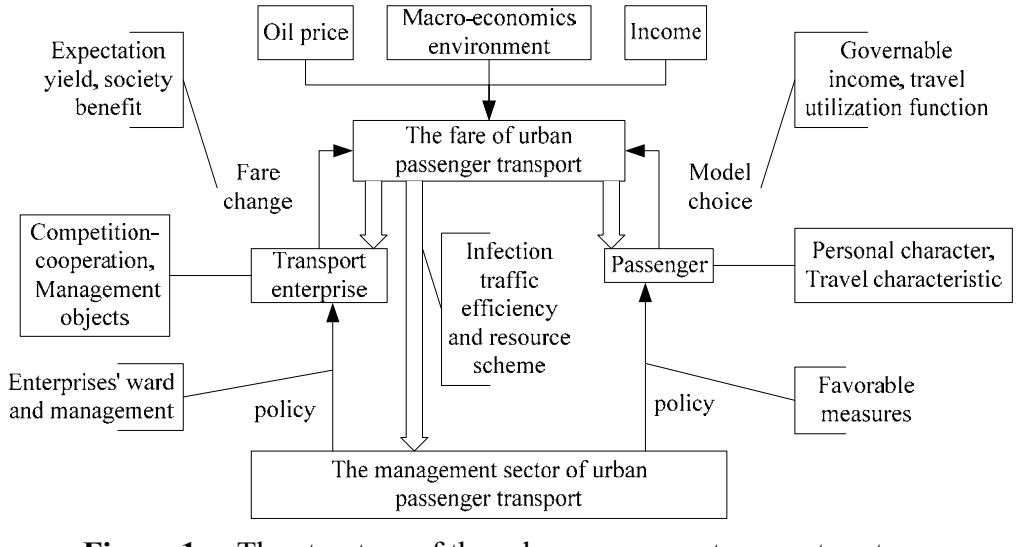


Figure 1. The structure of the urban passenger transport system.

The Complex Network Model of Urban Passenger Transport System

Modeling ideas. The small-world economy complex network treats the socio-economic as the transactions, financing, and production of network to research, (Corso et al. 2003); the nodes on behalf of market players (such as enterprises, factories, consumers, etc.), while edges on behalf of the main activities (such as from enterprises to purchase commodities, from bank financing, etc.), the direction of the edges represents the capital flows' direction. Many of the economic system variables obey the power-law tail distribution, including personal wealth, firm size, etc. (Corso et al. 2003; Axtell 2001; Stanley et al. 1996). The power-law can describe the interaction between a large numbers of elements in a complex system. Based on the research, the authors construct the topology model of the urban passenger transport system, which contains three categories nodes: management, transport enterprises, and passengers. The following is the detailed description of network topology structure.

Network Topology Structural

Network nodes' Determination. Nodes: there are three participants of the urban passenger transportation system; this three-level network is composed of three nodes.

- First level node - Passenger management $A_i\,(i=1,2,3)$, this is accordance with the current status of China's passenger service market. In the domestic passenger market, the public transport management department, taxi passenger transport management department, and the rail traffic management, commission accounted for more than 75 percent of the urban passenger transport market share, so the first level node predigests to three nodes.
- Second level nodes - Passenger provider $B_j\,(j=1,2,\cdots,10^2)$. (Axtell 2001; Stanley 1996; Naldi 2003) studies have shown that the size of corporate assets subject to a discrete Pareto distribution or Zipf rule, the tail of the cumulative distribution function has the following form:

$$\Pr\left[B \geq b_j\right] = \left[b_0 / b_j\right]^{\alpha}, \quad b_n \geq b_j \geq b_0, \quad \alpha > 0 \tag{1}$$

where, b_j is the asset size of the passenger transport providers; b_n is the largest providers' asset size; and b_0 is the smallest providers' asset size. Axtell (2001) found that in 10 years time, accompanying with the number of firms, firm size, number of staff changes, and the value of α is stabile close to 1.

- Third level nodes - Passengers $C_k\left(k=1,2,\cdots,10^6\right)$. The income distribution of the personal characteristics is assumed to obey the power-law distribution of wealth (Corso et al. 2003; Dragulescu and Yakovenko 2001; Coelho et al. 2008), which is subject to

$$T\left(w\right) = w^{-\beta} \tag{2}$$

The results of Corso et al. (2003) have shown that Eq. (2) develops into the following equation, which would be closer to reality.

$$T(w) = w^{-\beta} \exp^{-\theta w} \tag{3}$$

where β, θ is a parameter that varies with the time and location. Dragulescu and Yakovenko (2001) found that $\beta = 1.7 \pm 0.1$ in 1998 in the United States. Due to the data's limit, the statistical data cannot determine the value β, θ, and therefore assumes that the passenger number is 10^6 orders of magnitude to describe the distribution of the node.

Network Edges' Determination. Described as follows: (1) side: On behalf of market transactions, reflecting the passenger transport market's capital flows. (2) edge connectivity rules: Passengers based on accessibility and travel cost to choose the mode, and thus link to the relative management; for passenger transport providers, directly link with the corresponding management sectors. Connection rules are as follows: $A_i B_j$ means that each second node connects with only one first node; $A_i C_k$ means that each third node connects with only one first node; $B_j C_k$ means that each second node connect with only one third node, and preferential attachment to B according to the size.

Network Dynamic Evolution Model

Network Optimization Preferential Attachment Rule. A network node is dynamic and not fixed. As time goes by, the old nodes out, the new node enters, and the relationship between the nodes is re-selecting and connecting. The rule of additional node's connection is simply recorded as new nodes tend to connect to nodes with more edges, which give priority to connection rules that gave way to the k power-law distribution (Corso et al., 2003; Albert and Barabási, 2002), that is

$$p(i) \propto k^{\gamma}, \quad 2 < \gamma < \infty \tag{4}$$

Here, the value γ adjusts according to different economies. The choice proportional of node i is the direct ratio to the probability of penetration $k(i)$, that is

$$p(i,t) = \frac{k(i)}{\sum_j k(j)} \tag{5}$$

Node Life Function. Since the urban passenger transport market is a competitive and open market, new enterprises often enter the market while old enterprises' quit. The second node's change is particularly vital to the network's structure and function. Studies have shown that in a competitive industry, competition among enterprises fails to obey the Pareto distribution probability function (Campbell and Choudhury, 2005):

$$P_a(\rho, \mu) \sim f(x) = \rho \mu^{\rho} x^{-(\rho+1)}, \quad x \geq \mu > 0, \quad \rho \geq 0 \tag{6}$$

where P_a is the enterprise's competition and replacement probability.

Simulation of Urban Passenger Transport System's Complex Network Model

In 2008 there were 101 taxi passenger companies in Harbin, in line with the 10^2 assumption of the passenger provider, the asset size is in the range of assets $[b_0, b_n]$, in accordance with the enterprise scale function generator. The Harbin taxi company's size distribution is shown in Figure 2.

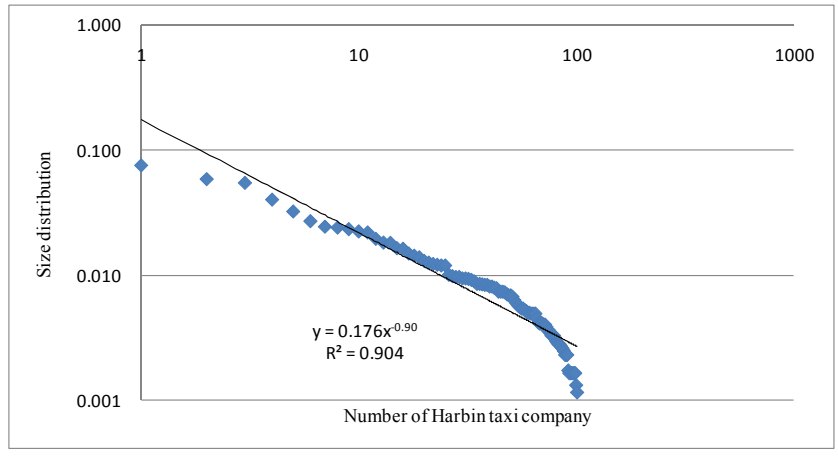


Figure 2. Harbin taxi company's size distribution.

Table 1. Harbin taxi passenger service market data.

Year	The number of taxicab (vehicle)	Increase taxicabs' number (vehicle)	Ratio	Urban population (10^3person)	Increase population (10^3 person)	Ratio
1998	7950	632	1.000000	4335	21	1.000000
1999	8948	1630	2.579114	4369	55	2.619048
2000	10411	3093	4.893987	4380	66	3.142857
2001	11556	4238	6.705696	4409	95	4.523810
2002	11566	4248	6.721519	4492	178	8.476190

This section adopts the taxicabs size and the population of Harbin during 1998 to 2002 to simulate and analyze Harbin taxi passenger service market system. Here the urban population growth data reflects the growth of the number of passengers; using taxicab growth data reflects the increase of taxi passenger transport companies.

The simulation analyzes the relationship between the taxi company's wealth and passengers, and results are as follows:

- From the network aggregation we find the large-scale node has a more linking side, which means the asset size passengers cluster is directly proportional, the larger the taxi passenger transport company, the more passengers will connect with it;

- In recent years, the taxi passenger transport service market in Harbin's experienced dynamic changes. Big enterprises transaction power series increased more than the small enterprises, which fully reflects the enterprises' scale and brand effect;
- Small enterprises have influenced relatively large enterprises by their policy changes, and passengers cluster changes are also relatively large.

For the urban passenger transport manager, it may be based on the current macroeconomic policy and industrial policy environment, by setting the relevant values (such as passengers changes, the number of taxi passenger transport enterprises' change, etc.), to predict the dynamic evolution of the market, so as to control risk and provide the basis for the development of relevant policies.

Conclusions

Using the complex network theory to establish the urban passenger transport system complex network model, and through the simulation of Harbin's taxi passenger service market changes to explain the enterprises scale effect and the small enterprises' eliminate facts. This model reveals the dynamic evolution features between passenger and transport enterprises, reflecting the policy change on taxi service enterprises.

The results suggest that the management department can base its policy on the current macroeconomic policy and industrial policy environment to predict the dynamic evolution of the market. The complex network model of the urban passenger transport market has provided a new measure, which overcomes the previous studies' linear approximation and inadequate research on dynamic nature of the problem.

From the system perspective, this study only accomplishes the establishment of urban passenger transport system's complex network model. To study the macro-economic changes and taxi fares changes, the dynamic characteristics and self-organized evolution of the system will need further analysis.

References

Albert, R., and Barabási, A.L. (2002). "Statistical mechanics of complex networks," *Reviews of Modern Physics,* 74(1), 47-97.

Axtell, R. L. (2001). "Zipf distribution of U.S. firm sizes," *Science,* 293, 1818-1820.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., and Hwang, D. U. (2006). "Complex networks: structure and dynamics," *Physics Reports,* 424, 175-308.

Cairns, R. D., and Liston-Heyes, C. (1996). "Competition and regulation in the taxi industry," *Journal of Public Economics,* 59, 1-15.

Campbell, S. V., and Choudhury, A. H. (2005). "The relationship between interest rates on the number of large and small business failures," *Academy of Accounting and Financial Studies Journal,* 9(3), 29-40.

Coelho, R., Richmond, P., Barry, J., and Hutzler, S. (2008). "Double power laws in income and wealth distributions," *Physica A,* 387, 3847-3851.

Corso, G., Lucena, L. S., and Thomé Z. D. (2003). "The small world of economy: a speculative proposal," *Physica A*, 324, 430-436.

Dragulescu, A., and Yakovenko, V. M. (2001). "Exponential and power-law probability distributions of wealth and income in the United Kingdom and the United States," *Physica A,* 299, 213-221.

Gwilliam, K. (2008). "A review of issues in transit economics," *Research in Transportation Economics*, 23, 4-22.

Mokhtarian, P. L., and Chen, C. (2004). "TTB or not TTB, that is the question: a review and analysis of the empirical literature on travel time (and money) budgets," *Transportation Research Part A*, 38, 643-675.

Morrison, P. F. (1997). "Restructuring effects of deregulation: the case of the New Zealand taxi industry," *Environment and Planning A*, 29(5), 913-928.

Naldi, M. (2003). "Concentration indices and Zipf's law," *Economics Letters,* 78, 329-334.

Ongkittikul, S., and Geerlings, H. (2006). "Opportunities for innovation in public transport: effects of regulatory reforms on innovative capabilities," *Transport Policy*, 13(4), 283-293.

Roumboutsos, A., and Kapros, S. (2008). "A game theory approach to urban public transport integration policy," *Transport Policy*, 15, 209-215.

Schaller, B., and Consulting, S. (2005). "A regression model of the number of taxicabs in U.S. cities," *Journal of Public Transportation*, 8(5), 63-78.

Scheiner, J. (2009). "Interrelations between travel mode choice and trip distance: trends in Germany 1976-2002," *Journal of Transport Geography,* doi:10.1016/j.jtrangeo.2009.01.001

Stanley, M.H.R., Amaral, L.A.N., Buldyrev, S. V., Havlin, S., Leschhorn, H., Maass, P., Salinger, M. A., and Stanley, H. E. (1996). "Scaling behaviour in the growth of companies," *Nature*, 379: 804-806.

Su, B.B., Chang, H., Chen Y. Z., and He, D. R. (2007). "A game theory model of urban public traffic networks," *Physica A*, 379, 291-297.

Van de Velde, D. M. (1999). "Organisational forms and entrepreneurship in public transport Part I: classifying organisational forms," *Transport Policy*, 6, 147-157.

Wang, J., An, S., and Xie, B. (2006). "State management methods for urban passenger transport system based on artificial system," *Dynamics of Continuous Discrete and Impulsive Systems-Series B-Applications & Algorithms 13*, 12(s), 950-954.

Wu, J. J., Gao, Z. Y., and Sun, H. J. (2007). "Effects of the cascading failures on scale-free traffic networks," *Physica A*, 378, 505-511.

Ying, J. Q. (2007). "Efficient fare and value capture: a unified theoretic approach," *Proceedings of the 11th World Conference on Transport Research*, Berkeley, Paper No. E5-353, 1-18.

Performance Evaluation of Bus Transit Routes Based on the BCC Model

Yadan Yan, Xiucheng Guo, Chaoping Liu

School of Transportation, Southeast University, Sipailou 2#, Nanjing, Jiangsu 210096, China; PH: 86-25-83795528; Email: wenhee@163.com

Abstract

Data envelopment analysis(DEA) is an optimization-based technique widely used to measure relative efficiencies of public transit agencies. This paper evaluates the performance of bus routes within a public transit system using the BCC model of DEA. We construct a concept framework of performance evaluation and select appropriate input and output indexes to compute the operation efficiency and service effectiveness of each bus route. Function-achievement is then analyzed and bus routes which need to be optimized and regulated are screened out. Twelve bus routes of Suzhou are selected for the practical application.
Keywords: bus routes; BCC model; performance evaluation; operation efficiency; service effectiveness

Introduction

Public transit is an integral and important component of urban transportation systems. Its advantages of convenience, speediness and effectiveness are easier to be given full play comparing to other traveling modes. Currently, such problems as congestion and slow speed of bus vehicles and deficit of bus transit enterprises emerge in many cities. This trend has stimulated much research interest on the performance evaluation of public transit systems (Taylou, Garrett, & Iseki, 2001; Murray, 2001; Tsamboulas, 2004). However, much existing research concentrates on evaluating the performance of overall operation efficiency of the whole public transit system using Data envelopment analysis(DEA). Little attention has been given to the operation efficiency and service effectiveness of single bus transit routes.

In practice, bus routes that need optimizing and regulating are usually screened out through judging such indexes as the line length, non-linear coefficient and coverage

rate subjectively and qualitatively when we plan the bus transit of a certain city. Planning schemes lack quantitative basis and pertinence.

Performance evaluation essentially should include two aspects, namely, the operation efficiency from the transit enterprises perspective and service effectiveness from the passengers perspective. This paper use the BCC model of Data envelopment analysis (DEA) to evaluate the performance of each bus transit route and bus routes that should be paid attention to be optimized and regulated are then screened out.

Methodology

Data envelopment analysis(DEA) is an optimization-based technique widely used to measure relative efficiencies. The main function of a bus route is to supply rapid and convenient service for passengers. It could be considered as an input-output system and its efficiency evaluation is to analyze the function-achievement under the conditions of certain input resources. The larger the relative efficiency is, the more efficient a bus route is.

BCC Model. In our case, each bus route is treated as a decision making unit (DMU). The DEA model adopted here is the output-oriented BCC model (Banker et al.,1984), as the overall objective of a bus route is to serve as many passengers as possible. Another reason is that it makes the variable returns to scale (VRS) assumption, which means that efficiency may increase or decrease with a change of size in input or output. The VRS assumption is more appropriate for bus routes with higher variability of productivity (Yong Lao, Lin Liu, 2009). Mathematically, VRS suggests that the estimated production frontier can pass anywhere relative to the origin in input-output space.

Firstly, the model variables are defined as follows:

- j: index of decision making units, j=1,2,…,n ;
- i: index of input, i=1,2,…,m ;
- r: index of output, r=1,2,…,s ;
- x_{ij}: the i th input for DMU_j ;
- y_{rj}: the rth output for DMU_j ;
- λ_j: the nonnegative scalers(weight) for DMU_j ;
- μ: the optimal output level.

Mathematically, the output-oriented BCC model(Banker et al.,1984) can be written as:

$$Max \quad \mu \tag{1}$$

$$\text{s.t.} \quad x_{i0} \geq \sum_{j=1}^{n} x_{ij}\lambda_j \qquad i=1,\ 2,\ldots,\ m \tag{2}$$

$$y_{r0}\mu \le \sum y_{rj}\lambda_j \qquad r=1,2,\ldots,s \tag{3}$$

$$\sum_{j=1}^{n}\lambda_j = 1 \tag{4}$$

$$\lambda_j \ge 0 \qquad j=1,2,\ldots,n \tag{5}$$

Since the model needs to be solved n times (once for every DMU), DMU_0 is used to represent one of the n DMUs under evaluation. x_{i0} and y_{r0} are the ith input and rth output for DMU_0, respectively. The DEA model essentially seeks to create the most efficient virtual DMU for a given DMU_0, whose productivity is based on the linear combination of peer DMUs' input and output. The objective function(1) attempts to maximize the level of output for the virtual DMU. Constraint(2) sets the limit of input to no more than the observed amount of input. Constraint(3) specifies that the output level is at least as good as that of DMU_0. Constraint(4) ensures that the DEA model's variable returns to scale(VRS) status. Constraint(5) imposes non-negativity restrictions for λ_j.

If μ=1.0 turns out to be the optimal solution, then DMU_0 is technically efficient, since the model is unable to find a virtual DMU with a higher level of output without increasing the current level of input. If the optimal solution μ is greater than 1.0, then the DEA model has identified a virtual DMU that can perform better than DMU_0. This indicates that DMU_0 is not technically efficient. The higher the value of μ is, the less efficient the DMU_0 is. The relative efficiency score of DMU_0 can be calculated as $1/\mu$(Yong Lao,Lin Liu,2009).

Input-output variables. Figure.1 provides a graphical illustration representing the implementation process of the bus service along a bus route. Bus transit enterprises provide vehicles, stops and lines which product such outputs as vehicle miles, schedule reliability and average traveling time. These service indexes are then treated as the input set from the passengers perspective whose output is passenger volume. And the final output of transit enterprises is revenue.

The whole process is divided into two phases, A→D and B→C. In order to evaluate the performance of a bus route from both the bus transit enterprise and passenger perspective, we ought to compute the operation efficiency and service effectiveness. Key affecting factors of operation efficiency are the resources provided by bus transit enterprises while key affecting factors of service effectiveness are passengers' service requirements. For bus routes with high operation efficiency but low service effectiveness, their service is not well received by passengers; for bus routes with high service effectiveness but low operation efficiency, they couldn't develop durably from the enterprise perspective.

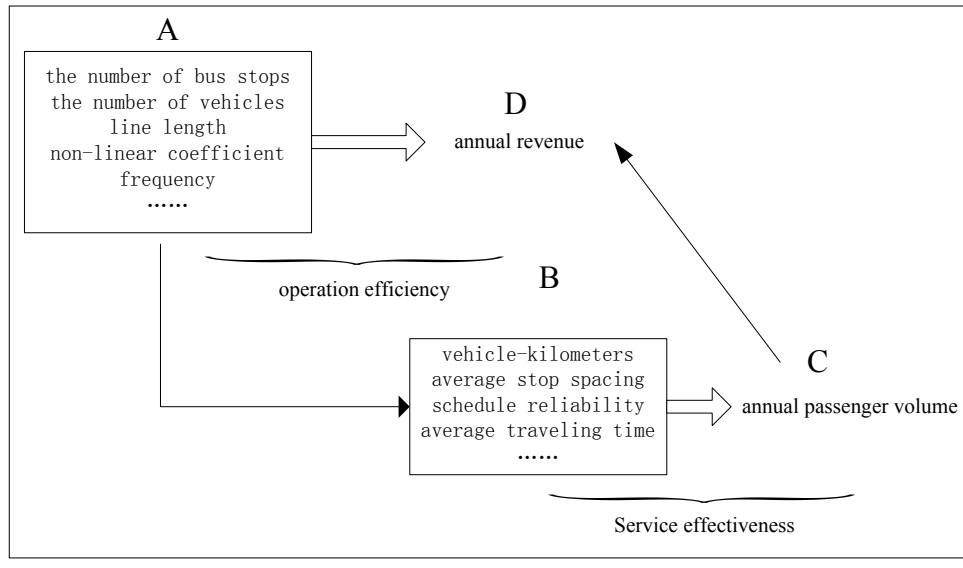


Figure 1. Concept framework of performance evaluation of bus routes.

From the operator's perspective, the objective is to use as few resources as possible and make as much profit as possible. From the passengers' perspective, bus routes should meet the demand by providing cheap and direct service to them. Criteria for using public transport can also include vehicle and transfer terminal comfort, regularity, service coverage and frequency. It is the main challenge in transit planning to find an equilibrium between these conflicting objectives. Accordingly, Table 1 displays the input-output specifications with which we evaluate bus transit routes. We use operation efficiency to measure the productivity of supply and service effectiveness to measure the benefits of demand for individual bus routes.

Because of the absence of actual costs of labor, fuel and other operational expenses for individual bus routes, it is reasonable to assume that the cost of operating a bus route is related to operation time, line length and the number of stops. In addition, factors such as service type(local or express routes) are not addressed in this paper.

Table 1. Input-output indexes of BCC model.

Performance	DMU	Input Indexes	Output Indexes
operation efficiency	each bus route	the number of bus stops, the number of vehicles, line length, non-linear coefficient, frequency	annual revenue
service effectiveness	each bus route	vehicle-kilometers, average stop spacing, schedule reliability, average traveling time	annual passenger volume

Case Study

We apply the methodology to evaluate the twelve bus routes in Suzhou. Data of the

year 2005 (Zhang Bin, 2006)are used.

Data. Corresponding values of input and output indexes are displayed in Table 2 and Table 3 respectively.

Table 2. Data of input –output indexes of operation efficiency.

	Inputs					Outputs
DMU	Number of bus stops	Number of vehicles	Line length	Frequency	Non-linear coefficient	Annual revenue
2W	23	12	13.55	5	1.36	2.00
7W	34	20	20.40	12	1.19	1.04
8W	18	14	12.80	4	1.43	2.44
12W	30	16	20.70	5	1.08	0.85
19W	31	9	29.20	5	1.07	1.41
23W	26	14	17.50	5	2.49	0.96
27W	40	12	25.90	4	1.22	0.80
32W	25	10	16.80	5	1.43	2.39
49W	22	12	14.10	4	1.86	1.48
61W	34	20	20.40	5	2.21	0.83
62W	45	18	39.10	6	1.49	0.41
70W	32	12	19.20	5	1.46	1.22

Table 3. Data of input-output indexes of service effectiveness.

	Inputs				Outputs
DMU	Vehicle-kilometers	Average stop spacing	Schedule reliability	Average traveling time	Annual passenger volume
2W	180.96	0.589	0.29	78.9	10.89
7W	222.45	0.600	0.27	88.7	16.25
8W	164.05	0.711	0.26	68.6	13.10
12W	248.36	0.690	0.33	76.2	13.53
19W	236.82	0.942	0.44	110.9	14.19
23W	326.44	0.673	0.34	112.9	9.45
27W	157.56	0.648	0.29	88.8	3.80
32W	246.30	0.672	0.30	79.4	10.21
49W	218.35	0.641	0.32	79.8	7.58
61W	206.84	0.600	0.58	67.3	11.65
62W	194.60	0.869	0.64	133.3	21.94
70W	125.06	0.600	0.18	100.2	7.13

BCC model results and analysis. The model is solved using the LINDO (Linear, Interactive, and Discrete Optimizer) software which is specially used for solving linear programming problems. The best way to obtain a comprehensive picture of bus

route performance is to compare operation efficiency with service effectiveness comprehensively. Whereas previous studies have indicated that, in general, there is a positive relationship between efficiency and effectiveness (Karlaftis, 2004), no clear positive or negative associations are present between operation efficiency and service effectiveness for bus routes in Suzhou (Table 4). Bus routes with high operation efficiency may or may not have high service effectiveness, and vice versa.

Table 4. BCC model results for bus routes in Suzhou.

	operation efficiency		service effectiveness	
DMU	μ_1	$1/\mu_1$	μ_2	$1/\mu_2$
2W	1.000000	1.000000	2.265793	***0.758452***
7W	1.000000	1.000000	1.000000	1.000000
8W	1.000000	1.000000	1.000000	1.000000
12W	1.547857	***0.355648***	1.280746	***0.862367***
19W	1.000000	1.000000	1.000000	1.000000
23W	1.000000	1.000000	1.000000	1.000000
27W	1.691532	***0.331263***	3.135360	***0.307982***
32W	1.000000	1.000000	1.821020	***0.671517***
49W	1.000000	1.000000	3.604940	***0.457052***
61W	1.000000	1.000000	3.251346	***0.554305***
62W	1.369375	***0.692629***	1.000000	1.000000
70W	1.424173	***0.24542***	1.000000	1.000000

According to computing results, twelve bus transit routes are further classified into four groups:

Line 7, Line 8, Line 19 and Line 23 are bus routes with high scores in both operation efficiency and service effectiveness. They could be preserved when regulating and optimizing bus routes in Suzhou.

Two bus routes, Line 12 and Line 27, have very low scores of operation efficiency and service effectiveness, are the worst performers that should be carefully re-planned.

Four bus routes, Line 2, Line 32, Line 49 and Line 61, have considerably high score of operation but low score of service effectiveness. These four routes are performing satisfactorily from the bus transit enterprise's perspective and are strategically important for them. But given the low level of service effectiveness, they should be appropriately adjusted to attract more passengers and optimized to supply better service.

The bus routes that are operationally efficiently but with low value of service effectiveness should be supported to maintain their current basic line alignment. Operation efficiency could be improved through enhancing its operation and management.

Conclusions

In this paper we construct a concept framework of performance evaluation and adopt BCC model to calculate the operation efficiency and service effectiveness from two perspectives respectively. Twelve bus routes in Suzhou are taken for practical example finally. There are limitations in this paper. Each bus route is treated independently and the transfer points between bus routes are not considered. Besides, selection of indexes and external environment are not comprehensively thought out. Future research is needed to tackle these complex problems.

References

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models of estimating technical and scale inefficiencies in data envelopment analysis. *Management Sciences*, 9(9), 1078–1092.

De Borger,B.,Kerstens,K.,&Costa,A.(2002).Public transit performance.What does one learn from frontier studies? *Transportation Review*, 22(1):1-38.

Geng Mingjun.(2004). *Research on optimization method of bus network in the near future*, Chengdu:Southwest jiaotong university.

Karlaftis, M. G .(2004). A DEA approach for evaluating the efficiency and effectiveness of urban transit systems. *European Journal of Operations Research*, 152: 354-364.

Ming-Minin Yu, Chih-Ku Fan.(2008).Measuring the performance of multimode bus transit:A mixed structure network DEA model. *Transportation Research Part E*, 1-15.

Murray, A. T. (2001). Strategic analysis of public transport coverage. *Socio-Economic Planning Science*, 35(3): 175-188.

Taylou, B. D., Garrett, M., & Iseki, H. (2001).Measuring cost variability in provision of transit service. *The Journal of the Transportation Research Board*, 1735, 101-112.

Tsamboulas, D. A. (2004). Assessubg performance under regulatory evolution: A European transit system perspective. *Journal of Urban Planning and Development*, 132(4), 226-234.

Lao,Y., Liu. L. (2009). Performance evaluation of bus lines with data envelopment analysis and geographic informatin systems. *Computers Environment and Urban Systems*, 1-9.

Zhang B. (2006). *Research on the index evaluation system of normal public Transit's route optimization and resourch distribution*. Nanjing Southeast University.

Creating Multi-routes Combination Bus Timetable

Bingqiang Situ[1], Wenzhou Jin[2] and Ming Wei[3]

[1] PhD, [2] Professor, [3] PhD, School of Civil Engineering and Transportation, South China University of Technology, 510640, Guangzhou, China, Email: benst2002@126.com, ctwzhjin@scut.edu.cn, oscar.weiming@gmail.com

Abstract

This paper addresses the problem of creating a combination bus timetable for a given bus network with multi-routes so as to minimize the waiting time of non-transfer passenger and transfer passenger. It attempts to arrange the departure time of the bus in any route to satisfy the demand of passenger with minimum waiting time, and enables the transfer passenger from one route to another with minimum waiting time at the transfer station. In this paper, the problem is formulated as a nonlinear programming problem, and the departure time of each bus in any route is the decision variable. This problem is solved by particle swarm optimization (PSO). The computational results presented in this paper show that the optimization method presented in this paper can create a multi-routes combination bus timetable, which satisfy the passenger demands, to minimize the waiting time. A comparison between traditional timetable and combination timetable shows that combination timetable can perform better satisfying the mutative passenger demands and reduce more waiting time.

Key Words: public transportation planning; combination bus timetable; time headway; transfer; genetic algorithm

Introduction and the Scope

The objective of this work is to create the combination bus timetable for multi-routes in a given bus network. It mostly presents the optimization models for this timetable

probleme to minimize the waiting time for two types of passengers. One type is the non-transfer passenger in a trip from their first arriving station to first getting on bus, and the other is the transfer passenger from one route to another at the transfer station. In order to minimize the waiting time, it needs to adjust the timetable of single route to satisfy the non-transfer passenger demand and also to combine several routes to adjust their timetables to minimize the waiting time for transfer passengers. A combination bus timetable model could achieve these aims for multi-routes in a given bus network.

Bus timetable is one of the vehicle scheduling problem (VSP), which is a classical optimization problem. It consists in assigning a set of scheduled trips to a set of vehicles in the public transportation, allocations of crews and vehicles to routes, and service frequency for each route, so VSP is a part of the operational planning of public transportation systems (Dantzig and Fulkerson 1954). Usually, network design includes the design of route layouts and the determination of associated operational characteristics in VSP. According to Ceder and Wilson (1986), network design elements are part of the overall operational planning process which includes five steps: design of routes, setting frequencies, developing timetables, scheduling buses, and scheduling drivers. Ngamchai and Lovell (2003) present a new model to optimize bus transit routes configuration and service frequency on each route. A route evaluation algorithm designs the service frequency on each route and applies headway coordination techniques at transfer points to decrease transfer time between routes if it is possible to reduce the overall cost.

Actually, bus timetable is an importation and difficult task of public transport, which has received much attention in the literature. Some passengers transfer from one vehicle to the other at a station (Palma and Lindsey 2001), and the cost is the main focus in the presented literature. We also force the two vehicles arrive or depart at a station at the same time for the passenger transfers. As noted in Ceder et al. (2001), it present a mixed integer linear programming model to generating a timetable for a given network so as to maximize their synchronization, and a heuristic algorithm is developed to solve the problem. But in many cases, reducing the waiting time is easier to be realized than synchronization. So Zhou and Yang (2004) minimize the waiting time when transferring among different routes base on the real-time information, and present a linear programming model with one transfer station and multi transfer stations. This method is especially for the routes with large headway. Heilporn et al. (2008) present two equivalent mixed integer linear programming models to minimize the total delay on the public transportation network for passenger travel. Situ and Jin (2009) minimize the time headway between vehicles of several routes arriving at a station, and the total time headways of their arriving (or departing) are less as far as possible. Furthermore, Situ and Jin (2010) minimize or

maximize the time headway between vehicles of several routes in cooperation or competition transportation network. The bus timetable should be chosen to best satisfy demand or transit assignment. Yu et al. (2006) present a bi-level programming model for the bus frequency design, which determines the optimal bus frequency aiming to minimize the total travel time of passengers. In the model, the upper-level optimize the bus frequencies from the result of passenger assignment, and the lower-level assigns the bus passengers. Schedule-based or Frequency-based assignment model for congested transit networks are presented to reduce the cost with passenger assignment, see for Poon et al. (2004) and Schmocher et al. (2008). The above models are often NP hard problems, and heuristic algorithms are developed to solve the problems.

The bus timetable mostly meets the non-transfer passenger demands in the real public transport, and then it meet the transfer passenger demands as far as possible. Because the non-transfer passengers are more than the transfer ones, and passengers would like one leg in their trip without any transfer, though some transfers are inevitable. Bus companies and government always encourage transfer in passenger's trip, and allowance or discount is even obvious. Actually, the transfer is the indispensable demand, but not the primary demand in many cities.
The problem addressed in this paper has not been dealt with completely satisfactory in the above literatures. These few articles are not looking at the two type passengers waiting time simultaneously for multi-routes in a bus network. The purpose of this paper, though treating the timetable problem in a mathematical method, is to establish a more useful and comprehensive combination bus timetable for single route and multi-routes. That is to provide recommendations on how to adjust the bus timetable to satisfy the two type passenger demands.

The paper is organized as follows. Section 2 provides a mathematical formulation of combination bus timetable, which includes non-transfer and transfer waiting time models. Section 3 gives a particle swarm optimization that solves the problem. A practical case applied of proposed solution methodology is presented in Section 4. A summary and directions for further research are provided in Section 5.

Formulating the Model

Assumption. We use the following assumptions in this paper.

- The time between any given stations is constant in a given planning horizon A.
- The departure time and arrival time are same at a given station, so that the transfer time is ignored, just considering the waiting time.
- Vehicles must run from origin to destination by stopping at each station of the

route, without any overtaking the vehicle of the same route.

- The capacity of vehicles is infinity. That is the passenger can get on the first coming bus, which is available, without waiting for the next or several buses.

Notation. The given bus network is presented by a directed graph, $G=\{L, I\}$,where L is a set of arcs representing the directed traveling path of the bus routes, i.e. one route has two paths with up and down direction; I is a set of bus stations in the network. $\bar{I}$ is a set of key transfer stations ($\bar{I} \subset I$). The problem data are the following: A_r is the rth planning horizon (the departure time of the buses must be set in the interval [$A_{r\text{-}1}$, A_r] which is a discrete interval); R_j is the number of departures to be scheduled for route j during the interval [$A_{r\text{-}1}$, A_r] ($1 \le j \le L$); t_j^i is the traveling time from the starting point of route j to the key transfer station i ($1 \le j \le L$, $1 \le i \le \bar{I}$) (traveling time can be referred to as the mean traveling time, and it is considered deterministic); Q_j and Q_j^{max} denote the demand and the maximal section demand of route j in the interval [$A_{r\text{-}1}$, A_r]; B_j and b_j represent the operating cost and supplier cost in route j; l_j is the load factor in route j; C denotes the vehicle capacity; T_j is the bus traveling time per trip in route j; H_j and h_j denote the maximal and minimum headway between two adjacent bus departures in route j.

The following should be noted: (a) this paper assumes throughout that the first departure in a day in each route j must take place in the earliest time (if the planning horizon in a day is [6:00, 22:00], the first departure is 6:00), and the last departure in each route j in each planning horizon must take place in A_r (for horizon [7:00, 8:00], the last departure is 8:00), so that the first departure in the next planning horizon must take place in the interval (A_r, A_{r+1}+H_j]; (b) the problem is impractical unless the following constraint holds, for each j:

$$R_j \times h_j \le A \le R_j \times H_j \tag{1}$$

R_j, H_j and h_j can be formulated respectively as: $R_j = \frac{Q_j^{\max}}{l_j \times C}$, $H_j = \frac{60C}{Q_j}$, $h_j = \frac{2T_j B_j}{b_j}$.

Decision Variables. The decision variables are the following:

- x_j^m represents the departure time of the mth bus in route j ($1 \le m \le R_j$, $1 \le j \le L$);
- the function $F_j(\cdot)$ is the cumulative distribution function in route j.
- λ_{jk}^{mni} is a binary variable that yields the value 1 if the following constraint holds:

$$\left(x_k^n + t_k^i\right) \le \left(x_j^m + t_j^i\right) \le \left(x_k^{n+1} + t_k^i\right) \tag{2}$$

(the arrival time of the mth bus in route j is between the ones of the (n-1)th bus and the nth bus of route k at transfer station i); otherwise, it yields the value 0.

Non-transfer model. Let function $P_j(\cdot)$ denote the distribution function in route j. The waiting time of non-transfer passenger for each route j is supposed as shown in

Figure 1(a), which describes the waiting time with function $P_j(\cdot)$. The model with distribution function is as follows:

$$Z_1(x_j^1,\ldots,x_j^{R_j}) = \min \int_{x_j^m}^{x_j^{m+1}} P_j(t)dt \tag{3}$$

$$\text{s.t.} \quad h_j \le (x_j^{m+1} - x_j^m) \le H_j, \quad 1 \le m \le R_j\text{-}1, \tag{4}$$

$$x_j^1 \le H_j, \quad x_j^{R_j} \le A. \tag{5}$$

Constraint (1) remains unchanged.

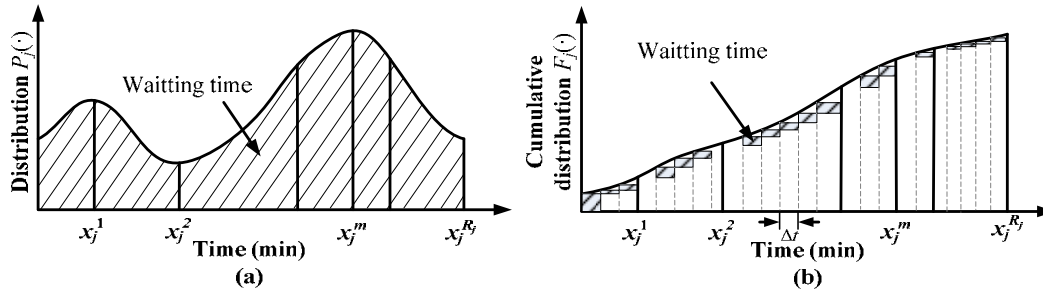


Figure 1. Waiting time of non-transfer passenger with $P_j(\cdot)$ and $F_j(\cdot)$.

As shown in Figure 1(a) and Eq. (3), it is difficult to get the accuracy value by the continuous variable. If the passenger arrives at the station in the interval (0, Δt], we assume the passenger arriving time is at Δt. This assumption is viable only if the Δt is little enough (Δt=1min in this paper). Base on this assume, the waiting time of non-transfer passenger for each route j is also supposed as shown in Figure 1(b), which describes the waiting time with function $F_j(\cdot)$. The model with cumulative distribution function is as follows:

$$Z_1'(x_j^1,\ldots,x_j^{R_j}) = \min \sum_{t=x_j^m+\Delta t}^{x_j^{m+1}-\Delta t} \left[F_j(t) - F_j(t-\Delta t)\right] \times \Delta t \cdot \tag{6}$$

Constraint (1), (4) and (5) remain unchanged.

Transfer Model. Let β_{jk}^i denote the transfer factor from route j transferring to route k at transfer station i. The two direction of each route can be represented two routes. Usually, there is no transfer passengers between these two direction routes, so transfer factor yields the value 0. Furthermore, it yields the value 0 when any two routes do not intersect at transfer station i. If constraint (2) holds, the number of transfer passengers from mth bus in route j to the nth bus in route k at transfer station i can be represented as

$$\beta_{jk}^i \times \left[F_j(x_j^{m+1}) - F_j(x_j^m)\right]$$

The model of the minimum waiting time for transfer passengers is as follows:

$$Z_2(x_j^1,\ldots,x_j^{R_j},x_k^1,\ldots,x_k^{R_k}) =$$
$$\min \sum_{j=1}^{L-1}\sum_{k=j+1}^{L}\sum_{n=1}^{R_k-1}\sum_{m=1}^{R_j-1}\sum_{i=1}^{\bar{I}}\left\{\lambda_{jk}^{mni}\times\beta_{jk}^{i}\times\left[F_j(x_j^{m+1})-F_j(x_j^m)\right]\times\left[(x_k^{n+1}+t_k^i)-(x_j^m+t_j^i)\right]\right\} \quad (7)$$

$$\text{s.t.}\quad x_j^1\le H_j,\quad x_k^1\le H_k,\quad x_j^{R_j}\le A,\quad x_k^{R_k}\le A, \quad (8)$$

$$\lambda_{jk}^{mni}=\begin{cases}1 & \text{if } (x_k^n+t_k^i)\le(x_j^m+t_j^i)\le(x_k^{n+1}+t_k^i),\\ 0 & \text{otherwise.}\end{cases} \quad (9)$$

Constraint (1) and (4) remain unchanged.

Combination Model. The combination multi-routes bus timetable model for minimun the whole waiting time of the non-transfer passengers and transfer passengers is as follows:

$$Z(x_j^1,\ldots,x_j^{R_j},x_k^1,\ldots,x_k^{R_k}) =$$
$$\min\left\{\begin{array}{l}\displaystyle\sum_{j=1}^{L-1}\sum_{k=j+1}^{L}\sum_{n=1}^{R_k-1}\sum_{m=1}^{R_j-1}\sum_{i=1}^{\bar{I}}\lambda_{jk}^{mni}\times\beta_{jk}^{i}\times\left[F_j(x_j^{m+1})-F_j(x_j^m)\right]\times\left[(x_k^{n+1}+t_k^i)-(x_j^m+t_j^i)\right]\\ \displaystyle+\sum_{j}^{L}\sum_{t=x_j^m+\Delta t}^{x_j^{m+1}-\Delta t}\left[F_j(t)-F_j(t-\Delta t)\right]\times\Delta t\end{array}\right\} \quad (10)$$

Constraint (1), (4), (8) and (9) remain unchanged. The function (10) is to be minimized by choice of the timetable $x_j^1,\ldots,x_j^{Rj},x_k^1,\ldots,x_k^{Rk.}$ The first line in this sum specifies the waiting time of transfer passenger between any two routes, and the second line specifies the waiting time of non-transfer passenger in each route.

Algorithm

The models presented in the paper are nonlinear programming problem, which would be solved by particle swarm optimization (PSO). PSO is inspired by the paradigm of birds seeking food (Kennedy and Eberhart, 1995). PSO has been applied to several science and engineering optimization problems and has been proven to be an efficient optimization algorithm. The decision x_j^m (solution candidates, $1\le m\le R_j$, $1\le j\le L$), which is the departure time, can be represented as a $_L$ chain which is a multidimensional particle (the dimension value of the particle is $\sum_{j=1}^{R_j}$). The algorithm guide swarms of particles towards the most promising regions ($h_j\le(x_j^{m+1}-x_j^m)\le H_j$, $[A_{r}\text{-}1,\ A_r]$) of the multidimensional search space. Then it compares the solution candidates to arrive at the optimal solution. The optimal value of function $Z^*(\cdot)$ represents the adapting value of optimal particle (optimal solution).

Case Study

The method presented in the previous section is applied to a case study on creating a combination bus timetable in Guangzhou University Town in Guangzhou City of

China. Guangzhou University Town is almost an independence island with some inner and though island routes. We choose 6 routes and 3 key transfer stations for this application. These routes are 381, 382, 383S, 383X, 387S and 387X ('S' and 'X' mean the up and down direction respectively, especially, 381 and 382 are the ring routes), which belong to Guangzhou No.3 Bus Company, shown in Figure 2. The 6 routes get through every zone (or the university) in Guangzhou University Town. The original field data applied in this case are mostly provided by Guangzhou No.3 Bus Company and Guangzhou Transportation Commission.

Table 1 presents the traveling time (say, in minutes) and the scheduling constraints in planning horizon [17:00-18:00]. We assume the transfer factors between every two routes (base on IC card data) are from 0.05 to 0.12, and some factors are 0. Figure 3 presents the field curves (collected from IC card data and income data) and fitting curves of distribution and cumulative distribution in the 6 routes. The parameters for PSO are set as follows: particle number is 40, iteration number is 100, learning factor *c*l, *c*2 are 1.494, and inertia weight is 0.729.

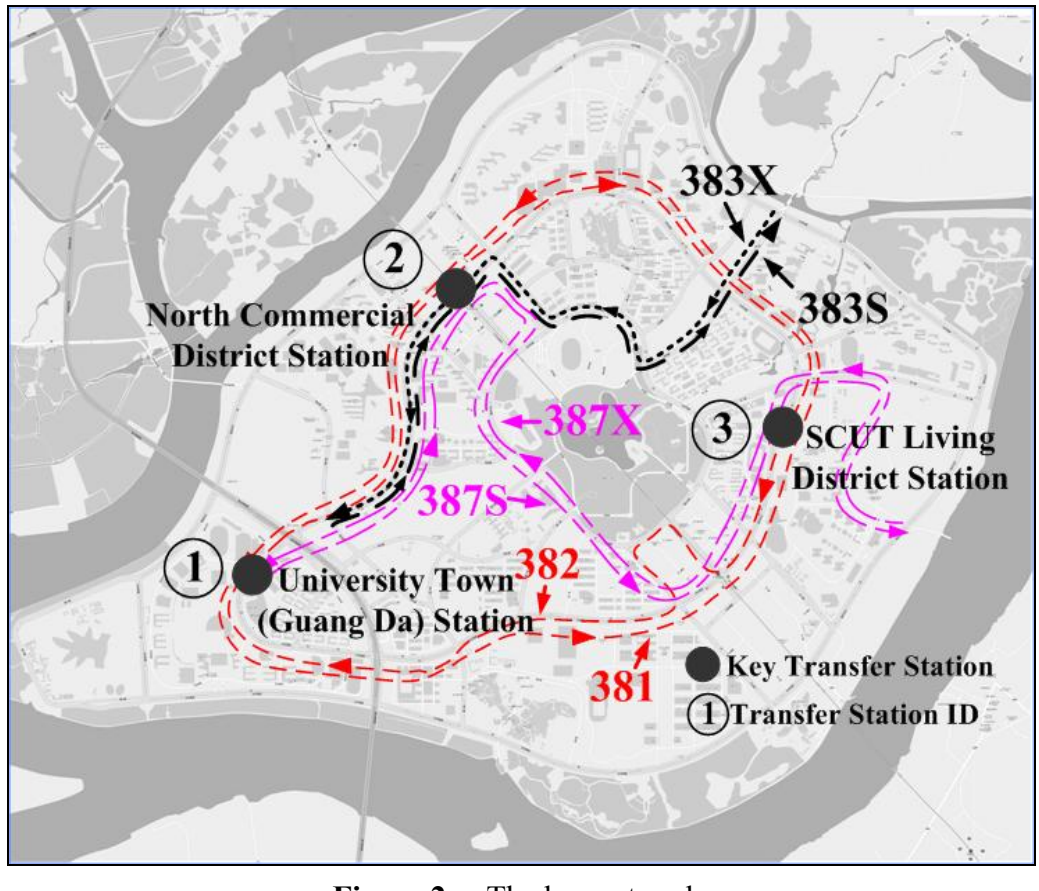


Figure 2. The bus network.

Table 1. Traveling time and scheduling constraints

Route No.	Key transfer station ID			Number of departure R_j	Headway constraint $[h_j, H_j]$
	①	②	③		
381	4.46	20.87	28.46	8	[3, 10]
382	31.56	17.55	7.50	8	[3, 10]
383S	*	5.78	*	5	[8, 20]
383X	*	10.43	*	5	[8, 20]
387S	0.00	11.89	20.25	4	[10, 25]
387X	23.35	14.14	5.23	4	[10, 25]

The route does not pass through the transfer station.

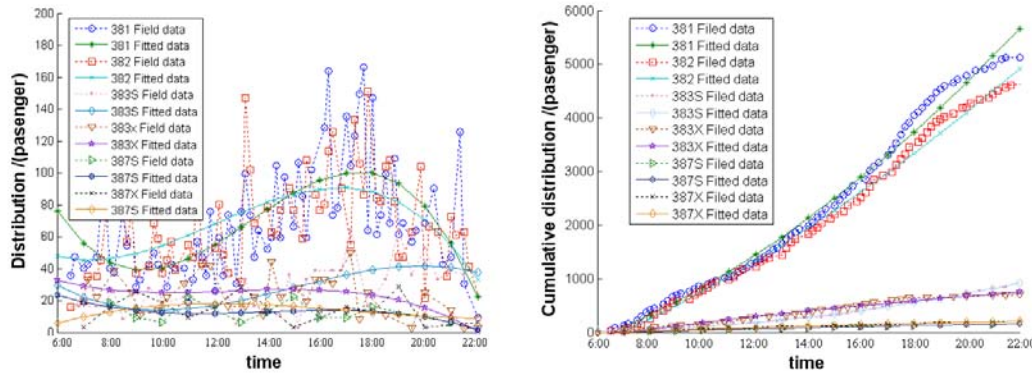


Figure 3. The field curve and fitting curve.

Results. With the value of variables shown as above, the combination bus timetable can be created based on Eq. (10) using Matlab to compute. Figure 4 shows that as the iterative time increases, the adapting value of optimal particle (optimal solution) reduces (the minimum Z^* is 3879 min). The combination timetable with multi-routes are showed in table 2.

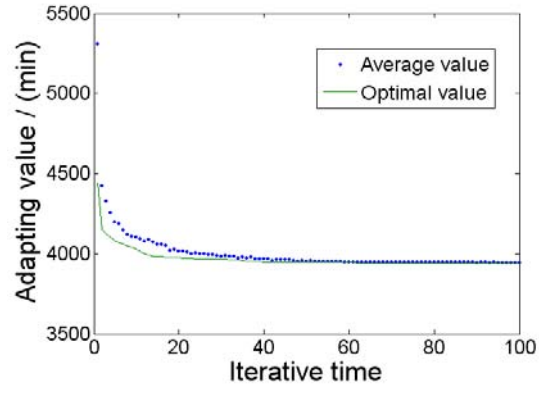


Figure 4. Curve of object value.

Table 2. Combination bus timetable in [17:00, 18:00].

Route No.	Departure time	Headway (min)
381	17:08 17:16 17:23 17:31 17:38 17:45 17:53 18:00	8-8-7-8-7-7-8-7
382	17:08 17:15 17:22 17:30 17:38 17:45 17:52 18:00	8-7-7-8-8-7-7-7
383S	17:12 17:24 17:36 17:49 18:00	12-12-12-13-11
383X	17:13 17:28 17:36 17:46 18:00	13-15-8-10-14
387S	17:15 17:31 17:41 18:00	15-16-10-18
387X	17:16 17:29 17:46 18:00	16-13-17-14

Table 3. Comparison between traditional timetable and combination timetable.

Planning horizon	Waiting time		Variation (%)
	Traditional	Combination	
6:00- 8:00	7006	6793	-3.04
8:00-10:00	7892	7627	-3.36
10:00-12:00	8657	8301	-4.11
12:00-14:00	8034	7600	-5.40
14:00-16:00	8740	8266	-5.42
16:00-18:00	8740	8250	-5.61
18:00-20:00	8650	6705	-22.49
20:00-22:00	10857	10148	-6.53

Discussion. A simple comparison between the waiting time of traditional timetable and combination timetable in each planning horizon is highlighted in Table 3: it shows, for the same number of departures, the combination bus timetable can reduce more passenger waiting time than traditional timetable. We can observe, in all planning horizons, the variation are negative (from -3% to -7% mostly), that is, the combination timetable performs better than the traditional one. Moreover, the variation is -22.49% in planning horizon [18:00-20:00], as the range of passenger demand in most routes are very great in this planning horizon (see for the Figure 3). As a consequence, the optimization method presented in the previous section can reduce the waiting time of passengers, and the combination bus timetable for multi-routes satisfies the passenger demand.

Conclusions

This paper has presented the new optimization models for creating a combination bus timetable with multi-routes, which allows us to reduce the waiting time of non-transfer passengers and transfer passengers. The mathematical models described, with deterministic traveling time, are formulated as the nonlinear programming problems solved by particle swarm optimization (PSO). For a given bus network, the computational results show that the optimization method presented in this paper can create a multi-routes combination bus timetable satisfying the passenger demands.

The comparison between traditional timetable and combination timetable show that combination timetable performs better at satisfying the mutative passenger demands, in order to reduce more waiting time of passenger.

Also note that, at the moment, the models assume that the time between any given stations is constant, so that a bus timetable could be created in advance. Nevertheless, performs in the field would make some difference to the timetable, as the traveling time would change frequently for the traffic jam, signal control or accident. Future work will be to find out the key station and key path which affect the traveling time base on complex network theory.

From the public transportation network point of view, the results show that there is room for improving the timetable to satisfy the passenger demand. Creating a multi-routes combination bus timetable is one of the effective methods. From the passenger point of view, the average waiting time is very little for each passenger, so they do not only care about the waiting time but also the fare, the comfort, and so on. As a consequence, it should increase the attractiveness of public transportation.

Acknowledgment

This work was supported by High-Technology Research and Development Program of China (863 program) under contract No.2007AA11Z201 and National Science Foundation of China under contract No.50878089. The authors thank the anonymous referee for her/his helpful comments, and thank Guangzhou No.3 Bus Company and Guangzhou Transportation Commission, and so on.

References

Ceder A., Wilson, N. H. M. (1986). “Bus network design.” *Transportation Research Part B*, 20, 331-344.

Ceder A., Golany B., Tal O. (2001). “Creating bus timetables with maximal synchronization.” *Transportation Research Part A*, 35, 913-928.

Dantzig G, Fulkerson D. (1954). “Minimizing the number of tankers to meet a fixed schedule.” *Naval Research Logistics Quarterly*, 1(3), 217-222.

Heilporn G., Giovanni L. De, Labbé M. (2008). “Optimization models for the single delay management problem in public transportation.” *European Journal of Operational Research*, 189(3), 762-774.

Kennedy, J., and Eberhar, R.C. (1995). “Particle swarm optimization.” *Proceedings of the IEEE International Conference on Neural Networks*, Perth, Australia, IEEE Service Center, NJ, IV: 1942-1948.

Ngamchai S., and Lovell D.J. (2003). "Optimal time transfer in bus transit route network design using a genetic algorithm." *Journal of Transportation Engineering*,129(5), 510-521.

Palma A. de, Lindsey R. (2001). "Optimal timetables for public transportation." *Transportation Research Part B*, 35(8), 789-813.

Poon M.H., Wong S.C., Tong C.O. (2004). "A dynamic schedule-based model for congested transit networks." *Transportation Reasearch Part B*, 38:343-368.

Schmocher J.D., Bell M.G.H., Kurauchi F. (2008). "A quasi-dynamic capacity constrained frequency-based transit assignment model." *Transportation Reasearch Part B*, 42:925-945.

Situ B.Q., Jin W.Z. (2009). "Study on the vehicle scheduling problem in transportation system." *International conference on computer science and information technology*, IEEE press, Vol 5, 175-178.

Situ B.Q., Jin W.Z. (2010). "Research on scheduling optimization model in cooperation and competition transportation network." *Journal of Highway and Transportation Research and Development*, 27(5).

Yu B., Yang Z.Z., Cheng C.T., Zuo Z. (2006). "Bi-level programming model for optimizing bus frequencies and its algorithm." *Journal of Jilin University* (*Engineering and Technology Edition*). 36(5), 664-668.

Zhou X.M., Yang X.G. (2004). "Study of dispatching at minimum waiting time of public transportation transfer under the condition of ITS." *China Journal of Highway and Transport*, 17(2), 82-84.

Ant Colony System for Multi-Depot Bus Scheduling Problem with Budget and Pollutant Emission Control

Ming Wei[1], Wenzhou Jin[2], Bo Sun[3] and Xiaoni Hao[4]

[1] Ph.D. Candidate, [2.]Professor, [3.]MSc Students, [4.]Instructor, School of Civil Engineering and Transportation, South China University of Technology, Guangzhou, China 510640, Email: oscar.weiming@gmail.com, ctwzhjin@scut.edu.cn, bowensunny@163.com, xnhao@scut.edu.cn

Abstract

To aim at maximizing utilization of vehicles, this paper discusses how to spend a limited budget to buy different types of buses to cover all assigned trips for bus scheduling to meet some requirements such as fuel restrictions, depot capacities, and an upper limit for air pollutants emitted from these buses. The mathematical model is formulated in a form of the integer programming problem in "Block Based" formulation ways, and solutions are obtained using the ant colony optimization (ACO) which uses new heuristic information and redesign a feasible solution construction according to the features of bus scheduling. Finally, numerical computations show that ACO can find good solutions effectively and efficiently.
Key Words: bus scheduling; ant colony optimization; air pollution; budget; fuel restrictions

Introduction

Recently, bus scheduling (VSP), where large numbers of trips with prescribed starting and ending times are assigned to buses located at different depots to minimize capital and operating costs, is one of the areas of concern in transportation research. Bodin and Golden (1981) proved that this scheduling was a difficult NP-hard problem when some constraints, such as the route time, made the problem more realistic. Studies in bus scheduling have stimulated the enthusiasm of research for many scholars, and related literature can be mainly divided into several categories according to different side constraints such as: 1) multi-depot VSP (Haghani and Banihashemi, 2002; Gintner et al., 2005; Kliewer et al., 2006; Wang and Shen, 2007); 2) multi-vehicle-type VSP (Gintner et al,. 2005; Kliewer et al., 2006; Li and Head,

2009); 3) VSP with route time constraints (Haghani and Banihashemi, 2002; Wang et al., 2006; Wang and Shen, 2007); 4) VSP with capacity of depot restrictions (Prakash et al. 1999; Kliewer et al. 2006; Liu 2007); and (Mes et al., 2007; Li et al., 2009). Studies referring to the reduction the air pollution are a scarce commodity. Until now, only Li and Head (2009) integrated VSP when evaluating new types of buses that use alternative energy sources to reduce emissions. Their results showed that the bus-scheduling model could significantly reduce the bus emissions while only slightly increasing operating costs. However, the paper only intended minimize the sum of fuel costs for the trips and didn't involve maximum utilization of vehicles, i.e., minimum capital costs for required buses and other operating costs such as waiting time and deadhead time, and fuel restrictions for buses. In reality, there is a gap between the model and some real-world situations of VSP.

In this paper, we study in depth a more real-world multi-depot VSP with the capacity of depot restrictions involving how to spend a limited budget in purchasing buses to cover all assigned trips therefore minimizing the capital costs for buses and the operating costs such as deadhead and waiting time to restrict air pollutants within a boundary or bounds. Due to fuel restrictions, buses can also be allowed to re-fuel at two adjacent gas points when their fuel supply is low in the middle of the day. The mathematical model is formulated in a form of an integer programming problem in "Block Based" formulation ways, and we apply ACO to resolve it. The rest of this paper is organized as follows. In section 2, we describe the problem and discuss its mathematical formula. In section 3, we introduce the proposed algorithm. In section 4, we examine the performance of the algorithm through a series of simulations. In section 5, we summarize our work.

Problem Description

The bus scheduling problem we examine in this paper can be formalized as follows. We are given n trips $T_1,\cdots, T_n$ that must be served by vehicles belonging to a given set of vehicle types K. These vehicles, purchased with a limited budget B and initially located at different depot d with a capacity r_d, can be refuelled to continue their tasks in the middle of the day ($d \in D$, D denotes a set of depots). Let B_d^k ($k \in K$) be the set of kth type of vehicles stationed at depot d which has an amount of air pollutants emitted and a maximum limit of working hours without fuelling. Each trip T_i requires uninterrupted driving from a starting relief point at a given starting time s_i to an ending relief point at a given ending time $e_i \geq s_i$. An ordered trip pair (T_i, T_j) is called infeasible if the same vehicle cannot cover T_j immediately after T_i, i.e., $e_i+w_{ij} \geq s_j$; otherwise, it is called feasible. $w_{ij} \geq 0$ denotes travel time of dead-heading trip between the ending relief point of trip T_i and the starting relief point of trip T_j.

The first $|D|$ trips $T_1,...,T_{|D|}$, each of which indicates the beginning and the ending of a duty in a depot, are dummy. Hence, we set $s_k=-\infty$, $e_k=+\infty$, and $w_{ik}=+\infty$ (i, k=1,…,$|D|$); moreover, let w_{kj} (j=$|D|$+1,…,$|D|$+n) represent the travel time incurred for a duty if depot k started with trip T_j. A duty for vehicle j beginning and ending at

depot k is denoted an ordered sequence set $F_j^k = (T_k, T_{j1}, ..., T_{jq}, T_k)$, which contains a dummy trips and $q \geq 1$ various non-dummy trips. The duty is feasible if it satisfies some operating constraints such as maximum limit on working hours and fuel restrictions related to vehicles and any two consecutive trips constitute a feasible pair. Bus scheduling is thus the problem of determining a set of feasible duties such that (a) all trips are assigned to exactly one vehicle, (b) each trip T_j (j=|D|+1,…, |D|+n) belongs to exactly one duty, (c) no more than r_d duties originate from each depot d, and (d) the capital costs for buses and the amounts of each air pollutant emitted by them are restricted to a given boundary. Our primary objective is to find a schedule of the vehicles that minimizes the capital costs for buses and the operating costs such as deadhead and waiting time.

Based on these descriptions, the formulation is written as follow:

$$\text{Min } c_0 \sum_{d\in D}\sum_{k\in K} |B_d^k| + c_1 \sum_{d\in D}\sum_{k\in K}\sum_{l\in B_d^k}\sum_{i,j=1}^{|D|+n} x_{ij}^l (s_j - e_i - w_{ij}) + c_2 \sum_{d\in D}\sum_{k\in K}\sum_{l\in B_d^k}\sum_{i,j=1}^{|D|+n} x_{ij}^l w_{ij} \quad (1)$$

$$\text{s.t.} \quad \sum_{i=1}^{|D|+n} x_{ij}^l = \sum_{j=1}^{|D|+n} x_{ij}^l = 1, \forall l \in B_d^k, \forall k \in K, \forall d \in D \quad (2)$$

$$\sum_{d\in D}\sum_{k\in K} C_k |B_d^k| \leq B \quad (3)$$

$$\sum_{d\in D}\sum_{k\in K}\sum_{l\in B_d^k}\sum_{i,j=1}^{|D|+n} e_{ij}^{lo} x_{ij}^l (e_i - s_i + w_{ij}) \leq U_o, \forall o \in O \quad (4)$$

$$\sum_{k\in K} |B_d^k| \leq r_d, \forall d \in D \quad (5)$$

$$\sum_{i=p}^{q} y_{pq}^l (e_i - s_i + w_{i(i+1)}) < wt_k, \forall l \in B_d^k, \forall p,q \in F_d^l \quad (6)$$

$$e_{q-1} + w_{(q-1)q} + ot_k \leq s_q, \forall l \in B_d^k, \forall p,q \in F_d^l, \exists y_{pq}^l = 1 \quad (7)$$

where c_0, c_1 and c_2 denote, respectively, degree of importance assigned to the capital costs for buses, total waiting time, and deadhead time; C_k denotes the price of kth type of vehicles;

o denotes the set of air pollutants emitted from buses. For each air pollutant $o \in O$, let U_o be the upper boundary on the amount of o emitted and e_{ij}^{lo} be the amount of pollutant $o \in O$ emitted per hour when some type of a vehicle makes a feasible trip pair (T_i, T_j). wt_k and ot_k denote working time and the time for an oil change of kth type of vehicle.

$$x_{ij}^l = \begin{cases} 1 & \text{if trip } i \text{ is run immediately before trip } j \\ & \text{by the vehicle } l, \\ 0 & \text{otherwise,} \end{cases}$$

$$y_{ij}^l = \begin{cases} 1 & \text{if vehicle } l \text{ can refuel at starting relief point of two} \\ & \text{adjacent trip } i \text{ and trip } j, \\ 0 & \text{otherwise,} \end{cases}$$

|x| denotes the number of elements in the collection x.

The objective (1) minimizes the sum cost of capital and operating costs. Constraints (2) express the flow conservation. Constraints (3) require that the capital costs are not more than a given budget. Constraints (4) ensure the amounts of each air pollutant emitted by all the buses do not exceed a given boundary. Constraints (5) ensure that backup vehicles dispatched from the depot cannot exceed a given number. Constraints (6) guarantee that a vehicle has adequate fuel to cover a series of trips between two adjacent fueling points. Furthermore, constraints (7) guarantee that there is enough time for an oil change at its starting relief point before a trip by a vehicle.

Application of ACO on Bus Scheduling

Ant colony optimization was a heuristic algorithm introduced by Dorigo et al. (1996) to simulate the foraging behavior of real ants. The basic idea (Dorigo and Stützle 2004) is that a large number of artificial ants cooperate in finding good solutions to difficult discrete optimization problems in three procedures: 1) build solutions by making use of pheromone trails and heuristic information, 2) update pheromone trails, and 3) implement centralized actions which cannot be performed by single ants. Successfully ACO has demonstrated in many applications, especially combinatorial optimization problems like vehicle routing (Li and Tian 2008), traveling salesman problem (Dorigo et al., 1996; Ji et al., 2009), and graph-coloring (Costa and Hertz, 1997, Salari and Eshghi, 2005).

In this paper, ACO is adopted in solving bus scheduling, and new heuristic information and a feasible solution construction method are proposed in algorithm according to the features of the problem. Some changes are induced as follows:

Solution Construction. Solutions are constructed by iteratively assigning trips to vehicles. All trips are firstly sorted in increasing order of their starting time, and all ants are initially placed on trip 1. At each construction step an ant chooses the next still unassigned trip by turns and assigns it to a feasible vehicle using the action choice rule of ACO. Care is taken in assigning each trip only to a feasible vehicle, i.e., the trip is assigned to a new vehicle only if no feasible one exists, generating in this way a feasible assignment.

- **Pheromone trails**. The pheromone trail τ_{il} represents the desirability of assigning trip i to vehicle l.
- **Heuristic information**. Besides pheromone trails, heuristic function is an important guide when assigning a trip to a vehicle. Maximum utilization of vehicles is the goal in this paper. Obviously, if the vehicle l travels from trip i (last trip executed by vehicle l at present) to trip j directly, the shorter deadhead time w_{ij} and the shorter waiting time s_j-e_i-w_{ij} are preferred. The heuristic function takes the form as: $\eta_{il} = \frac{1}{\psi \cdot w_{ij} + \delta \cdot (s_j - e_i - w_{ij})}$, where ψ and δ denote waiting and deadhead time's degree of importance, respectively.

- **Transition rules**. The state transition rule in ACO is a pseudo-random-proportional-rule. The ant k positioned at trip i will be assigned to vehicle j determined by the Eq. (8).

$$j = \begin{cases} \arg\max_{l \in N_i^k} \{\tau_{il} \mid \eta_{il} \mid^{\beta}\}, if \quad q \le q_0 \\ J, \qquad\qquad\qquad otherwise \end{cases} \tag{8}$$

J is a random variable selected according to the below probability distribution formulation.

$$p_{ij}^k = \frac{[\tau_{ij}]^{\alpha}[\eta_{ij}]^{\beta}}{\sum_{l \in N_i^k}[\tau_{il}]^{\alpha}[\eta_{il}]^{\beta}} \text{ , if } j \in N_i^k \tag{9}$$

where q_0 is a constant and q is a random variable uniformly distributed in [0,1]. α and β, respectively determine the relative influence of the pheromone trail and the heuristic information. N_i^k is the feasible neighborhood of ant k when being at trip i, that is, the set of vehicles that can be assigned to trip i.

Update of Pheromone Trail. After ants have constructed their solutions, the pheromone trails, which give a positive feedback mechanism of ACO to trigger the discovery of good solutions, are updated. It's critical to design good pheromone updating rules.

- **Global pheromone trail update**. Only best-so-far ant is allowed to update pheromone after each iteration, including both evaporation and new pheromone deposit. Thus, the search focuses very quickly around best tour when a strong additional reinforcement to the arcs belonging to the best solution found since the start of the algorithm is adopted in this way. The update is implemented by the following equation:

$$\tau_{ij} \leftarrow (1-\rho)\tau_{ij} + \rho Q_1 / C^{bs} \text{ , } \forall (i,j) \in T^{bs} \tag{10}$$

where $0 < \rho < 1$, and Q_1 are parameters. C^{bs} is the length of the tour T^{bs} built by the best-so-far ant.

- **Local pheromone prail update**. In addition to the global pheromone trail updating rule, an ant uses a local pheromone update rule to reduce its pheromone trail τ_{ij} of an arc (i,j) after immediately having crossed an arc (i,j) during the tour construction so that the arc becomes less desirable for the following ants. In other words, this preserves the diversity of exploring arcs, and algorithm may not show a stagnation behavior. The local rule is given by equation as follows:

$$\tau_{ij} \leftarrow (1-\xi)\tau_{ij} + \xi Q_2 / C^{nn} \tag{11}$$

where $0 < \xi < 1$, and Q_2 are two parameters. C^{nn} denotes the length of a nearest-neighbor tour.

Such an update strategy of pheromone trail was easily verified by Dorigo and Stützle that the pheromone trails can never have a value higher than Q_1 / C^{bs} and lower than Q_2 / C^{nn}, i.e., $\forall (i,j):\ Q_2 / C^{nn} \le \tau_{ij} \le Q_1 / C^{bs}$. Hence, it have the effect of limiting the probability of assigning trip *i* to vehicle *j* to the interval [p_{min},p_{max}] to avoid a search stagnation (Dorigo and Stützle 2004).

Pseudo-code for ACO. The process of using ACO to resolve bus scheduling problem is described in pseudo-code in Figure 1.

Computational Example

To verify our study's rationality of existence, we use ACO to solve the above problem in matlab7.0, and optimal algorithm parameter setting is experimentally found that we take a colony of 10 ants, 200 steps of iterative procedure, q_0=0.95, α=1, β=2, $\rho = \xi$=0.1 , Q_1=200, O_2=50, c_0=100, $\delta = c_1$=1.5, and $\psi = c_2$=2.5.The details of test instances are shown in Tables 1–4 as follows: there are a total of 249 trips assigned to a fleet of vehicles consisting of two types of vehicles, which are dispatched from three depots with the same capacity restrictions r_1= r_2= r_3=15 and bought with a limited budget B= ￥10000000. Assume that some strict standards are imposed on vehicle emissions including particulate matter (PM, U_{PM} =5kg), nitrogen oxides (NO_x, U_{NOx} =50 kg), carbon dioxide (CO_2, U_{CO2}=7000 kg). Average emission rate of each type of vehicle can be determined by using some measuring techniques.

```
Procedure ACO for bus scheduling
    Initialize parameters and set pheromone trails
    Sort trips by their start-time
    While(termination condition not met) Do
       Place each ant on initial trip
       For ant=1:AmountOfAnts
            ConstructAntsSolutions
            LocalPheromoneTrailUpdate
       End
       GlobalPheromoneTrailUpdate
    End
EndProcedure
```

Figure 1. Procedure of ACO to resolve bus scheduling.

Table 1. Vehicle information.

Vehicle type	Price	Runtime	Time for an oil change	NO_x	PM	CO_2
A	￥300000	450 min	15 min	0.26kg/60min	0.003kg/60min	25.5kg/60min
B	￥400000	480 min	20 min	0.05kg/60min	0.001kg/60min	21.9kg/60min

Table 2. Route information.

Route	Working time	Interval	Start-point	End-point	Runtime	Vehicle type
1	7:20-21:00	20 min	1	3	60	A or B
2	6:45-20:30	15 min	2	3	55	A or B
3	7:00-20:30	15 min	3	1	65	B
4	7:00-21:00	20 min	3	2	60	A or B
5	7:30-20:30	15 min	1	2	70	A

Table 3. Deadhead time information.

Departure depot	Arrival depot	Deadhead time	Departure depot	Arrival depot	Deadhead time
1	2	15 min	2	3	20 min
1	3	20 min	3	1	20 min
2	1	20 min	3	2	25 min

Table 4. Timetable information**.**

Trip	Departure time	Route	Trip	Departure time	Route
1	7:20	1	…	…	3
2	7:40	1	152	20:15	3
…	…	1	153	20:30	3
41	20:40	1	154	7:00	4
42	21:00	1	155	7:20	4
43	6:45	2	…	…	4
44	7:00	2	195	20:40	4
…	…	2	196	21:00	4
97	20:15	2	197	7:30	5
98	20:30	2	…	…	5
99	7:00	3	248	20:15	5
100	7:15	3	249	20:30	5

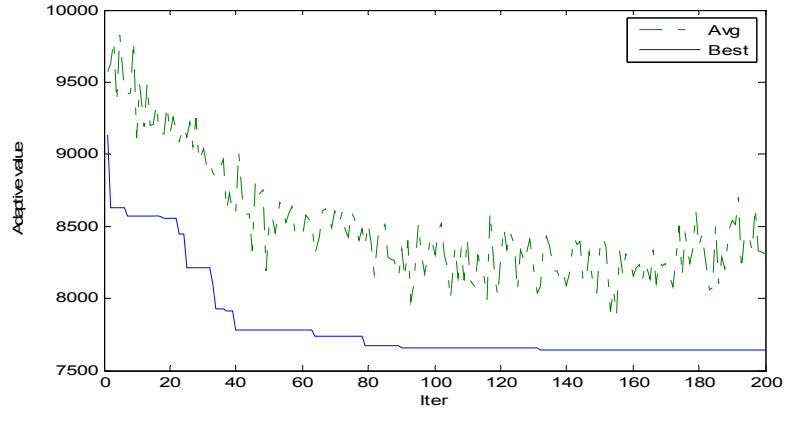


Figure 2. The optimal solution and its mean value of the iterative process.

Table 5. The results of vehicles arrangement.

No.	Vehicle routes	Vehicle type
1	43-156-51-162-59-118-18-127(fuel)-26-138- deadhead -144-37-194	B
2	44-103- deadhead -112-13-170- deadhead -128-25-140(fuel)-34-191	B
3	99	B
4	154-48-160-56-166-64-172-72-132(fuel)-28-185- deadhead -148-40	B
5	45-104-7-164- deadhead -120-19-176-79(fuel)-183-87-146-39-196	B
6	100- 4-161- deadhead -116-16-173-74-134(fuel)-30-143- deadhead -149-41	B
7	1-158- deadhead -206- deadhead -212-66- deadhead -222(fuel)- deadhead -228-82- deadhead -238- deadhead -244-98	A
8	155-50-109-11-168-67-174- deadhead -133(fuel)-29-142-36-151	B
9	46-105-8-114-15-123- deadhead -129-27(fuel)-139- deadhead -145-38-195	B
10	101-5-110-12-119- deadhead -125-23-135(fuel)- deadhead -141-35-192	B
11	197- deadhead -203-57- deadhead -213- deadhead -219-73- deadhead -229(fuel)-83-186- deadhead -243-97	A
12	2-159-55-165-63-171-71-177(fuel)- deadhead -231- 85- deadhead -241-95	A
13	47-106-9-115- deadhead -121-20-130- deadhead -137(fuel)-32-189- deadhead -152	B
14	102-6-111- deadhead -117-17-126-24-182(fuel)- deadhead -33-147- deadhead -153	B
15	198- deadhead -204-58- deadhead -214- deadhead -220- deadhead -178(fuel)- 80-184-88-190-96	A
16	3-107- deadhead -113-14-122-21-131	B
17	157-52-163-60-169-68-175-76-136(fuel)-31-188- deadhead -150-42	B
18	199-53- deadhead -209- deadhead -215-69- deadhead -225- deadhead -181(fuel)-84-187-92-193	A
19	49-108-10-167- deadhead -124-22-179	B
20	200-54- deadhead -210- deadhead -216-70- deadhead -226- deadhead -232(fuel)-86- deadhead -242- deadhead -248	A
21	201- deadhead -207-61- deadhead -217- deadhead -223-77- deadhead -233(fuel)- deadhead -239-93- deadhead -249	A
22	202- deadhead -208-62- deadhead -218- deadhead -224-78- deadhead -234(fuel)- deadhead -240-94	A
23	205- deadhead -211-65- deadhead -221-75-180	A
24	227-81- deadhead -237-91- deadhead -247	A
25	230- deadhead -236-90- deadhead -246	A
26	235-89- deadhead -245	A

The optimal solution is that the bus company will spend ¥9200000 in buying 12 vehicles belonging to vehicle type A (12 vehicles initially located at depot 1, 0 vehicles initially located at depot 2, and 0 vehicles initially located at depot 3) and 14

vehicles belonging to vehicle type B (1 vehicle initially located at depot 1, 6 vehicles initially located at depot 2, and 7 vehicles initially located at depot 3) assigned to cover all trips through refueling 18 times. These vehicles will have a total of 1110 minutes waiting time, 1320 minutes deadhead time and 15465 minutes working time, and give off 0.53 kg PM, 40.01 kg NO_x and 6572.60 kg CO_2 to pollute the atmosphere. The results of vehicle arrangements are in Table 5 and the curves of best tours found by ACO are illustrated in Figure 2.

Conclusions

Bus scheduling is a fundamental issue of concern in public transit management. In this paper, we apply ACO to resolve bus scheduling problems and reduce spending when purchasing buses to cover all trips to maximize utilization of vehicles when they meet all constraints such as fuel restrictions, multiple depots and their capacities etc, and new heuristic information and a feasible solution construction method are proposed in algorithm according to the features of the problem. Numerical experiments show that ACO is more efficient and robust in solving this problem. However, our study doesn't consider some important real-world aspects regarding unexpected events, such as traffic jams, traffic accidents, and vehicle accidents in the existing formulations and solution approaches. The original schedule may be deteriorated to the extent in this situation where the fleet plan may need to be adjusted in real-time depending on the current state of what is certainly a dynamic system. Further enhancements to the formulations used in this research and developing more efficient heuristic approaches for solving the problem will be another important area for future research.

Acknowledgments

This work was supported by High-Technology Research and Development Program of China (863 program) under contract No. 2007AA11Z201 and National Science Foundation of China under contract No. 50878089. These supports are gratefully acknowledged. However, all facts, conclusions, and opinions expressed here are solely the responsibility of the authors. The authors thank the anonymous referee for her/his helpful comments and so on.

References

Bodin, L., and Golden, B. (1981). "Classification in vehicle routing and scheduling," *Networks*, 97-108.

Costa, D., and Hertz, A. (1997). "Ants can colour graphs," *Journal of the Operational Research Society*, 295-305.

Dorigo, M., Maniezzo, V., and Colorni, A. (1996). "The ant system: optimization by a colony of cooperating agent," *IEEE Transactions on Systems, Man and Cybernetics Part B*, 29-41.

Dorigo, M., and Stutzle, T. (2004). "Ant colony optimization," *The MIT Press*, 37-38.

Gintner, V., Kliewer, N., and Suhl, L. (2005). "Solving large multiple-depot multiple-vehilce-tyoe bus scheduling problems in practice," *OR Spectrum*, 507-523.

Haghani, A., and Banihashemi, M. (2002). "Heuristic approaches for solving large-scale bus transit vehicle scheduling problem with route time constraints," *Transportation Research Part A*, 309–333.

Ji, J., Zhen, H., and Liu, C. (2009). "A fast ant colony optimization algorithm for traveling salesman problems," *Computer Research and Development*, 968-978.

Kliewer, N., Mellouli, T., and Suhl, L. (2006). "A time-space network based exact optimization model for multi-depot bus scheduling," *European Journal of Operational Research*, 1616-1627.

Liu, Z. G (2007). "Regional bus operation bi-level programming model integrating timetabling and vehicle scheduling," *System Engineering Theory and Practice*, 135-141.

Li, X. Y., and Tian, P. (2008). "Research on ant colony optimization algorithm for the open vehicle routing problem," *System Engineering Theory and Practice*, 81-93.

Li, J.-Q., and Head, K. L. (2009). "Sustainability provisions in the bus-scheduling problem," *Transportation Research Part D*, 50-60.

Li , J.-Q., Mirchandani, P. B., and Borenstein, D. (2009)." A Lagrangian heuristic for the real-time vehicle rescheduling problem," *Transportation Research Part E*, 419-433.

Mes, M., van der Heijden, M., and van Harten, A. (2007). "Comparison of agent-based scheduling to look-ahead heuristics for real-time transportation problems," *European Journal of Operational Research*, 59-75.

Prakash. S., Balaji, B. V., and Deepak, T. (1999). "Optimizing dead mileage in urban bus routes through a nondominated solution approach," *European Journal of Operational Research*, 465-473.

Salari, E., and Eshghi, K. (2005). "An ACO algorithm for graph coloring problem," *2005 ICSC Congress on Computational Intelligence Methods and Applications*, 22-27.

Wang, W., Shen, J., and Liu, Z. (2006). "Ant colony algorithm for transit vehicle scheduling problem with route time constraint," *ICTTS*, 509-516.

Wang, H., and Shen, J. (2007). "Heuristic approaches for solving transit vehicle scheduling problem with route and fueling time constraints," *Applied Mathematics and Computation*, 1237-1249.

A Corridor-based Earthquake Evacuation Model

Fei Song[1], Yang Wang[2] and Xiaokuan Yang[3]

[1] Graduate Student, [2] Graduate Student, [3] Professor, Transportation Research Center, Beijing University of Technology, China; Email: songfei@emails.bjut.edu.cn, wangyang8265@emails.bjut.edu.cn, xiaokuan@bjut.edu.cn

Abstract

The earthquake happened in Wenchuan, Sichuan, province on May 12, 2008, resulted in a tremendous loss of property and lives. Emergency evacuation and transfer are so important that government should give them high priority. The emergency evacuation corridor planning will be more significant for restoration of disaster areas. This paper proposes a methodology for the carrying capability of regional earthquake evacuation corridors based on macro-simulation and to prove the availability and rationality of a predetermined regional earthquake evacuation plan. The proposed approach includes three processes: 1) Estimate the evacuation demand of earthquake affected area; 2) Confirm the OD-pairs based on the location of shelter and the preplanned direction of evacuation; 3) Build up the earthquake evacuation network through the macro-simulation software OREMS as well as input the outcome of procedures 1 and 2. Meanwhile, the carrying capability of preplanned regional earthquake evacuation corridor through evacuation clearance time and average evacuation speed are the output of OREMS. In the end, the proposed approach was employed in the Beijing case and proved the availability and rationality of Beijing earthquake evacuation corridor planning. Criterion for selecting the earthquake evacuation road and the methodology for searching availability and rationality of earthquake evacuation corridor planning are proposed in this paper.

Key Word: earthquake evacuation; corridor-based; macro-simulation

Introduction

Earthquake evacuation has been studied frequently all over the world. The huge earthquake disaster challenged the capability of the government to handle emergencies, especially after the earthquake occurred in the Wenchuan, Sichuan, province on May 12, 2008. Moreover, in light of the fact the survival rate had a significant relationship with evacuation time, shown in the data of earthquake in Tangshan, the emergency evacuation planning will be more significant for rescue and restoration of a disaster area.

The earthquake evacuation corridor was different from the traditional transportation system that serves the normal traffic demand. The reason is that subways and railroads are possibly damaged during earthquake. Even though these facilities are still in place, they can't be used for evacuation due to safety concerns. Second, due to the fact that the capacity of ferry (boat) or helicopter is limited and the geographic conditions for them are constrained, these transportation means are seldom considered in the process of an evacuation plan. Therefore, the surface transportation system seems much more flexible and prone to sustainability. It is evident that the surface transportation system is very important for earthquake evacuation. It is necessary for all levels of government to predetermine earthquake evacuation plans and to evaluate the availability and rationality of regional earthquake evacuation corridor planning.

In review of previous studies, it appears that most studies of emergency evacuation technology are concentrated on two aspects: study on methodology of emergency evacuation planning and study on management strategy of emergency evacuation traffic. Meanwhile, evacuation planning includes three parts, such as evacuation demand estimate, evacuation distribution estimate, and evacuation-related traffic assignment. Current practice in hurricane evacuation demand estimation involves the estimation of the total evacuation demand in the first step followed by the estimation of its temporal distribution in the second. A study by Fu (2007) proposed two ways of estimation of hurricane evacuation demand. Participation rate models are the primary method to estimate evacuation demand, and response curves are used for the temporal distribution. The sequential logit model is the other way to provide estimates of dynamic disaggregate travel demand for hurricane evacuation. The study demonstrated the superiority of the sequential logit model over the traditional two-step method in modeling hurricane evacuation travel demand. Evacuation distribution estimation also involves an aggregate model and disaggregate model (Cheng and Wilmot 2008). However, there was still no case study to prove the accuracy of these two models.

In general, evacuation efficiency can be improved by building the programming model to optimize the destination and route assignments in the network. An evacuation assignment model could be divided into two categories, static and dynamic, according to the evacuation demand and impedance function that can be either time-dependent or not (Elise 2008). Moreover, the dynamic assignment model can be used not only in evacuation planning but also in improving the management strategy of evacuation traffic (Wei 2008; Liu et al. 2008). Due to the fact that the dynamic evacuation traffic assignment model needs a more high-powered computer, it is hard to apply this model in the actual urban road network.

Although there was some research on management strategy of emergency evacuation traffic, these research almost always focuses on the effectiveness of a single strategy in the situation of hurricane evacuation (Sisiopiku 2007; Wolshon and Marchive 2007). The research lacked a case study on developing an earthquake evacuation corridor. Criterion for selecting the earthquake evacuation road and the methodology for searching availability and rationality of earthquake evacuation corridor planning are proposed in this paper.

Principle and Methodology

Principle in the Selection of Earthquake Evacuation Corridor. According to the relative rule of law, the earthquake evacuation corridor must be provided with high capacity and high anti-seismic ability. The earthquake evacuation corridor selection is dependent on the following attributes:

(1) Safety and flexibility: In case of earthquakes, some routes of road networks usually become unavailable. The corridor selected for evacuation must have a strong ability to withstand earthquake damage. Evacuation routes should not cross an earthquake fault zone or should have a strong seismic grade. As a result, primary roads in the city and high-grade highways, or state roads are chosen as the evacuation corridor. Moreover, alternative channels are needed if part of the evacuation routes are destroyed by the earthquake.

(2) High capacity: Once the earthquake occurs, the sphere of influence is large, and there may be many disaster victims who are concentrated in one area. A large number of relief supplies are needed. Therefore, the evacuation corridor should be able to withstand a high level of service and traffic capacity.

(3) Integrate with urban network planning: Earthquake evacuation corridor should integrate with urban network planning, which can adapt to changes in the city's population and raise the defense capabilities of the city as soon as possible.

(4) Convenient for surveillance: Since an evacuation corridor should be monitored, the roadways with monitoring facilities or other ITS facilities should be selected as a priority.

(5) Adaptable either during normal conditions or emergencies: The evacuation corridor itself should be a part of the urban status road network. Evacuation routes should be used for daily traffic in non-emergency situations.

Methodology of Assess the Carrying Earthquake Evacuation Corridor.

Macro-simulation model (OREMS). The critical step after an earthquake is to transfer people to the shelter quickly via the predetermined earthquake evacuation corridor. This paper used the macro-simulation model (OREMS) which is based on the UE (User Equilibrium) principle to optimize the earthquake evacuation procedure. This objective model could be mathematically described as follows:

Objective:

$$\min Z(x) = \sum_a \int_0^{x_a} t_a(\omega)\, d\omega$$

s. t.

$$\sum_k f_k^{rs} = q_{rs} \qquad \forall r,s$$

$$f_k^{rs} \geq 0 \qquad \forall k,r,s$$

$$x_a = \sum_r \sum_s \sum_k f_k^{rs} \delta_{a,k}^{rs} \qquad \forall a$$

$$t_a(x_a) = t_a^0 \left[1 + \alpha \left(\frac{x_a}{C_a} \right)^{\beta} \right]$$

where,

$Z(x)$—the objective function, the sum of the integrals of the link performance functions;

x_a— flow on link a, vehicle;

$t_a(x_a)$—impedance function, minute;

t_a^0—travel time on link a in free speed, minute;

f_k^{rs}—flow on path k connecting O-D pair $r-s$, vehicle;

q_{rs}—flow rate between origin r and destination s, veh/hr;

$\delta_{a,k}^{rs}$—indicator variable;

$$\delta_{a,k}^{rs} = \begin{cases} 1 & \text{if link } a \text{ is on path } k \text{ between O-D pair } r\text{-}s \\ 0 & \text{otherwise} \end{cases};$$

C_a—capacity of link a, veh/hr; and

α, β—pending coefficient.

Algorithm of optimization model. The macro-simulation model of an earthquake evacuation traffic assignment is a nonlinear programming model with linear

constraints. The convex combination algorithm is especially useful for determining the equilibrium flows for transportation networks. The algorithm itself, when applied to the solution of the UE problem, can be summarized in the following steps:

- Initialization. Perform all-or-nothing assignment based on $t_a = t_a(0), \forall a.$ This yields $\{x_a^1\}$. Set counter $n = 1$.
- Update. Set $t_a^n = t_a(x_a^n), \forall a.$
- Direction finding. Perform all-or-nothing assignment based on $\{t_a^n\}$. This yields a set of flows $\{y_a^n\}$.
- Line search. Find α_n that solves $\min_{0 \le \alpha \le 1} \sum_a \int_0^{x_a^n + \alpha(y_a^n - x_a^n)} t_a(\omega) d\omega$
- Move. Set $x_a^{n+1} = x_a^n + \alpha_n (y_a^n - x_a^n), \forall a.$
- Convergence test. If a convergence criterion is met, stop (the current solution, $\{x_a^{n+1}\}$, is the set of equilibrium link flows); otherwise, set $n = n + 1$ and go to Step 1.

Through the convex combination algorithm, we can receive the best solution of the evacuation traffic assignment model. The outputs of macro-simulation model (OREMS) include evacuation time and evacuation speed. The availability and rationality of a preplanned earthquake evacuation corridor could be assessed by these simulation results.

Case Study

Objective Evacuation Network. Statistics Yearbook of Beijing 2004 shows that the population density in the central district of Beijing is 42 times greater than the average population density in all of Beijing. This data implied that the central district of Beijing is more prone to earthquakes. In this paper, a central urban area is identified as the study area.

Earthquake Evacuation OD. Beijing statistics show that cars per capita was 0.2. Automobile occupancy was predicted 4 persons/vehicle during earthquake evacuation. The capacity of a bus used in earthquake evacuation was 180 persons/bus. The bus would be converted to equivalent car. The conversion coefficient was 2.5. Based on these rules, the earthquake evacuation demand expressed as the number of equivalent car, which was shown in Table 1.

According to the relative rules regarding emergency shelter planning, recommended area per capita of earthquake emergency shelters was 2.0 m^2. Earthquake shelters include all parks and stadiums in the central district of Beijing. The capacity constraint of shelters is also expressed as the number of equivalent car (Table 2).

Table 1. Earthquake evacuation demand in central district of Beijing.

District	Population (million)	Evacuation demand (vehicle)
Xicheng	0.83	168 306
Dongcheng	0.69	139 917
Chongwen	0.41	83 139
Xuanwu	0.61	123 694
Total	2.54	515 056

Table 2. The capacity constrain of earthquake shelter.

District	Park	Green area (m^2)	60% Green area (m^2)	Capacity (person)	Capacity (pcu)
Xicheng	7	1 400 906	840 543.9	420 270	85 220
Dongcheng	7	1 318 798	791 278.8	395 638	80 224
Chongwen	4	2 349 379.5	1 409 627.7	704 813	142 920
Xuanwu	4	449 244	269 546.4	134 772	27 327
Other	11	3 535 938.5	2 121 563.1	1 060 777	215 101
Total	33	9 054 266	5 432 559.9	2 716 270	550 792

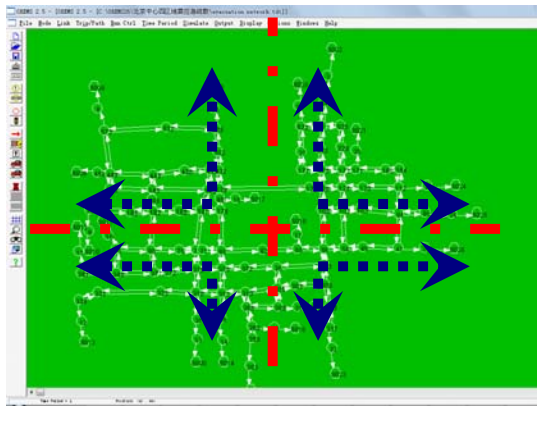


Figure 1. Earthquake evacuation direction.

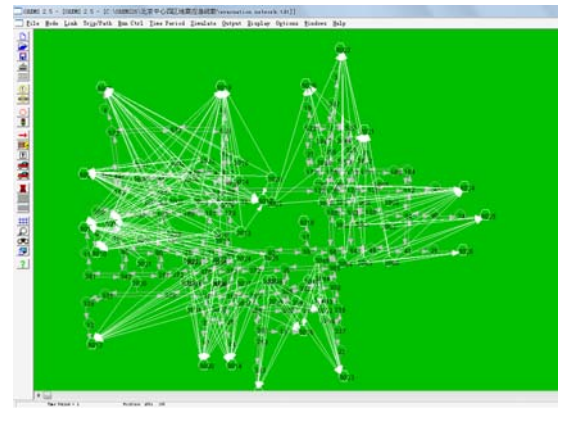


Figure 2. Evacuation demand distribution.

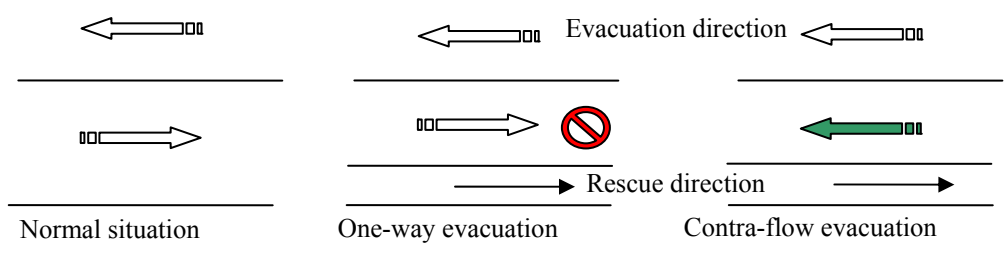


Figure 3. One-way evacuation and contraflow evacuation.

Table 3. Analysis result of carrying capability of evacuation corridor.

Evacuation strategy	Evacuation time	Average evacuation speed
One-way evacuation	4:02 hours	24.6 mile/hour
Contra flow evacuation	2:33 hours	32.7 mile/hour

Validation Analysis of Carrying Capability. After setting up the evacuation network and inputting the evacuation demand OD by the OREMS, the objective area was divided into four quadrants by Chang Avenue and the middle Axial Road. This paper adopts two evacuation strategies, one-way and contra-flow, to evacuate people to the shelter. Contra-flow evacuation strategies means that the opposite direction road can also be used for evacuation, which can improve the ability to evacuate. However, the traffic management is difficult when it comes to the end of the road.

After applied the macro-simulation model, the result of analysis for the carrying capability of the earthquake evacuation corridor was shown in Table 3.

The output of OREMS shows that the carrying capability of earthquake evacuation corridor based on selection principle fulfilled the earthquake evacuation demand. For the same network, the evacuation time is relatively less in contra-flow evacuation strategy. It would improve the evacuation efficiency significantly and reduce loss of life caused by the earthquake.

Although contra-flow evacuation strategy could achieve good evacuation results, there are three critical things that should be accomplished before an actual evacuation is needed. (1) A venue for the community to get on evacuation vehicles must be preplanned. (2) The traffic management people have to be assigned to the key intersection to regulate the evacuation traffic. (3) The traffic in reverse lanes must be clear before contra-flow evacuation begins. The evacuation time obtained from the proposed methodology does not include this mobilization time. Forecast the length of the mobilization time needed for further research. Notifying the public about proper evacuation routes is also important for the success of earthquake evacuation.

Conclusions

Earthquake evacuation modeling is researched relatively less than transportation modeling. However, it's important to make the emergency evacuation corridor plan for restoration of a disaster area. This paper proposed a corridor-based earthquake evacuation model through the macro-simulation software, OREMS, especially for Beijing.

This paper examines a set of principles to select a suitable earthquake evacuation corridor and proposes a methodology for the research on the carrying capability of a regional earthquake evacuation corridor based on macro-simulation to prove the availability and rationality of a preplanned regional earthquake evacuation corridor. In the end, the proposed approach was employed in the Beijing case study and proved the availability and rationality of Beijing earthquake evacuation corridor planning. The evacuation time and average evacuation speed can be obtained in the emergency simulation, and the result shows that the contra-flow strategies are are more effective during an evacuation There are still some inadequacies in this paper. This paper does not discuss the effect by the time-dependent variation of evacuation demand. Therefore, future study in this area is crucial for transportation safety during an emergency.

References

Cheng, G. X., and Wilmot, C.G. (2008). "A destination choice model for hurricane evacuation." *Annual Meeting of the Transportation Research Board*, Washington, D.C., 08-0840.

Elise, M. H. (2008). "The maximal dynamic expected flows problem for emergency evacuation planning." *Annual Meeting of the Transportation Research Board*, Washington, D.C., 08-2924.

Fu, H.Q. (2007). "Static vs. dynamic and aggregate vs. disaggregate: a comparison between practice and research in hurricane evacuation travel demand modeling." *Annual Meeting of the Transportation Research Board*, Washington, D.C., 07-0646.

Liu, Y., Chang, G. L., and Liu, Y. (2008). "A corridor-based emergency evacuation system for Washington D.C." System Development and Case Study. *Annual Meeting of the Transportation Research Board*, Washington, D.C., 08-0674.

Sisiopiku, V.P. (2007). "Application of traffic simulation modeling for improved emergency preparedness planning." *Journal of Urban Planning and Development.* 133(1).

Wei, H., Zeng, Q. A, and Hu, H. (2008). "Integrated urban evacuation planning framework for responding to man-made disasters over surface transportation network." *Annual Meeting of the Transportation Research Board*, Washington, D.C., 08-1112.

Wolshon, B., and Marchive, E. (2007). "Emergency Planning in the urban-wildland interface: subdivision-level analysis of wildfire evacuations." *Journal of Urban Planning and Development*. 133(1).

Designing and Implementing a Service Oriented Traveler Information System

Jun Liu[1], Dongmei Lin[2]and Chunyan Li

[1] Prof., MOE Key Laboratory for Urban Traffic Complexity System Theory and Technology, Beijing Jiaotong University, Beijing 100044, P.R.China, Email: jliu@bjtu.edu.cn
[2] PhD candidate, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, P.R.China, E-mail: lindongmei@gmail.com

Abstract

The objective of the research was to develop an advanced traveler information system framework that supported individual requirement for travelers. Advanced traveler information system is related to comprehensive sources and diversified forms of traffic data distributed in heterogeneous systems. The difficulty of data integration of heterogeneous systems demands a proper method to solve the problem. Due to SOA's contribution to the integration of heterogeneous systems, the paper focuses on the designing and implementation of SOA-based traveler information system. This system uses a Service Oriented Architecture (SOA) that contains four roles. These include: a service registry, an integrating platform, service requesters and service providers. Then four key technologies are proposed, including requirement identification, service matching, service composition and service delivery. Furthermore, the paper describes the process of the system and a way of realizing the system by the application of Web Service. Finally, we implement a traveler information system for the city of Beijing. It uses the Web service technology to encapsulate real-time traffic service, route planning service and map inquiry service; then a new service is composed of real-time traffic data and route planning data by applying service composition algorithm. The result is displayed with a GIS map on mobile software. The architecture of this system not only improves the efficiency of system interoperability, but also provides a platform that can effectively realize service integration when needed. Therefore, it is an appropriate way of designing a traveler information system.

Key words: ITS; ATIS; traveler information system; SOA; web service; service composition

Introduction

With the development of social economy, the requirement of travelers for the quality of traveling and information services is getting increasingly higher. Advanced Traveler Information Systems (ATIS) have become an efficient way of reducing traffic congestion and improving travel efficiency, especially metropolises of China.

In recent years, information technologies have been applied to a set of management information systems in the areas of highway, railway, aviation and urban transport, which as well provide travelers some traffic information. However, many of the traveler information systems cannot really be classified as advanced since they have a high reliance on manually derived and anecdotal data (McQueen, 2002). These existing systems either, to some extend, provide traveler information or are the source of traffic data, hence they should be properly integrated when designing an advanced traveler information system.

ATIS is, at its core, a means to effectively combine and coordinate a variety of transport modes, with the goal of improving traffic management and service level. To satisfy the individual requirements, ATIS provides travelers multiple traffic information services including pre-trip route planning, ticket planning, traffic-related information access, en-route guidance, etc. Limited to administrative mechanism, existing systems usually provide services restricted to a single transport mode. Although some systems can combine bus and rail transit services, the comprehensive, individual requirements are still not satisfied. To meet the various needs of travelers, an advanced traveler information system is related to comprehensive sources and diversified forms of traffic data distributed in heterogeneous systems, as mentioned above; it must acquire, interpret and convert the various data. The difficulty of data processing demands a proper method to solve the problem.

The wide application of Service Oriented Architecture (SOA) technology in the interoperability of heterogeneous systems has shed light on the design of ATIS. In this paper, we designed and implemented a SOA-based traveler information system.

Service Oriented Architecture

There are several definitions of SOA. Here we quote what CBDI defines. CBDI (Sprott, 2005) defines SOA as the policies, practices, frameworks that enable application functionality to be provided and consumed as sets of services published at a granularity relevant to the service consumer. Services can be invoked, published and discovered, and are abstracted away from the implementation using a single, standards-based form of interface. We can see from the definition that SOA has the characteristic of loose coupling. Loose coupling enables service requester to have no knowledge of the technical details of the provider's implementation, such as the programming language, deployment platform, and so forth. The service requester typically invokes operations by way of messages—a request message and the response—rather than through the use of APIs or file formats.

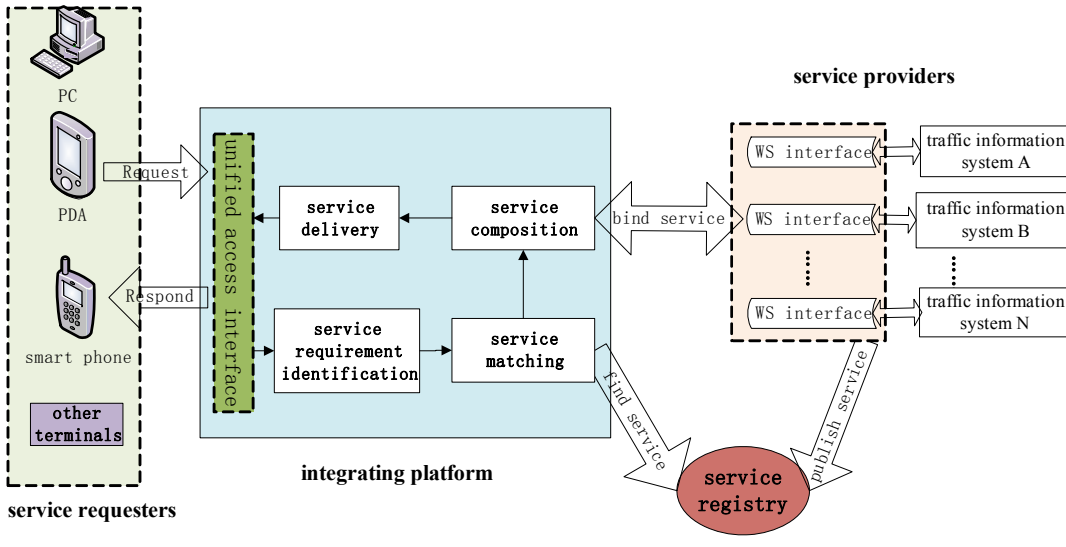


Figure 1. SOA-based architecture of traveler information system.

Due to SOA's contribution to the integration of heterogeneous systems, it is a suitable architectural approach to developing advanced traveler information system.

System Architecture

This traveler information system architecture consists of four roles: service providers, a service registry, an integrating platform and service requesters. In the following representation (Figure 1), service providers are responsible for providing basic functions. Figure 1 also shows the architecture of the SOA-based integrating platform.

Service Provider. Any information system that provides traffic information service(s) makes it a service provider as long as the services are encapsulated as Web Service (WS) interfaces. Some traffic information services have already been developed. As in Figure 1, existing providers have the ability to publish services to the service registry, and then services can be found and invoked by users and other application programs.

Service Registry. In this architecture, service registry is relatively an independent system, with the purpose of effectively managing different kinds of traffic information services. With this registry, service providers can publish their services, and service requesters can search services that they need. Therefore, service registry is the core of this system architecture that supports the integrating platform.

Integrating Platform. However, while these existing services can provide basic functions, they cannot meet complex travelers' needs. There is a requirement gap between existent traffic information services and travelers' individual requirements. Therefore, the integrating platform (hereinafter to be referred as platform) provides comprehensive services to travelers. It collects data from service providers, generate

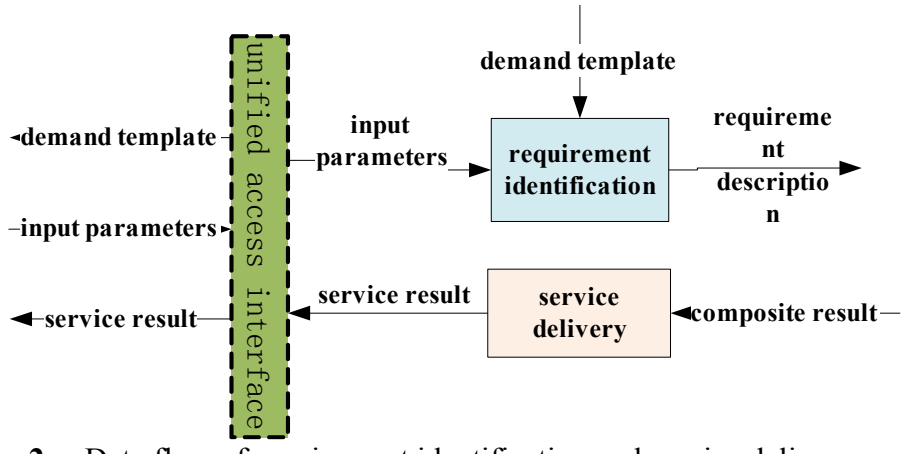


Figure 2. Data flow of requirement identification and service delivery modules.

new services for travelers, and in this case the platform serves as a service requester to other providers and in the meantime a service provider to travelers. Travelers can access the platform through several terminals. They can reach the platform by either website or WS interface. On website, a unified input interface is given, and service results are displayed respectively by category. Via WS interfaces, the platform can be accessed through the use of PDAs, smart phones and GPS navigators. Simple needs may be satisfied by basic functions from existent systems, while complicated needs be met by, as mentioned above, newly generated services. Regardless of where the services are originally located, all services seem running in the same platform, which is the biggest benefit for users.

Service Requester. Service requesters are the final users of this architecture, travelers who need traffic information. Using any supported terminals they send service requests to the platform. After four processes (Figure 1), the platform analyzes the demands and finally delivers requesters the service result that can meet their needs.

Key Technologies and Methods of the System

In addition to the role issue, the heterogeneous system integration poses another problem. To generate new services that satisfy travelers' individual requirements, the platform performs as the essential part of this architecture. This paper proposed are four key technologies in the system and their functions, namely requirement identification, service matching, service composition and service delivery.

Requirement Identification. Requirement identification is the link between travelers and the platform. It recognizes travelers' needs by analyzing the features, and figures the mapping relation out between travelers' requirements and traffic information services. With the help of an XML-based demand template of requirement identification (Zhu, 2009), the platform can match traveler's individual needs to existent services or composite services that can be generated by the platform. As a result, an XML-based requirement description is formed. Figure 2 presents the data flow of this module.

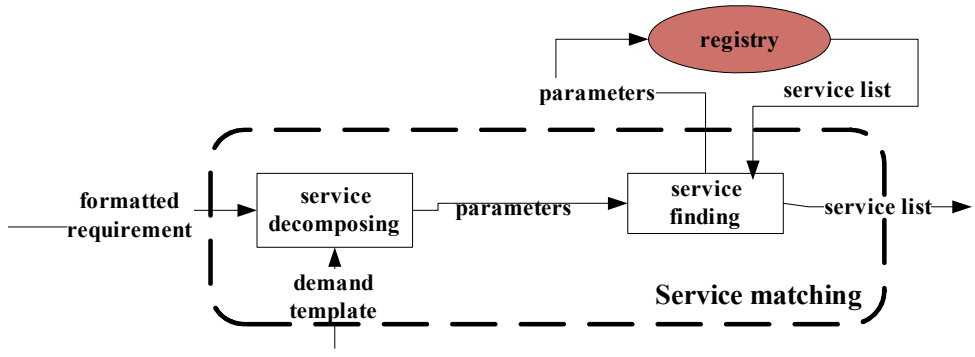


Figure 3. Data flow diagram of service matching module.

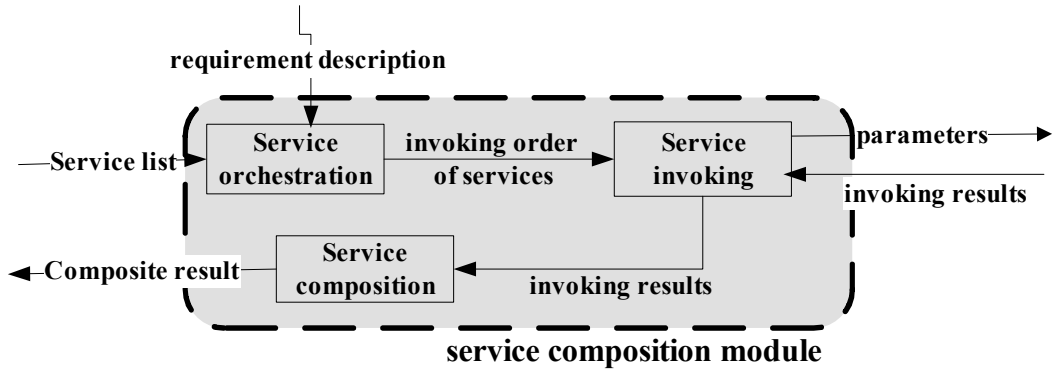


Figure 4. Data flow diagram of service composition module.

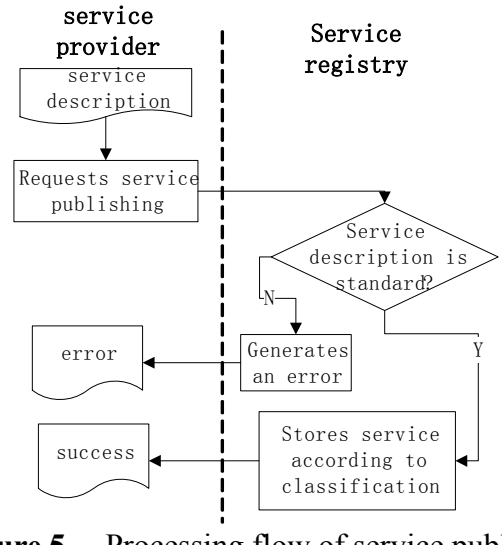


Figure 5. Processing flow of service publishing.

Service Matching. After the process of requirement identification, the platform continues its work by locating the services. To find, select and locate service providers is what service matching does. This technology focuses on how to describe, register and discover a traffic information service; it achieves the matching of traffic information services and data resources.

As mentioned in the introduction chapter, travelers' requirements diversify so that they may not be satisfied with basic services. Here in service matching module, it firstly decomposes the comprehensive needs into basic functions that constitute a complicated function, according to the Classification List of Traffic Information Services stored in service registry. Then it gets a list of who provide each basic function from the registry. At last it selects at least one provider of each basic function from alternative providers. Hence this module is divided into two parts – service decomposing and service finding; Figure 3 describes the data flow of service matching.

Service Composition. Service composition is concerned with designing a combination solution based on a set of existing services. As for travelers' complicated needs, they are designed to be achieved by composite services. Its role is to achieve individual needs of travelers with composite services. The paper provides a composition method based on Web services. This module is responsible for three major tasks. (1) Service orchestration. According to the result of service matching as well as travelers' requirement, it arranges the order of basic services that should be invoked and prepares the environment and parameters. (2) Service invoking. The platform invokes services in a predefined order, and gets the service results. (3) Service composition. After first two tasks, necessary basic data are acquired. The platform composes these data to form a brand new service which matches an individual requirement. Figure 4 shows the data flow of service composition module.

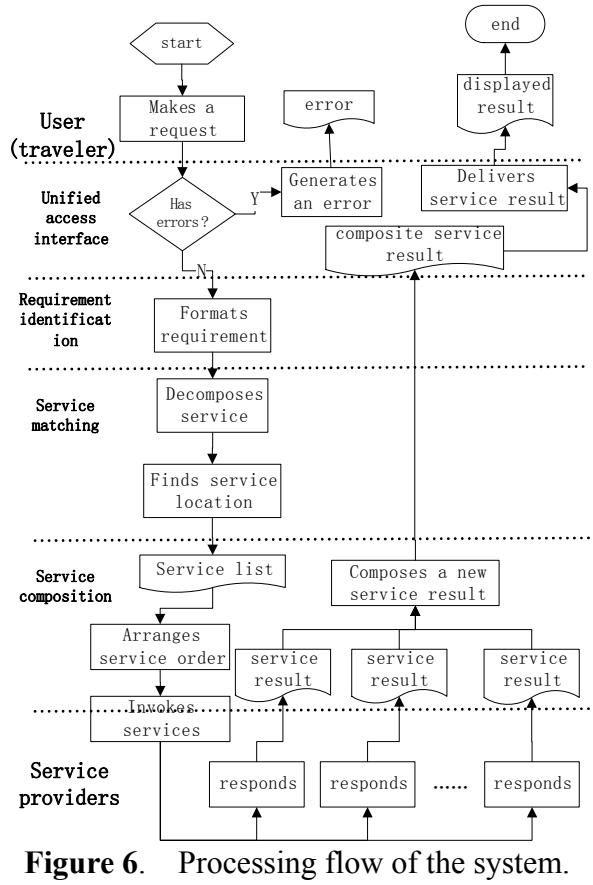


Figure 6. Processing flow of the system.

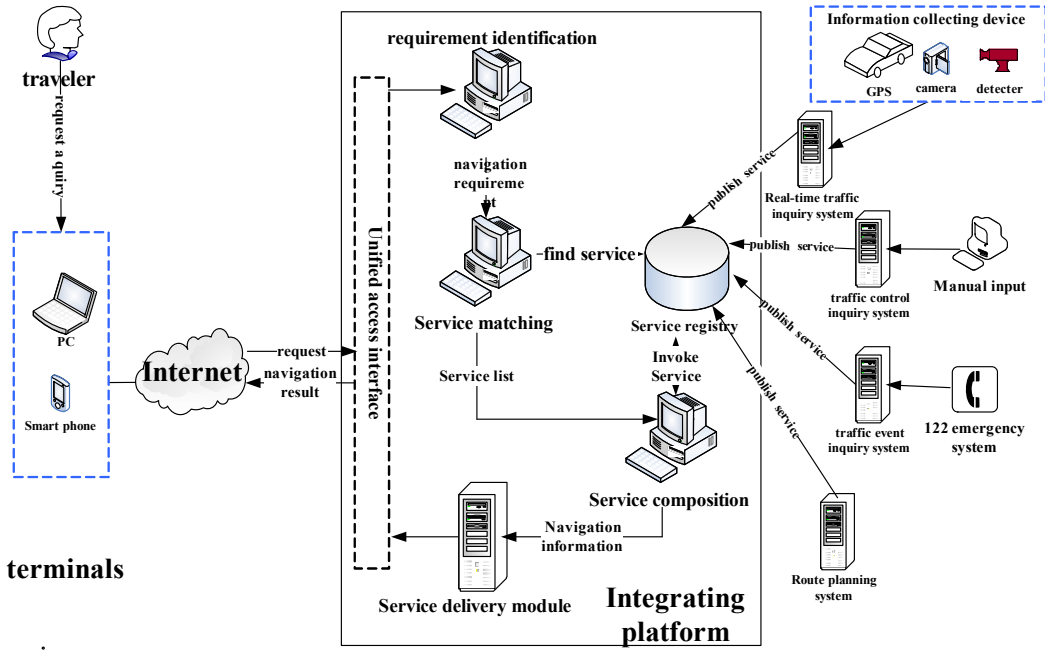


Figure 7. System structure.

Service Delivery. To achieve a seamless delivery of traffic information, an approach to the unified service description and access is the essence of service delivery. A recommended specification of data format is given in Zhu's article (Zhu, 2009), and we suggest any terminal developers follow the format. Though this paper does not discuss the specific method, it is worth further researching.

Processing Flow. In addition to four key technologies, the processing flow of the system is also proposed. Two processes are described in Figures 5 and 6. One the process of service publishing, and the other is the process of how this system works when a user makes a request.

System Implementation

According to the portrayed architecture, an advanced traveler information system for Beijing is implemented. It contains a website and a mobile software to provide traffic information, and it covers all parts of the architecture (Figure 7). This system is realized with the application of Web Service; every module is encapsulated with web service interface so they communicate through XML documents.

The system makes use of basic functions provided by existent systems. GPS positioning is provided by "7fish" positioning system, routing planning is offered by "Sogou" map system, and real-time traffic information is from a system of STONE ITS company. When a traveler makes a request through mobile phone to acquire the real-time traffic information, the platform gets the request, invokes related services, composes a result with basic services, sends the result back to the phone, and finally displays the result on the phone (as in Figure. 8).



Figure 8. Real-time traffic information displayed on mobile software

The successful implementation of this system indicates that the architecture is a reasonable solution of integrating traffic services and designing advanced traveler information system.

Result and Conclusion

The traffic information system presented in this study is a service-oriented architecture whose mission is to improve flexibility of traveler information service. This system has been designed and implemented. The kernel of this architecture is based on web services technology, allowing integrating platform to deploy new types of traffic services. As traveler's terminals send service requests to the platform, the system invokes basic services to acquire detailed information. The information is then parsed by the platform and composes new types of traffic services.

For servers this represents a system for enhancing ductility in the traffic systems that have been developed. In this study, architecture proposed with the advantages of SOA technology allows for the reuse of traffic services. Whenever a new business requirement needs to be implemented, the system can adjust the existing systems to fulfill the new demand.

This research has successfully demonstrated a traveler information system using Service Oriented Architecture. As new systems need to be applied, the architecture can support the deployment process effectively. The deployment management system not only makes service delivery more effective, but also reduces the gap between the traffic systems and individual needs. The system is expected to enhance the quality of advanced services and become a reference model, as a convenient means of service integration.

Acknowledgements

The research was funded by Science and Technology of China, Program for National High-Technology (863) Research (2007AA11Z220).

References

Colan, M. (2004). *Service-Oriented Architecture expands the vision of Web services*, *Part* 1. http://www.ibm.com/developerworks/library/ws-soaintro.html*?S_TACT=105AGX52&S_CMP=cn-a-ws*#N10053.

Erl, T. (2005). *Service-Oriented Architecture: Concepts, Technology, and Design*, Upper Saddle River, NJ: Prentice Hall PTR.

McQueen, B., Schuman, R., and Chen, K. (2002). *Advanced Traveler Information Systems*, Norwood, MA: Artech House, 200-202.

Sprott, D., and Wilkes, L. (2004). *Understanding Service-Oriented Architecture*, CBDI Forum, https://www.alaska.edu/oit/eas/ea/Microsoft-MSDNguidelines/MSDN-NET-Architecture-Center-Service-Oriented-Architecture-Unde.pdf.

Zhu, L., Chen, X., and Liu, J. (2009). “XML-based Public Traveler Information Service Demands Identification Template.” *Journal of Transport Information and Safety,* 2009, 27(4), 50-54.

Mobile Sensors as Traffic Probes: Addressing Transportation Modeling and Privacy Protection in an Integrated Framework

Xuegang (Jeff) Ban[1] and Marco Gruteser[2]
[1] Department of Civil and Environmental Engineering, Rensselaer Polytechnic Institute (RPI), 110 Eighth Street, Room JEC 4034, Troy, NY 12180-3590; Email: banx@rpi.edu
[2] WINLAB / Electrical & Computer Engineering, Rutgers University, Tech Centre of New Jersey, 671 Route 1 South, North Brunswick, NJ 08902; Email: gruteser@winlab.rutgers.edu

Abstract

Mobile traffic sensor data, once widely available, will significantly enhance current transportation modeling applications such as arterial performance measurement. Receiving and processing mobile sensor data however may involve severe privacy concerns if not properly designed. In contrast to the fact that the current research on transportation modeling and privacy protection is rather separated, we propose in this paper a framework on privacy-aware transportation modeling (PATM) and application-aware privacy protection (AAPP). The proposed framework focuses on the interactions between transportation modeling and privacy preserving, being aware of privacy when developing transportation models as well as application needs when designing privacy preserving mechanisms. Using two case studies, we show how PATM and AAPP may be applied in privacy-preserving mobile data collection while satisfying the application needs at the same time. The paper is concluded by discussing how to design a unified approach to guarantee privacy and data needs for various applications.
Key Words: Transportation Modeling; Traffic Modeling; Privacy Protection; Location Privacy; Mobile Traffic Sensors; GPS Cellular Phones; Virtual Trip Lines; Arterial Performance Measurement; OD Estimation;

Introduction and Motivation

Mobile Traffic. *sensors* refer to those that move with the traffic flow. They are in contrast to fixed-location sensors like loop detectors that dominant current traffic

detection system. Broadly speaking, mobile traffic sensors include any monitoring or data collection system with a dedicated device equipped with vehicles. In this sense, they include for example probe vehicles (such as those equipped with Electric Toll Collection (ETC) tags), cellular phones, portable GPS devices (like GPS cellular phones), GPS navigation systems, Bluetooth Mac Address Matching (BMAM, see Wasson et al. 2008), vehicles in Vehicle Infrastructure Integration (VII, now called IntelliDrive (2009)), etc. Most mobile sensors need to communicate with either satellites (GPS) or cellular towers (cell phones) or dedicated roadside infrastructures (ETC, VII, BMAM) to derive the mobile component's position, speeds, and other relevant information.

In this paper, we focus on mobile sensors that can provide detailed tracking capability including GPS and cellular techniques. The current penetration of these mobile sensors is low (about 17% according to HarrisInteractive (2007)). As indicated by the International Telecommunication Union (ITU, 2009), the penetration of cell phones is nearly 100% in developed countries and nearly 50% in developing countries in 2007. Since most future cellular phones may be equipped with GPS, the high penetration of mobile sensors is expected in the near future. In addition, the implementation of IntelliDrive will make the high penetration of mobile sensors possible.

Mobile Sensor Data Collection in Transportation Applications. Mobile sensors can potentially provide detailed traces (trajectories) of individual vehicles, which contain rich information about traffic conditions especially when the data are widely available. As mobile data are usually *samples* of real traffic flow, traditional modeling techniques based on fixed-location sensor data such as flow, occupancy, etc., may not work directly. As a result novel modeling techniques are needed to use mobile data more effectively (Zhou, 2004; Ban et al. 2009; Ban and Hao, 2010). In the past issues related to deploy mobile sensors (such as coverage and penetration) and traffic modeling using mobile data have been investigated (Turner and Holdener, 1995; Quiroga and Bullock, 1998; Cheu et al. 2002; Qiu et al. 207; Smith and Fontaine, 2007). On the other hand, receiving and processing raw mobile trajectory data leads to severe privacy concerns (Congressional Record, 2001; Karger and Frankel, 1995; Agre, 1995; Warrior et al. 2003). The reality is that both agencies and private sector organizations who for various reasons have collected a large volume of trajectory data are extremely reluctant to share this information, and privacy protection is one of the main reasons. The first author personally experienced the difficulty of obtaining ETC data from transportation management agencies: due to privacy concerns, the data must be deleted within 24 hours. Trucking companies like FedEx are continuously monitoring the movement of their delivery trucks. These data however are not shared as they are considered valuable business property and might contain sensitive information about their drivers. As a result, almost no mobile data are now publicly available.

To protect privacy, it is a common practice to place restrictions on sharing data via policies or regulations. It is also often beneficial for organizations to avoid the

collection of personally identifiable information. Possession of such data leads to the following disadvantages. First, managing such data incurs overhead, at least for organizations that need to comply with privacy regulations. Examples include US government organizations covered under the Privacy Act of 1974 and businesses in the EU, which need to comply with the EU Data Directive 95/46/EC. These regulations, for example, require organizations to provide notice to data subjects, to allow access to records by data subjects, and to keep the data secure. Second, holding privacy sensitive data presents public relations and liability risks if such data is compromised. Privacy breaches occur frequently as public records indicate (Privacy Rights Clearinghouse, 2009), due to stolen laptops, dishonest insiders, or external hackers, among others. Third, even without such breaches, it is difficult to guarantee that the data will not be used for purposes other than the originally intended one. Data may be subpoenaed for civil cases or new owners of an organization may want to put the data to new uses. Travelers aware of such risks and concerned about their privacy may avoid services that collect personal data, thus leading to lower participation rates.

Due to these privacy concerns, several recent applications involved with mobile data collection have put more emphasis on privacy protection:

An AITS project originally led by researchers from Rensselaer Polytechnic Institute applied GPS devices to collect traffic and behavioral data from a group of recruited drivers (List et al. 2005). To protect privacy, trajectory data were *not* collected directly; instead the concept of *monuments* was developed, which are pre-defined locations to collect mobile sensor data such as travel times between two monuments on arterial streets (He et al. 2002; Demers et al. 2006).

Nokia recently developed a mobile data collection system based on GPS cellular phones (Herrera et al. 2009; Ban et al. 2009). Privacy protection is a major consideration of the system and the concept of virtual trip lines (VTL) was developed to represent pre-defined locations along roadways (Hoh et al. 2007). No trajectory information is collected by the system; instead only speeds at VTL locations are collected (travel times between VTLs may also be collected at sparse locations as discussed in Section 3).

VII (IntelliDrive) contains a subsystem on archiving probe vehicle data. Privacy protection is considered by encrypting the data transmitted between vehicles and roadside devices, as well as between the devices and data archival servers. Overall, the privacy issues of VII are addressed via nine privacy principles (Jacobson, 2007). The recent VII Proof of Concept Test (RTIA, 2008) concludes that privacy is "still an important issue and requires further evaluation and research." Especially, privacy needs to be ensured via the concept of "*privacy by design*" – to ensure privacy by designing the data collection system so that privacy sensitive information is not collected or revealed. This is *in contrast to* traditional mobile data collection practices for which data are collected first and then processed to prevent the release of privacy sensitive information.

The above discussions and existing mobile sensor related studies revealed that (1) mobile data are extremely valuable for traffic modeling and traveler information; (2) appropriate privacy protection mechanisms need to be in place in order to share large volume of mobile sensor data among transportation agencies, private sectors, academia, and the public; and (3) although there have been sparse investigations and implementations, the privacy issue has not been well addressed in the transportation community.

Location Privacy. Although past surveys indicated that the public has been indifferent to their location privacy, as the public is more aware of the adverse consequences of privacy leaks, they will be more concerned about protecting their privacy before sharing data (Krumm, 2008). Duckham and Kulik (2006) defined four general methods to ensure location privacy including *regulatory strategies, privacy policies, anonymity, and obfuscation.* The first two are related to policies, while the last two are on the computational (technical) aspects. The policy approaches, although effective in certain cases, tend to be *conservative* as they usually prohibit the release of information which may not be necessary. Also, since privacy issues are multiple-facet and complicated, they have already created many controversies in designing proper policies and regulations for protecting privacy (Solove, 2008). In this paper, we focus on technical approaches for privacy protection (i.e. anonymity and obfuscation) rather than policy-related mechanisms. This is called *location privacy* – which define quantitatively what privacy violation is and how to collect/process mobile data (Krumm, 2008). Anonymity aims to maintain the privacy of location data by replacing the associated name with an untraceable ID; Obfuscation means to degrade the quality of location measurements by aggregation, adding noise, etc. to reduce the possibility of privacy violation. Anonymity by simply removing identifiers from location data has been found not effective to guarantee privacy (Gruteser and Hoh, 2005). Hoh et al. (2007) further showed, using a dataset of week-long GPS traces from 239 drivers, that they were able to find home locations of 85% of a subset of 65 drivers. Therefore novel privacy schemes are usually needed to protect location privacy.

A Framework to Consider Privacy and Modeling Simultaneously. In summary, without proper considerations of privacy, travelers will be reluctant to share their location data; the public and private sectors will also be concerned due to liability reasons. Novel data collection schemes are needed to promote data sharing among different entities including public and private sectors, researchers, and the public. As a result, for privacy researchers, how to design privacy-preserving mobile data collection schemes is critical. As the mobile data quality may be compromised after applying privacy-protection techniques (e.g. data noise is added), this also calls for transportation researchers to develop new modeling approaches that can extract meaningful transportation knowledge based on privacy-preserving mobile data.

In this paper, we propose a modeling framework that emphasizes privacy-aware transportation modeling (PATM) techniques and application-aware privacy protection (AAPP) mechanism. The framework can hopefully address both

transportation application data needs and privacy protection. We present two specific examples on how the framework can be applied to transportation applications. The first example is for arterial performance measurement using VTL data. We show how crucial performance measures such as delays, queue lengths etc. for signalized intersections can be estimated using VTL data. The second example is regarding how to improve freeway origin-destination (OD) demand estimation by fusing both detector data and privacy-preserving mobile sensor data. We conclude by further discussing how to apply the proposed framework to other types of transportation applications.

Privacy-aware Transportation Modeling and Application-aware Privacy Protection

Transportation modeling and privacy protection have been studied separately so far. On the one hand, with the primary goal being extracting information as much as possible, transportation researchers are aggressive in acquiring information: they tend to assume available data to the finest detail and consider privacy protection as a restricting factor for them to obtain "detailed" data. Developing privacy protection techniques, on the other hand, ensures policy makers, public agencies, private sectors, researchers, and the driving public that individual privacy is protected. This promotes sharing of mobile data: the public feels that their sensitive information will not be released and agencies/private sectors/researchers are more comfortable to share data as they have the mechanism to ensure privacy. However privacy experts have focused on protecting individual's privacy, while mostly using very basic models of application data requirements, which may lead to the unnecessary prohibition of the release of some insensitive data.

As a result, there needs a fundamental change to the current thinking and modeling practices of transportation applications and privacy preserving. That is, we need a new paradigm that facilitates the close collaboration between privacy experts and transportation researchers to develop a holistic framework to (1) design effective privacy protection mechanisms, and (2) meanwhile to develop novel modeling approaches to extract information/knowledge using privacy-preserving mobile data. We denote this paradigm privacy-aware transportation modeling (PATM) and application-aware privacy preserving (AAPP), as depicted in Figure 1.

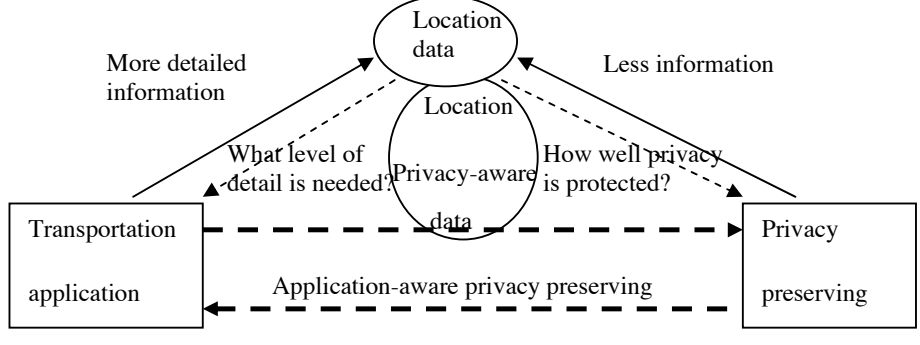


Figure 1. A Holistic View for Privacy Protection and Transportation Modeling.

Figure 1 depicts the traditional way of transportation modeling and privacy preserving (on the top) and the newly proposed holistic modeling framework (on the bottom). From the traditional perspective, with one pushing for more information on the modeling side and another prohibiting the release of information from the privacy preserving side, consensus can hardly be reached on which level of detail the data should be released. To address these challenges, our proposed approach focuses on the interactions between transportation modeling and privacy preserving, being aware of privacy when developing transportation models as well as application needs when designing privacy preserving mechanisms. As a result, both privacy protection and application needs can be satisfied to the maximum extent possible.

The challenge of the newly developed PATM and AAPP framework is that as privacy protection mechanisms will likely compromise the original mobile data, privacy experts need to know what is the best way to do so and transportation researchers need to develop novel transportation modeling techniques for knowledge/information extraction based on the privacy preserving mobile data. While knowledge extraction and privacy preserving do have seemingly conflicting objectives, the authors believe that one can satisfy both in most cases using innovative designs for privacy-protection and modeling. This is due to the fact that privacy does not necessarily need to be compromised to satisfy application data requirements: An application-aware design of privacy algorithms can retain features important for the application, while still achieving privacy by removing features that are less important. For example, cloaking techniques to depersonalize location data can achieve the same level of privacy by retaining high spatial resolution and reducing temporal resolution or by providing high temporal resolution and reducing spatial resolution (Gruteser and Grunwald, 2003). Similarly, as shown in previous research (Ban et al. 2009; Ban and Hao, 2010), in order to extract transportation information like crucial traffic measures (such as real time queue length or delay at a signalized intersection), it is NOT necessary to know the detailed trajectory of every vehicle. Rather, some aggregated mobile data (such as individual travel times between an upstream and a downstream location of the intersection) can suffice, which imposes little privacy concerns (Hoh et al. 2008).

As will be presented in the next two sections, in the current practice, the proposed scheme can be generally initiated from the transportation modeling side. However one has to keep in mind that both sides need to be considered. The challenge is to reach a point that both objectives are met in the best possible way without compromising the other party severely. For this purpose, an iterative process is necessary.

Virtual Trip Lines for Arterial Performance Measurement

The first case study is based on the mobile sensor data collection system developed by Nokia that collects virtual trip lines (VTL) data. VTLs are pre-defined (and *virtual*) locations on roadways. When crossing a VTL, a vehicle equipped with mobile sensors will report its speed. As shown in Hoh et al. (2008), using VTL to

regulate location and speed reports reduces privacy concerns, compared to probe vehicle systems which periodically transmit location reports. Privacy is increased because VTLs can be placed into less privacy sensitive areas and areas where traffic information is most important (such as major highways or arterial intersections). When location updates are anonymized and trip lines are placed with sufficient spacing, it also becomes difficult to track the path of an individual vehicle for an extended duration. Thus this approach avoids the collection of potentially sensitive trip information and in particular trip endpoint information.

Arterial modeling has for long suffered from insufficient data coverage. Mobile devices such as GPS cellular phones provide a great potential for low-cost, wide-area data collection of arterials. For arterial performance measurement, the key is the performance of signalized intersections as they contribute to the majority of the delay on arterial streets. With both data collection and privacy protection in mind, PATM and AAPP can be naturally applied to extend the VTL approach. The process can be generally initiated from the application side. First, transportation researchers can examine the original VTL data as described in Hoh et al. (2008) which only contains speed information and a trip line identifier, and the placement of VTLs does not allow tracking of a vehicle across multiple VTLs. They then conclude that it is very difficult to obtain intersection performance measures based on the original VTL data. Transportation researchers can further examine the data needs for estimation arterial performances and find that ravel time or delay information of individual vehicles passing an intersection can be used to estimate intersection delay patterns, arrival volumes, and queue lengths. They subsequently introduce the requirement to obtain travel times across an intersection. The privacy experts then need to check whether collecting intersection travel times could severely impact privacy. In general, privacy increases if an adversary cannot determine the complete path of a traveler. Since cars have to follow roadways, on roadway segments confusion is only possible if the spacing of trip lines is sufficiently large so that the order of cars changes in between with high probability. At intersections, however, privacy can increase significantly if it is not possible to determine whether a car took a turn or traveled straight ahead. While introducing travel time monitoring across an intersection reveals turn information for cars, a high degree of privacy can still be preserved if a small subset of intersections is chosen for monitoring so that any given vehicle trip still passes through other unmonitored intersections with high probability. As a result, collecting travel times for major intersections only does preserve high level of privacy violation. At this point, a consensus is reached between transportation researchers and privacy experts that travel times of an intersection can be collected that can be used for arterial intersection performance measurement purpose without severely violate privacy.

The VTL case study in this section serves to illustrate how PATM & AAPP can be designed to provide both adequate privacy protection and data quality for applications. The process is in general iterative before a proper balance can be reached.

Intersection modeling based on VTL travel times are discussed in detail in Ban et al. (2009) and Ban and Hao (2010). For the sake of completeness, we briefly present the major findings here, focusing on how to estimate intersection arrival volume which was not discussed previously in detail. As intersection delays are usually a function of incoming flows and signal timing settings, it is possible to reconstruct flow and signal timing information from measured travel times or delays. This however requires a reverse thinking process compared with traditional arterial modeling methods based on loop detectors: one needs to reconstruct traffic states (volume, queue length, etc.) from travel times while traditional methods all use traffic volume as an input (to estimate delays). The key is to identify discontinuities or nonsmoothness of intersection delays from sampled travel times, which indicate a traffic state change or signal status change.

In Ban et al. (2009), we showed that if the penetration of mobile sensors is sufficiently high (e.g. >= 20%), the time-dependent intersection delay pattern can be estimated via the sampled travel times. Figure 2(a) below depicts how this can be done. The figure shows a typical signalized intersection with two VTLs installed upstream (VTL1) and downstream (VTL2) respectively. Under the assumption that a queue never passes VTL1, we can use the bold solid triangles in the figure to represent how queue forms and dissipates based on shockwave theory (Lighthill and Whitham, 1955). The horizontal part of the triangles represents the duration of red time. If delays due to vehicle decelerations and accelerations are ignored and the arrival rate is uniform within one cycle, delays can be fully determined by the triangles. In the figure, dashed lines represent trajectories of vehicles, while dotted lines are boundaries when the discontinuities of delays happen. We aim to characterize vehicle delays as a function of the time when it passes VTL1. Here we assume that data have been collected and thus one can perform "post-processing" to re-construct a mapping from the time that a vehicle passed VTL1 to its experienced delay at the intersection. As shown by the trajectories of vehicles (dashed lines), if a vehicle approaches the intersection in red time or if the queue length is not zero (e.g. trajectory a in the figure), the vehicle will join the end of the queue first and thus be delayed. The delay encountered by the vehicle is the horizontal part of trajectory a. Otherwise, if a vehicle arrives during green time and there is no queue (e.g. trajectory b), the vehicle will pass the intersection with no delay. The (red) delay curve at the bottom of Figure 2(a) will spike up at the time that allows a vehicle to travel to the intersection in free flow just before the start of the red time. More importantly, by analyzing the geometry of the triangles, one can observe that if a vehicle passes by VTL1 at a time that would make it get to the intersection just after the start of the red time, delay for this vehicle will be the maximum for the specific cycle. After that, delays will be reduced linearly until no delay is reached. This is represented by the line segments marked as "1" of the delay curve at the bottom of Figure 2(a). The slope of the delay reduction part, denoted as delay reduction rate s, can be calculated analytically as (Ban et al. 2009):

$$s = \frac{u_f (w - u_w)}{w(u_f + u_w)} = \frac{v}{k_f}\left(\frac{1}{u_f} + \frac{1}{w}\right) - 1 \tag{1}$$

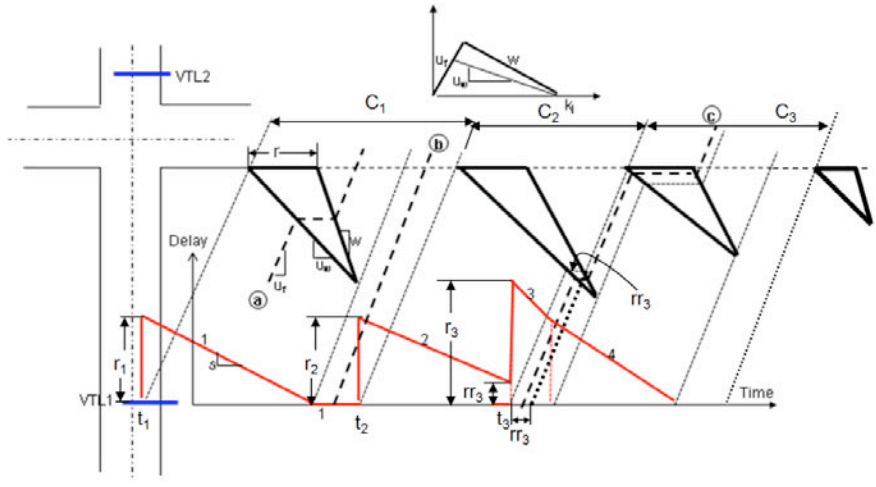


Figure 2(a). Intersection Delay Pattern.

Here w is the shock wave speed, uf is the free flow speed, uw is the wave speed when a vehicle joins the queue, kj is the jam density, and v is traffic flow which is assumed to be constant within a cycle. The three parameters uf, w, kj are specific to actual arterial locations, which also determine the fundamental diagram of the location. Notice that since w >= uw always hold (refer to the fundamental diagram at the top of Figure 2(a)), s is positive, meaning that delay always reduces from its maximum (when traffic light turns red) to some minimum value (when light turns green and no queue exists) for normal situations.

In Ban et al. (2009a), we further developed a two-step estimation method to construct the delay pattern (red lines at the bottom of Figure 2(a)) using sampled travel times. The delay pattern turns out to be critical as it can be used to further estimate traffic volume and queue length. Equation (1) clearly shows that if the delay pattern can be estimated, the delay reduction rate s is known. Then the only unknown in (1) is the traffic volume v, which can be calculated as follows:

$$v = \frac{(s+1)k_f}{\frac{1}{u_f} + \frac{1}{w}} \tag{2}$$

It was found in Ban et al. (2009) that under relatively high penetration (>20%), VTL data can be used to estimate the signalized intersection delay patterns relatively accurately. Under even higher penetration (>60%), the signal timing parameters (like cycle length, duration of red and green times, etc.) can also be derived. Here we present the estimation results of intersection arrival volumes using VTL data based on equation (2). Figure 2(b) depicts the estimated volume (dashed line with '+' signs) and observed volume (solid line with '.' signs) for a signalized intersection in micro-simulation (Ban et al. 2009). Although significant discrepancies between the two lines can be observed, we still see that the estimated volume can follow the trend of the obse ed volume fairly closely. If we aggregated volumes over 5-minute, 15-

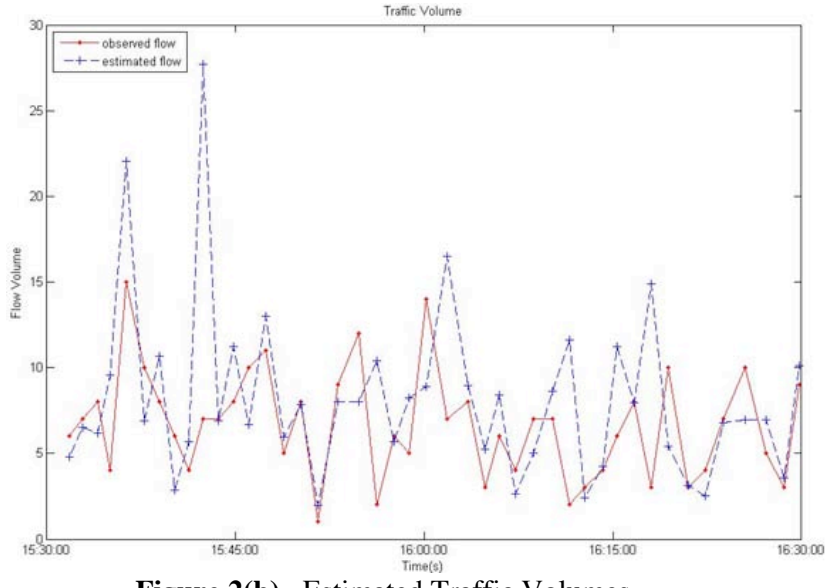


Figure 2(b). Estimated Traffic Volumes.

minute, 30-minute, and the entire 1-hour period, the estimation accuracy can be much improved. The percentage root mean square errors (PRMSE) are 29.5%, 12.8%, 7.4%, and 8.9% respectively. The equation for computing PRMSE is listed below:

$$PRMSE = \frac{|\hat{v} - \bar{v}|}{\bar{v}} \times 100 \quad . \tag{3}$$

Here $\hat{v}$ and $\bar{v}$ are the estimated and observed volumes respectively for an appropriate period of time. As shown in Ban and Hao (2010), the time-dependent queue length of signalized intersections can also be estimated using VTL data. This shows the potential of using privacy preserving mobile data to estimate real time performances of signalized arterial intersections.

Mobile Data for Freeway Origin-Destination Demand Estimation

Most existing OD demand estimation methods are based on fixed locations sensors like loop detectors, which have been extensively studied for both the static case (Cascetta and Nyugen, 1988; Bell, 1991; Yang et al. 1992) and the dynamic case (Cascetta et al. 1993; Chang and Tao, 1996; Zhou and Mahmassani, 2007; Nie and Zhang, 2008). Here we illustrate how privacy-preserving mobile sensor data can be used to enhance existing OD estimation results. A well-recognized problem of using traffic counts from loop detectors to estimate OD demands is that the resulting model is usually under-determined: i.e. the number of unknown OD variables is far more than the number of locations (detector stations) for colleting traffic counts. Correspondingly, there are usually multiple OD demand patterns that can produce exactly the same set of observed traffic counts. We illustrate this using the example of estimating freeway (static) OD demands as shown in Figure 3. The figure depicts

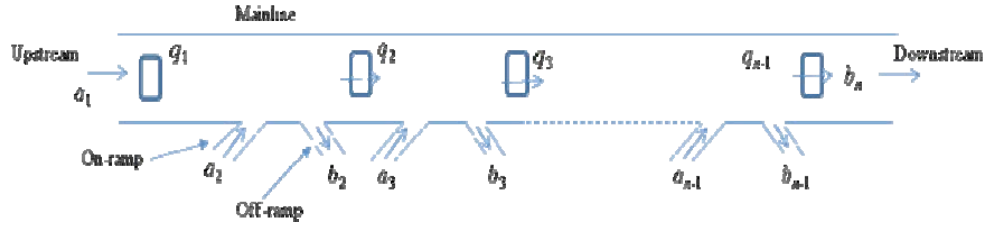


Figure 3. Freeway OD Estimation.

a freeway segment with n ramp locations (the first location denotes the upstream freeway mainline and the last location is the downstream freeway mainline). Denote ai the inflow and bi the exit flow for $1 \le i \le n$, and qi is the mainline volume between location i and i+1 for $1 \le i \le n-1$. Assume xi,j is the demand from location i to location j for $1 \le i \le n, 1 \le j \le n$. As this is a freeway segment (i.e. one way), we have $x_{ij} = 0$ for all $i \ge j$. The OD estimation problem is to determine xi,j given ai, bi, and qi.

We should point out here that this static version of the freeway OD estimation problem is rather preliminary as more sophisticated dynamic models have been proposed in the literature (Chang and Wu, 1994). Therefore the problem is selected mainly to illustrate the concept in this paper. A simple optimization model for the freeway OD estimation problem can be given as below:

$$\min_{x_{ij}} \; z = \sum_{i=1}^{n-1} \left| \sum_{l=1}^{i} \sum_{j=i+1}^{n} x_{lj} - q_i \right| \qquad (4\text{-}1)$$

$$a_i = \sum_{j=i+1}^{n} x_{ij}, i = 1,2,\cdots,n-1 \qquad (4\text{-}2)$$

$$b_j = \sum_{i=1}^{j-1} x_{ij}, j = 2,\cdots,n \qquad (4\text{-}3)$$

$$x_{ij} \ge 0, i = 1,2,\cdots,n; j = i+1,\cdots,n \qquad (4\text{-}4)$$

Here the objective function is the absolute deviation from the observed mainline traffic counts (qi) and the estimated traffic counts from OD demands $\sum_{l=1}^{i} \sum_{j=i+1}^{n} x_{lj}$. Equation (4-2) is the constraint that the inflow at location i must exit at downstream locations (i+1 to n); equation (4-3) is the constraint that the exit flow at location j must come from upstream locations (1 to j-1). The model is clearly a linear programming problem by noticing that the absolute value in (4-1) can be easily converted to linear terms by introducing extra variables. The model has n(n-1)/2 unknowns for xi,j and 3n-5 observed values for ai, bi, and qi. If n is large, the number of unknowns is larger than the number of observed values – in which case the model is under-determined. For example, here we assume n=4 and the observed values are given in Table 1. It turns out that there are multiple optimal solutions that can exactly match the observed values (i.e. the optimal objective value in (4-1) is zero) and Table 1 shows two of them.

Table 1. OD Estimation Results.

						Res 1				Res2	
i	ai	bi	qi	1	2	3	4	1	2	3	4
1	3000		3000		500	1000	1500		500	1500	1000
2	100	500	3500			500	500				1000
3	500	1500	2500				500				500
4		2500									

The above example shows that traditional models based on traffic counts may not produce the "true" OD demand matrix. This can cause problems. Assume the true OD matrix is "Res1" but "Res2" is generated by the estimation model. In this case, due to the error of predicting OD demand, certain control strategies like ramp metering or pricing may fail due to excessive or fewer demands at specific locations of the freeway.

If mobile data are available, they can provide a sample of the ratio between x2,3 and x2,4. Assume the ratio is $\bar{\beta} = \bar{x}_{2,3} / \bar{x}_{2,4}$ where $\bar{x}_{2,3}$ and $\bar{x}_{2,4}$ are observed demands. The objective function of (4-1) may be revised as:

$$\min_{x_{ij}} \; z = \sum_{i=1}^{n-1} \left| \sum_{l=1}^{i} \sum_{j=i+1}^{n} x_{lj} - q_i \right| + w \left| x_{2,3} - \bar{\beta} x_{2,4} \right| \tag{5}$$

Clearly the revised objective is a weighted summation of the total deviation at mainline traffic count locations (the first term) and the deviation between x2,3 and x2,4 and their observed values. Assume the observed demands are representative, i.e. $\bar{\beta} = 1$ for this particular example (both x2,3 and x2,4 are 500 for "Res1" as shown in Table 1). The revised model generates exactly "Res1" indicating that the model performs better than model (3). In reality, $\bar{\beta}$ may not be completely representative.

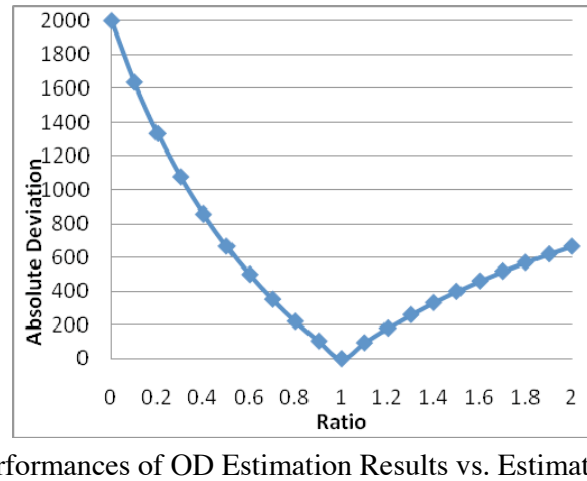


Figure 4. Performances of OD Estimation Results vs. Estimate of Ratio

Figure 4 shows how the results vary based on different $\overline{\beta}$'s. It can be seen from the figure that even the estimates of $\overline{\beta}$ have significant errors, the model results are still acceptable. For example, for $0.6 \le \overline{\beta} \le 1.5$ (i.e. $\overline{\beta}$ is estimated from 40% lower to 50% higher than the true value), the absolute error of the estimated demand matrix from (4) and the true OD ("Res1" in Table 1) is less than 450, 10% from the total OD demands.

The example above shows that by combining aggregated mobile sensor data (i.e. the ratio of demands for different OD pairs) with traditional traffic counts, OD estimation results can be significantly improved. To obtain the aggregated ratio information, one does not need detailed vehicle trajectory data. This provides more flexibility for designing privacy protection mechanisms. One option is to apply random perturbation to the OD pair reported by each vehicle and then apply privacy preserving data mining techniques to reconstruct the original distribution of OD values (Agrawal and Srikant, 2000). If portions of the trajectory should be preserved also, a random path swapping (RPS) technique can be applied to the mobile data collection process. RPS randomly decides to switch (or not switch) the non-overlapping sections of the paths of two randomly selected individual traces between the same OD pair. This can be depicted Figure 5. If the OD zones are large, such a scheme can provide a basic level of privacy because (1) it provides "misleading" privacy information which can confuse an privacy adversary, and (2) even for the same individual, the adjusted traces over multiple days will unlikely to form any pattern (or if any pattern is formed, it should be very different from the true pattern), which may not be used by the adversary to identify the individual. RPS however does not modify a large portion of the route and does not change the total traffic demand for any OD pair so that $\overline{\beta}$ can be easily retrieved from the collected mobile data. These schemes therefore represent another example of collecting privacy-preserving mobile data to fulfill the needs of certain transportation applications (i.e. freeway OD estimation).

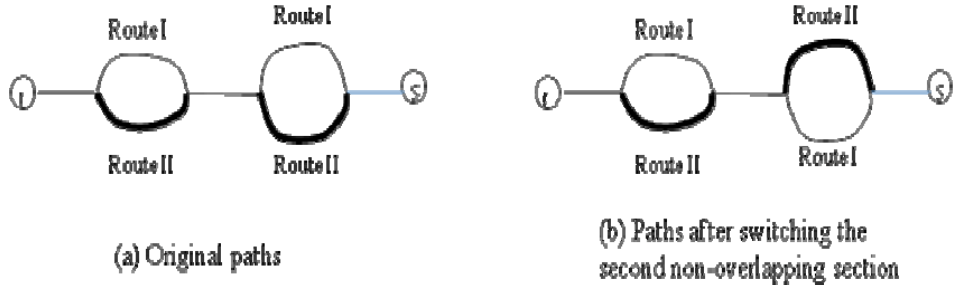


Figure 5. Random Route Switching Algorithm for Privacy Preserving.

Conclusions and Future Study

In this paper, we proposed a privacy-aware transportation modeling (PATM) and application-aware privacy protection (AAPP) framework for privacy-preserving transportation modeling using mobile sensor data. The proposed framework focuses on the interactions between transportation modeling and privacy preserving – being aware of privacy when developing transportation models as well as application needs when designing privacy preserving mechanisms. The framework was illustrated using two case studies: arterial performance measurement using VTL travel times and freeway OD estimation using aggregated OD ratio information. It was shown that the PATM & AAPP framework can be generally initiated by investigating the application data needs and then adjusted by privacy preserving principles. The process is usually iterative in order to reveal the minimum data requirements and to design the most effective way to provide such data without severely comprising privacy. Therefore in most cases, innovations are needed to design non-traditional data collection schemes, as well as to develop novel transportation modeling techniques based on the new data formats.

The widely available mobile traffic sensors present great potentials, while at the same time great challenges, for privacy protection and modeling using mobile data. The proposed PATM & AAPP framework aims to address these issues; the results reported in this paper are just the first step in the sense that the framework was only illustrated using two case studies. However, it suffices to show that privacy protection and transportation modeling are two equally important aspects to use mobile sensor data, which need to be addressed simultaneously. In the future, the framework will be extended and tested on more applications. More importantly, theoretical investigations are needed to provide a comprehensive framework on how to systematically protect privacy while preserving critical information as much as possible. For this, novel privacy protection mechanisms and mobile-data based traffic modeling methodologies are needed. In this process, close collaborations between privacy protection experts and transportation modeling researchers is the key.

References

Agrawal, R., and Srikant, R. (2000). “Privacy-preserving data mining.” *SIGMOD* , 29(2), 439-450.

Agre, P.E. (1995). “Transport informatics and the new landscape of privacy issues.” *Computer Professionals for Social Responsibility (CPSR) Newsletter,* 13(3).

Balke, K., Charara, H., & Parker, R. (2005). “Development of a Traffic Signal Performance Measurement System (TSPMS).” *exas Transportation Institute*, Report 0-4422-2.

Ban, X., Herring, R., Hao, P., and Bayen, A. (2009). "Delay pattern estimation for signalized intersections using sampled travel times." *Transportation Research Record 2130*, 109-119.

Ban, X., and Hao, P. (2010). "Real-time queue length estimation for signalized intersections using sampled travel times." *In Proceedings of the 89th Annual meeting of Transportation Research Board (CD-ROM)*.

Bell, M. (1991). "The estimation of origin–destination matrices by constrained generalized least squares." *Transportation Research Part B 25*, 13–22.

California Center for Innovative Transportation (CCIT), 2006. "Final report of task order 3: Corridor management plan demonstration." *California Center for Innovative Transportation, University of California, Berkeley*.

Cascetta E., and Nguyen S. (1988). "A unified framework for estimating or updating origin/destination matrices from traffic counts." *Transportation Research Part B 22*, 437-455.

Cascetta, E., Inaudi, D., Marquis, G.(1993). "Dynamic estimators of origin–destination matrices using traffic counts." *Transportation Science 27* (4), 363–373.

Chang, G.-L., Wu, J., (1994). "Recursive estimation of time-varying origin–destination flows from traffic counts in freeway corridors." *Transportation Research Part B 28 (2)*, 141–160.

Chang, G.-L., Tao, X., (1996). "Estimation of dynamic O–D distribution for urban network." *Proceedings of the 13th International Symposium on Transportation and Traffic Theory*, 1–20.

Cheu, R.L., Xie, C., Lee, D., (2002). "Probe vehicle population and sample size for arterial speed estimation." *Computer-Aided Civil and Infrastructure Engineering* 17 (1), 53–60.

Comert, G., and Cetin, M. (2008). "Queue Length Estimation from Probe Vehicle Location and the Impacts of Sample Size." *To appear in the European Journal of Operational Research*.

Congressional Record (2001). " Location privacy protection act of 2001." *Available on Internet: http://www.techlawjournal.com/cong107/privacy/location/s1164is.asp*.

Darroch, J.N. (1964). "On the traffic-light queue." *The Annals of Mathematical Statistics*, 380–388.

Demers, A., List, G., Wallace, W.A., Lee, E.E., Wojtowicz, J. (2006). "Probes as a path seeker: a new paradigm." Transportation Research Record 1944, 107-114.

Duckham, M., and Kulik, L. (2006). " Location privacy and location-aware computing, in Dynamic & Mobile GIS: Investigating Change in Space and Time." *J. Drummond, et al., Editors. CRC Press: Boca Raton, FL USA.*, 34-51.

Gruteser, M., and Hoh, B. (2005). "On the anonymity of periodic location samples." *In Proceedings of the Second International Conference on Security in Pervasive Computing.*

Gruteser, M., and Grunwald, D. (2003). "Anonymous usage of location-based services through spatial and temporal cloaking." *In Proceedings of First ACM/USENIX International Conference on Mobile Systems, Applications, and Services (MobiSys), San Francisco, CA.*

Haight, F.A., (1959). "Overflow at a traffic flow." *Biometrika* 46 (3-4), 420–424.

HarrisInteractive (2007). "National Study Shows GPS Adoption Rates Relatively Low, but Offers Recommendations to Accelerate Market." *Available on Internet:* Penetrationhttp://www.harrisinteractive.com/NEWS/allnewsbydate.asp?NewsID=1241.

He, R., Liu, H.X., Kornhauser, A.L., and Ran, B. (2002). "Study travel time variability from probe vehicle data." *Proceedings of the seventh International Conference on: Applications of Advanced Technology in Transportation, Cambridge, MA, United States*, 16-23.

Herrera, J.C., Work, D.B., Herring, R., Ban, X., and Bayen, A. (2009). "Evaluation of traffic data obtained via GPS-enabled mobile phones: the Mobile Century field experiment." *Transportation Research Part C, in press.*

Hoh, B., et al., Preserving Privacy in GPS Traces via Uncertainty-Aware Path Cloaking, in 14th ACM Conference on Computer and Communication Security (ACM CCS 2007). 2007: Alexandria, VA USA.

Hoh, B., Gruteser, M., Herring, R., Ban, J., Work, D., Herrera, J.C., and Bayen, A.. (2008). "Virtual trip lines for distributed privacy-preserving traffic monitoring." *In The Sixth Annual International conference on Mobile Systems, Applications and Services (MobiSys 2008), Breckenridge, U.S.A.*, June 2008.

IntelliDrive (2009). Available on Internet: see http://www.intellidriveusa.org/index.php.

International Telecommunication Union (ITU, 2009). Internet Link (Accessed on Dec. 02, 2009): http://www.itu.int/ITU-D/ict/statistics/ict/graphs/mobile.jpg.

Jacobson, L. (2007). "Vehicle Infrastructure Integration Privacy Policies Framework (ver 1.0.2)." *Reported of The Institutional Issues Subcommittee of the National VII Coalition*.

Karger, P.A., and Frankel, Y. (1995). "Security and privacy threats to ITS." *In Proceedings of the Second World Congresson Intelligent Transport Systems, volume 5, Yokohama, Japan*.

Krumm, J. (2008). "A survey of computational location privacy." *Personal and Ubiquitous Computing*.

Lighthill, M.J., Whitham, G.B.(1955). "On kinematic waves I Flood movement in long rivers. II A theory of traffic flow on long crowded roads." *Proceedings of Royal Society (London)* A229, 281–345.

List, G., Wallace, W.A., Demers, A., Salasznyk, P., Lee, E.E., Wojtowicz, J. (2006). " Field experience with a wireless GPS-based ATIS system." *In Proceedings of the 12th ITS World Congress (CD-ROM), San Francisco*.

Liu, X., and Ma, W. (2008). "A real-time performance measurement system for arterial traffic signals." *Presented at the 87th Annual Meeting of Transportation Research Board*.

Liu, X., Wu, X., Ma, W., and Hu, H. (2009). "Real time queue length estimation for congested signalized intersections." *Transportation Research, Part C, in press*.

McNeil, D.R., (1968). "A solution to the fixed cycle traffic light problem for compound Poisson arrivals." *Journal of Applied Probability* 5, 624–635.

Newell, G. F. (1960). "Queues for a fixed-cycle traffic light." *Ann. Math. Statist*. 31 589–597.

Newell, G.F., (1965). "Approximation methods for queues with application to the fixed-cycle traffic light." *SIAM Review* 7, 223–240.

Nie, Y., and Zhang, H.M. (2008). "A variational inequality formulation for inferring dynamic origin–destination travel demands." *Transportation Research Part B*, 42, 635–662.

Privacy Rights Clearinghouse (2009). "A Chronology of Data Breaches." Available on Internet: http://www.privacyrights.org/ar/ChronDataBreaches.htm (Accessed: 12/18/2009).

Qiu, Z., Chen, P., Jing, J, and Ran, B (2007). "Cellular probe technology applied in advanced traveler information." *International Journal of Technology Management*, in press.

Quiroga, C.A., Bullock, D. (1998). "Travel time studies with global positioning and geographic information systems: an integrated methodology." *Transportation Research Part C* 6(1-2), 101-127.

RITA (2008). "VII Proof of Concept Test Final Reports." Available on Internet: http://www.resourceguide.its.dot.gov/default.asp?SID=3&SSID=9.

Skabardonis, A., Geroliminis, N.(2008). "Real-time Monitoring and Control on Signalized Arterials." *Journal of Intelligent Transportation Systems*. 12 (2), 64–74.

Smaglik, E. J., Sharma, A., Bullock, D.M., Sturdevant, J.R., and Duncan, G. (2007). "Event-Based data collection for generating actuated controller performance measures." *Transportation Research Record 2035*, 97-106.

Smith, B.L., and Fontaine M.D.(2007). "Private-sector provision of congestion data." *NCHRP report* 70-1.

Solove, D. J. (2008) Understanding Privacy. Harvard University Press, Cambridge, Massachusetts.

Turner, S., Holdener, D.(1995). "Probe vehicle sample sizes for real-time information: The Houston experience." I*n: Proceedings of the Vehicle Navigation & Information Systems Conference, Seattle, WA, United States,*3–10.

Warrior, J., McHenry, E., and McGee, K. (2003). "They know where you are [location detection]." *Spectrum, IEEE* , 40(7), 20-25.

Wasson, J.S., J.R. Sturdevant, and D.M. Bullock. Real-Time Travel Time Estimates Using Media Access Control Address Matching. ITE Journal, 78(6), (2008), 20-23.

Webster, F.V.(1958)."Traffic signal settings." Road Research Laboratory Technical Paper No. 39, HMSO, London.

Yang, H., Sasaki, T., Iida, Y., Asakura, Y.(1992). "Estimation of origin–destination matrices from link traffic counts on congested networks." *Transportation Research B* 26, 417–434.

Zhou, X.(2004). "Dynamic origin–destination demand estimation and prediction for off-line and on-line dynamic traffic assignment operation." *Ph.D. Thesis, University of Maryland, College Park.*

Zhou, X., Mahmassani, H.S.(2007). "A structural state space model for real-time traffic origin–destination demand estimation and prediction in a day-to-day learning framework." *Transportation Research Part B* 41 (8), 823–840.

Deriving Rules for Trip Purpose Identification from GPS Travel Survey Data and Land Use Data: A Machine Learning Approach

Zhongwei Deng[1] and Minhe Ji[1, 2,]

[1] GIScience Key Laboratory, Ministry of Education, East China Normal University
[2] China East-West Cooperation Research Center, East China Normal University

Abstract

Passive GPS travel survey has proven an innovative alternative to the traditional paper-and-pencil method, as it can record exact time and location of travel activities without incurring much burden on survey respondents. Lying in the core of this technique is how to derive accurate travel information (e.g., trip purpose) from GPS data. Previous studies rely on simple rules that are manually extracted from the inherent structure of GPS data streams. These methods, however, are difficult to generalize for other applications and lack helpful hints for new research. This paper presents a machine learning approach to deriving trip purpose from GPS track data coupled with other relevant data sources. This approach employs a number of attributes (i.e., time stamp and land-use type of trip ends, a set of spatiotemporal indices of travel, and demographic and socioeconomic characteristics of respondents) to construct a decision tree for purpose of classification. Each attribute provides partial evidence to the depiction of a given purpose, but this depiction may or may not be monotonic, and none of them can work alone toward the goal. A reasoning procedure using the adaptive boosting technique was designed to explore how these attributes could work together to achieve trip purpose derivation. This technique generated multiple decision trees to improve the classification results through a mechanism of voting from these trees. Each tree was constructed in the depth-first fashion with the root node and the split of subsequent nodes being determined on the basis of the gain-ratio computed for the relevant attributes. This procedure was implemented in the C5.0 machine learning environment with 226 GPS trip records collected from 36 respondents. The experimental results seemed rather promising: using 10 iterations for adaptive boosting, an overall classification accuracy of 87.6% was achieved.

Key Words: passive GPS travel survey;trip purpose; machine learning; C5.0

Introduction

High-quality travel survey data are essential to urban transportation planning and regional demand modeling. Traditional household travel survey (HTS) methods have, however, long been recognized as suffering from such problems as low response rates and poor data quality. All these methods, including face-to-face interview and self-administrated questionnaire as well as their computer-assisted variants (i.e., CATI, CAPI, CASI), share something disadvantageous – they all rely on human memory for spatiotemporal data collection (Evert et al. 2004). The vagueness and inconsistency of human memory may typically degrade the data being collected and even ruin the entire travel diary record. Due to easy memory failure as respondents are subject to constant and long-lasting pressure, long-term travel surveys are seldom performed successfully (Wolf et al. 2001). Positioning and navigation technologies such as GPS can provide a viable solution to those problems. An average GPS instrument with second-by-second data collection capacity and spatial accuracy of 3-15 m, may serve as an ideal device for personal travel survey. Accurate spatiotemporal records of personal movements provide a nice description of travel behavior, which can be potentially utilized to construct travel diaries with a needed origin and destination information. Apparently, the burden of memorizing time and location for each trip during travel can be completely eliminated, potentially leading to a higher response rate and better data quality. This GPS-based travel survey method has become a research trend since the mid 1990s, as it is now widely used as a supplement to and even considered as a replacement for traditional methods (Wagner 1997; Stopher et al. 2001; Wolf and Lee 2008).

The biggest challenge of using GPS for travel survey is how to effectively extract travel information from the raw track data. The content of HTS typically consists of two parts, i.e., the static background information (e.g., demographic and work-related information for all household members and household vehicle ownership) and the travel behavior information. Since GPS track data only record movements regardless of the object's identity, a GPS-based HTS also needs to collect static information through questionnaires. The significant improvement over the traditional methods lies in the second part, namely, many elements can be directly deduced from GPS records, such as speed, distance, trip duration, and even trip routes. However, the most crucial information, e.g., travel mode and trip purpose, requires more efforts to derive. Previous research coupled GPS data with land-use information to derive trip purpose using rules manually derived from the GPS data being collected (Stopher et al. 2008). Although satisfactory results can be achieved using this approach, the manually derived rules are data-dependent and subject to researchers' experience. Different data sources and researchers will apparently lead to different rules with no relevant theory to govern the rule derivation process for mining needed travel information.

This paper describes a novel approach of using machine learning to derive rules from a combination of GPS and GIS data for trip purpose derivation. GPS survey data are typically ill-structured as they result from human behavior through space and time.

This can be perfectly dealt with via machine learning, a much preferred solution to ill-structured decision making problems (Christopher 2006). On the other hand, the spatiotemporal patterns of surveyed travel activities present considerable uncertainty in their prediction from GPS records. It is, therefore, logical to bring in such relevant data sources as land use and even locations of service facilities for uncertainty reduction. The output of this machine learning process is a group of decision trees for trip purpose identification with rules being extracted from a set of selected attributes. The rest of the paper first summarizes the attributes to be used for trip purpose derivation. This includes spatiotemporal indices computed from the GPS survey data and the ancillary land-use data. A proof-of-concept case study with GPS travel survey data collected from 36 respondents is then presented. The learning procedure and post-classification analysis were conducted in the C5.0 environment.

Methodology

Idea Formulation. In trip purpose derivation, we consider a typical deduction process for a sample question, "what is the purpose of Kevin's trip?

Situation 1: "Kevin went from home to school."
Based on prior knowledge about travel needs in this statement, we can infer at least five purpose candidates for Kevin, i.e., "schooling," "working," "business visit," "sending a child to school," and "tour or entertainment." In the process the basic information for deduction is "school" and "home," which is related to land-use types that are associated with different trip purposes. Therefore, land-use information is critical for trip purpose derivation.

Situation 2: "Kevin went from home to school at 7:30 AM."
With this additional information, we can naturally remove "business visit" from the candidate list because this visit behavior seldom takes place in the early morning. Different activities are usually performed within the timeframe in which they are deemed appropriate. Therefore, temporal information is another key factor to consider in trip purpose derivation.

Situation 3: "Kevin went from home to school at 7:30 AM and stayed there for five minutes."
This statement adds a piece of behavioral information about duration of stay. Combining this with the land-use and arrival-time information, the trip purpose option of "sending a child to school" can be inferred with high confidence. Likewise, other travel characteristics like trip distance, trip duration, and travel mode are useful indicators for trip purpose derivation.

Situation 4: "Kevin went from home to school at 7:30 AM and stayed there for five minutes, and he is a lawyer and a parent of a 7-year-old boy."
With this additional information about Kevin's occupation and family status, we can definitely infer that the trip purpose is "sending a child to school". In an HTS of any

kind, such static information as demographic characteristics and work-related status is also important and should be available for trip purpose derivation.

As shown in the above discussion, trip purpose derivation can be treated as a logical deduction process, which is based on a given set of prior knowledge in the problem domain and empirical evidences that are provided or revealed from the GPS-based HTS data. These evidences should include location, time, land-use, travel behavior, and socioeconomic characteristics. The associative and supplementary nature of these evidences may not require all of them in presence during deduction; however, the reliability of the deduction depends largely on the completeness of information.

C5.0. Success of using decision trees for building knowledge-based systems by inductive inference from examples has been demonstrated in many practical applications (Wu et al. 2008). Systems take as input a collection of cases, each belonging to one of a small number of classes and described by its values for a fixed set of attributes, and output a classifier that can accurately predict the class which a new case belongs. In this study, we used C5.0, a univariate decision tree algorithm widely used and tested for classification purposes with data of diverse types. A complete description of this algorithm is beyond the scope of this paper, and the interested reader is referred to Quinlan's work (1993). Hereby we only summarize the key components of this algorithm.

C5.0 adopts the approach of information gain ratio to estimate splits at each internal node of the tree. Information gain ratio measures the reduction in entropy in the data produced by a spilt. Using this metric, the test at each node within a tree is selected using the subdivision of the data that maximizes the reduction in entropy of the descendant nodes. Given a training data set S, a test is performed on a variable A to create branches and to partition data set S into v subdivisions, the information gain of attribute A refers to S can be described as (Quinlan 1993):

$$Gain(S, A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (1)$$

where v is one of the values of attribute A, $Values(A)$ denotes the full set of values of attribute A, and S_v is the subset of S for which the values of A are equal to v (*i.e.,* $S_v = \{s \in S | A(s) = v\}$). The first term of the equation is the original entropy of set S, and the second term indicates the entropy after split by attribute A; so Gain (S, A) denotes the reduction in entropy when the value of attribute A is known.

To create a decision tree efficiently, it is important to branch on the attribute that gains the most information. If the above measure is directly used for selecting the most informative variable, however, the test will be biased toward attributes with a larger number of values. To compensate for this effect, Gain (S, A) is normalized with the equation below:

$$SplitInformation(S, A) \equiv -\sum_{i=1}^{c} \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \quad (2)$$

then the gain ratio can be obtained as the splitting metric as follows:

$$GainRatio(S, A) \equiv \frac{Gain(S, A)}{SplitInformation(S, A)} \quad (3)$$

where $S_1 \sim S_c$ are the subset of S split by c different values of attribute A, and $SplitInformation(S, A)$ reflects the breadth and equality of split by the attributes.

Using this framework, A is recursively split such that the gain ratio is maximized at each node of the tree. This procedure continues until each leaf node contains only observations from a single class or no gain in information is yielded by further splitting.

Adaptive Boosting. Adaptive boosting is an innovative approach to improving classification accuracy in C5.0. This algorithm generates several classifiers to be used in the classification of a new case in a weighted voting fashion. That is, each classifier votes for its predicted class, and the votes are counted to determine the final class (Quinlan 1996). The multiple classifiers are generated in a given number of iterations, with the misclassified cases in the previous classifier being made correct in the making of the current classifier. In C5.0 the number of classifiers is equal to the number of trials specified by the user with the Boost option. Although it requires more computation than building a single decision tree, the improvement has proven significant. Trials over numerous datasets, large and small, have shown that on average 10-classifier boosting reduces the error rate for test cases by about 25% (Yoav et al. 1996).

Attributes Generation. In a specific GPS-based HTS situation, the attributes discussed in 2.1 can be extracted from GPS data, a related household background survey, and a GIS database. The processing steps are as follows. GPS trajectories are first segmented into individual trips, for which such travel characteristics as speed, distance, trip duration, and travel mode are derived (for a complete description, please refer to Chen and Ji (2009)). Then the trip ends are imported into ArcGIS™ and spatially overlaid with a land-use layer. This step allows individual trip ends to associate with land-use types of their locations. Similarly, the time stamp of a trip end is semantically grouped into a few categories (i.e., early morning, mid morning, noon time, early afternoon, late afternoon, and evening). The association of socioeconomic data to a trip end is made directly from the specific respondent's background information recorded when taking the trip.

As C5.0 is used in this study, the task of trip purpose identification becomes a pattern recognition process. All attributes resulting from above are then treated as the input variables for decision tree learning, and the trip purposes identified for the training

Table 1. Input and target variables for trip purpose identification.

	Types	Variables
	Land-use	Land-use
Input	Socioeconomic	Occupation, income, family structure, age
	Temporal	Weekdays, weekend days, time of a day
	Travel characteristics	Speed, mode, distance, duration
Target	Trip purpose	Trip purpose

samples are used as the target variable. All these variables together comprise the training sample dataset to be engaged in the construction of decision trees.

Experiments

Data Collection. In this case study, a group of 50 volunteers participated in a three-day GPS-based household travel survey in Shanghai, China. The respondents were requested to tote a portable GPS data logger when performing their daily travel activities and fill out a web-based HTS questionnaire at the end of each survey day. In addition to the regular household information, data about travel mode and trip purpose were also collected for building the training set. A complete description of the process is documented elsewhere (Deng et al. 2009; Chen and Li 2009). The data collection period lasted 2.5 months, and a total of 36 respondents completed the survey.

Data Preparation. This step involved the transformation of collected GPS track data into necessary information for trip purpose derivation. Definitions for several key variables were provided, and training samples were constructed for machine learning implementation in the C5.0 environment. These important variables are categorized as below.

Time intervals associated with activity constraints. The GPS data logger records date and time information, with time being recorded to the magnitude of a second. For the purpose of our analysis, the date information was converted to "weekday" and "weekend day." The time of a day also could be divided into semantic intervals with social and schedule constraints as delimiters. Since C5.0 can deal with continuous variables, on the other hand, there was no need to classify time into semantic intervals.

Characteristics of travel behavior. The positional information of surveyed GPS points combined with time can be processed to derive such behavioral indicators as travel speed, travel mode, trip distance, and trip duration. Trip ends can be identified by comparing the travel speed of a given trajectory to an empirical threshold value (e.g., consecutive points with travel speed less than 5 km per hour for 120 seconds).

Land-use. Followed from China's "Urban Land-use Classification Scheme and Planned Construction Land-use Standards," this study adopted 10 land-use types (i.e.,

residential, industrial, warehouse, highway transportation, street and square, municipal utilities, green space, specially designated, water bodies, and others). With both the trip-end data layer and the land-use layer in a GIS database, they were combined through a "point-in-polygon" analysis so that each trip end was attributed with the land-use type that was spatially associated with its location.

Socioeconomic information. This type of information was not considered as a classifying factor in this case study, since all respondents came from the same institution and thus possessed similar socioeconomic characteristics.

Trip purpose. The typology of trip purpose was defined in accordance with the classification scheme issued with the Fourth Comprehensive Transport Survey in Shanghai, the most recent HTS that took place in September 2009. Specifically, trips derived from GPS track data would be classified into seven categories such as *go to work*, *go to school*, *go home*, *pick-up or drop-off*, *shopping or recreation*, *business visit*, and *other activities*.

Classification Results. The data preparation process resulted in a training sample set of 226 trips with complete attributes, which was presented to C5.0 for rule learning. Adaptive boosting was used for minimizing classification errors. A range of 10 iterations were run in the boosting algorithm for accuracy comparison, and the results were documented in Table 2. In the table, the Trial column indicates the number of iterations for boosting, Size is the number of non-empty leaves on the tree, and Errors shows the number and percentage of cases misclassified. Clearly the use of the adaptive boosting technique was rather effective - the percent error was reduced from the original 19.5% to 12.4% after boosting with 10 trials.

Further evaluation of the learning effectiveness for trip purpose identification was performed through an error matrix (Table 3). The *go home* class accounted for the largest share of the total trips, and its classification accuracy achieved 95.6%. In comparison, the classes of *go to work*, *shopping or recreation*, and *pick-up or drop-off* also presented high accuracy levels (~ 90%). Unsatisfactory results occurred with *business visit* (80.8%) and *other activities* (73.3%). The findings from this classification are highly consistent with the socioeconomic status and demographic characteristics of the respondents. All 36 surveyed volunteers were university faculty members, and the life of teachers in higher education is relatively simple - *go home*, *go to work*, *pick up or drop off* their kid(s) at school – and these activities dominate their life, so that their travel behaviors have a strong relationship with specific land-use types and conform to specific time constraints. These routine trip purposes had very distinctive and homogeneous spatiotemporal patterns and therefore are relatively easy to identify and predict. In contrast, spatiotemporal patterns of low-frequency activities such as *other activities* and *business visit* for this group of respondents are far more difficult to learn.

Table 3. Error matrix for trip purpose classification.

(a)	(b)	(c)	(d)	(f)	(e)	Classified as	Accuracy
66		2		1		(a): class *go home*	95.6%
2	61	3	2			(b): class *go to work*	89.7%
1	1	30			1	(c): class *pick-up or drop-off*	90.9%
3	6	1	23	1	1	(d): class *other activities*	73.3%
	1			9		(e): class *shopping or recreation*	90.0%
	1		1		9	(f): class *business visit*	80.8%

Table 2. Errors percentage of boosting 10 trials.

Trial	Size	Errors (%)
0	25	44 (19.5)
1	24	66 (29.2)
2	32	61 (27.0)
3	24	73 (32.3)
4	25	59 (26.1)
5	26	65 (28.8)
6	32	67 (29.6)
7	31	59 (26.1)
8	23	60 (26.5)
9	28	61 (27.0)
boost		28 (12.4)

Conclusions

In the efforts of automating trip purpose identification from GPS-based HTS data, combining evidence from different sources through AI-based approaches seems a

viable direction. This paper formulated a machine learning procedure to incorporate four different types of information into a multiple decision tree based classification. These information types included GPS time stamps, land use, temporal behavior, and socioeconomic/demographic characteristics. A case study with over 200 trip records collected from 36 survey respondents was conducted to evaluate the feasibility of the aforementioned classification logic. With the use of the adaptive boosting strategy, the trip purpose identification task achieved a rather high correctness of 87.6%, a significant error reduction of 7.1% from classification with no boosting.

The experimental results also lead to following findings.

1) The four information types provided a rather complete framework of evidence and knowledge for trip purpose reasoning and identification. Even when part of information is missing, such as some sample records with no socioeconomic data, the learning and classification procedure still produces reasonably good results. It

implies that the supplementary nature of evidences from the selected information types can generate a sufficient level of robustness against missing data effects.
2) The strategy of combining relevant information from GPS data and other sources is workable in a carefully designed logical sequence and could be very well represented and implemented as a machine learning procedure. Although the process of defining attributes for trip purpose derivation and translating them into the "attribute-value" form can be rather tedious, the benefit of doing so is worth the effort: it can effectively minimize arbitrariness and subjectivity as typically seen in many manual rule-setting approaches and ensure the highest possible classification accuracy.

3) Decision trees and machine learning can handle ill-structured problems and data fuzziness rather well. In this case study, C5.0 demonstrated its ability of tolerating classes with scarce training samples and missing data as it built rules from inherent data structures and associations. The benefits of machine learning include its ability to discover, generate, and improve rules for trip purpose identification from data themselves. Only when the generation of classification rules for travel behavior analysis no longer depends on human interference, can the GPS-based HTS method be widely accepted for transport applications.

It is also worth noting that this study has examined only C5.0 for machine learning with a small set of homogeneous samples. The results may not be generalized for applications with a diverse setting and/or a large sample size, although a better performance can usually be achieved when more data are involved. Further research will include experiments with other machine learning methods, more diverse travel situations, and larger sample sizes.

References

Chen, W., and Ji, M. (2009). "A prompted recall interview platform for GPS-based HTS: design and development." *Paper presented to the international workshop on GIS for Transportation*, 1:157-163.

Christopher, M. B. (2006). "Pattern recognition and machine learning." *Series: Information Science and Statistics. Springer Verlag: New York.*

Deng, Z., Ji, M., and Chen, W. (2009). "Coupling passive GPS tracking and web-based survey to improve the response rate and quality of household travel surveys." *Journal of transportation system engineering and information technology*, 1-6.

Evert, H. V., Werner, B., and Erhard, E. (2004). "Survey design: the past, the present and the future." *Report to AVV Transport Research Centre, Netherlands*, 1-21.

Quinlan, J. R. (1993). "C4.5: Programs for Machine Learning." *San Mateo, California: Morgan Kaufmann.*

Quinlan, J. R. (1996). "Bagging, boosting, and C4.5." *In Proceedings of the 13th National Conference of Artificial Intelligence, Portland, Oregon*, 725-730.

Stopher, P. R., Philip B., and Stephen, G. (2001). "Using passive GPS as a means to improve spatial travel data: further findings." *Paper presented to the 23rd Conference of Australian Institutes of Transport Research.*

Stopher, P. R., Clifford, E., Zhang J., and FitzGerald, C. (2008)."Deducing mode and purpose from GPS Data." *Paper report to Institute of Transport and Logistics Studies*, 1-13.

Wagner, D. B. (1997). "Lexington area travel data collection test: GPS for personal travel surveys." *Final Report for OHIM, OTA, and FHWA*, 5-21.

Wolf, J., Guensler, R., and Bachman, W. (2001). "Elimination of the travel diary: Experiment to derive trip purpose from GPS travel data."*Transportation Research Record*, 1768:125-134.

Wolf, J., and Lee, M. (2008). "Synthesis of and statistics for recent GPS-enhanced travel surveys." *Submitted to the International Conference on Survey Methods in Transport: Harmonization and Data Comparability.*

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., and Motoda, H. (2008). "Top 10 algorithms in data mining." *Knowledge Inference Systems*, 5(4):1-37.

Yoav, F., and Llew, M. (1996). "Experiments with a new boosting algorithm." *Machine Learning: Proceeding of the Thirteenth International Conference*，148-156.

Prediction of Freeway Travel Time in Incident Management Evaluation Based on Genetic Neural Network

Ding He[1], Fujian Ni[2] and Shunxin Yang[3]

[1] Dr., [2] Prof., [3] Dr., School of Transportation, Southeast University, No. 2 Sipailou, Nanjing 210096, China, Email: richard3515@163.com, nifujian@gmail.com, shunxin.yang@gmail.com

Abstract

In order to evaluate the freeway incident management systems, the normal travel time for calculating the incident delays and benefits usually needs to be predicted. The evaluation result is very sensitive to incident delay, so it is inaccurate to use month, week, or day average value for describing the normal travel time. The artificial neural network with strong adaptability and high accuracy was proposed to solve the problem. Firstly the structures of a predicting model and data acquisition plan were proposed. Then a genetic algorithm was introduced to improve the traditional BP(Back propagation) neural network considering that BP network easily falls into the local convergence, and the genetic algorithm has a good global search capability. GA-BP(Genetic algorithm improved back propagation) network showed better convergence speed and predicting accuracy than BP in the study of Ningtong freeway travel time. The final error of GA-BP can be accepted, and the forecast results meet the actual need.
Key Words: travel time; neural network; genetic algorithm

Introduction

Calculating the incident delay savings to analyze the benefit is widely used in the study of incident management evaluation all over the world. The Service Patrol Program in Washington (Nee and Hallenback 2001), the CHART system in Maryland (Chang and Shrestha 2000), and other programs were evaluated with delay calculation. In this process, the prediction of the normal travel time in the same section and the same time without incident condition turns to an exigent problem because the time does not really happen and has great ambiguity.

Back propagation neural network (BP network) has learning association ability, parallelism, and a certain degree of fault tolerance and can approximate any complex non-linear system (Boozarjomehry and Svrcek 2001). But the learning process based on gradient descent leads to the slow convergence speed and more possibility of

falling into local convergence (Yan and Zhang 2000). The genetic algorithm (GA) has a good global search capability to optimize neural network algorithm (Chen et al. 2001). In this study, an improved genetic algorithm neural network (GA-BP) to predict the normal travel time was proposed, which can effectively solve the local convergence of the traditional BP network defects and improve learning speed and prediction accuracy of the network. Positive results are achieved with the validation of the actual sample data.

Literature Review

The prediction of travel time plays a rather imporatant role in transportation research. Rapid and precise prediction helps with evaluation, disposal decision, and traveller information in incident management. Recently diverse methods such as state equation (Hiramatsu et al. 2009), novel logliner model (Huang et al. 2008), pattern matching (Bajwa and Islam 2003), and knowledge based (Lee et al. 2009) have been used to deal with the problem. The methods above are mostly based on history data, experience, or specific parameters On the other hand, the way of artificial neural network was rarely reported in this area.

Artificial neural networks developed rapidly in the 1980s; it's a large-scale parallel-linked system, which simulates the structure of the human brain with self-adapting modelling and studying functions (Hu et al. 1993). The technique has a series of affiliated research and BP network, one of the most often used artificial neural networks, gets many successful applications in different areas. The BP network has a good adaptability of incomplete information and does not depend on specific parameters as much, which can better fit the complicated traffic environment. Safe evaluation (Chen et al. 2008), the prediction of traffic flow (Liu et al. 2009), and other studies proved the feasibility and accuracy of the BP method in traffic studies.

The Prediction of Ttravel Ttime

Incident Delay Analyses Based on Toll Station Data qnd the Prediction of Travel Time. Toll stations are usually set at the entrances or exits of freeway mainlines and the interchanges. Therefore, most freeways in China are isolated to other highways along the onward direction except in the location of toll stations. The actual travel time under incident conditions (gained with toll station data) minus the normal travel time without the effect of incident equal to the single vehicle delay, and then the total delay and benefit can be calculated. The evaluation shows a big sensitivity on incident delay. According to the Ningtong freeway report about economical benefit analyses of the incident management system, the benefit margin of a single incident can reach 300 thousand RMB with 8 minutes more single vehicle delay. The data from a toll station can be considered to accord with the study precision while month or week or day average travel time can't describe the normal travel time accurately enough especially under rain, fog, or snow conditions. Otherwise, travel time has many important applications besides in the field of evaluation. Neural network can be used to forecast this item.

Network input and output parameters and the data acquisition. Firstly, input and output parameters of neural network should be determined. The normal travel time of predicting a road section (the output parameter) closely relates to traffic volume, road

surface condition, weather conditions, and other random interference factors (Wang and Guo 2000). It's somewhat difficult to find the criterions of the classification and the evaluation of road surface conditions, weather characteristics, and other random interference factors while the travel time near the road section can easily express the effect of these factors. Traffic volume data can be found in the formal traffic observation record of freeway management center to ensure the reality. This study mainly focused on the incident management system of the Ningtong freeway and with the limitation of traffic volume and other conditions predicting travel time of passenger cars and trucks during the whole time point（from 3:00-4:00, for example） of time was used to replace the normal travel time during the corresponding time period. If traffic volume and other conditions confirm with some hypothesis, the travel time during 30 or 10 mins and so on can also be predicted. After the comprehensive analysis of traffic data from July, August, and September in 2008, different types (divided along the custom of toll station) of cars and trucks showed no obvious difference on travel time. So just the travel time of cars and trucks was used in the study.

The road network should be separated into sections before the construction of a neural network. To enhance the predicting accuracy, the freeway in every single toll station was divided along the forward direction. For alleviating the random error, the average filtering method (taking the average value of survey data in 60 min of whole time point as the sampling value) was introduced to eliminate major random noises.

Predicting Model of GA Improved Neural Network

BP neural network. Back propagation neural network (BP network) has good self-learning and self-association ability. The study process is composed of forward and back propagation. During the forward propagation, the input message is transmitted and treated in the hidden layer, and the status of one layer only affects the next layer. If the output layer can't get the expected result, the error signal will be transmitted back along the old link path, which is called the back propagation. After modifying the connection weight of different layers, the signal reaches the input layer, and then the new forward propagation begins. With the circulation of the operations we can get the anticipated error.

According to the Kolmogorov theorem, a 3-layer forward neural network can approximate any continuous nonlinear function by any accuracy (Lingras and Mountford 2001). This study then adopted a 3-layer BP network with one hidden layer. The intput parameters are set to be the normal travel time, traffic volume of passenger cars and trucks in the previous and next road section of the predicting section, and the traffic volume of the predicting section during the whole time point. The output parameters used the normal travel time of passenger cars and trucks of the predicting section during the corresponding time.

Genetic algorithm to improve the network(GA-BP). The genetic algorithm has a strong global search ability, high efficiency, and can ignore the continuity and derivability of the function. These features can improve the defection of local convergence and slow convergence speed of BP network. These are the basic steps how GA improves the BP network:

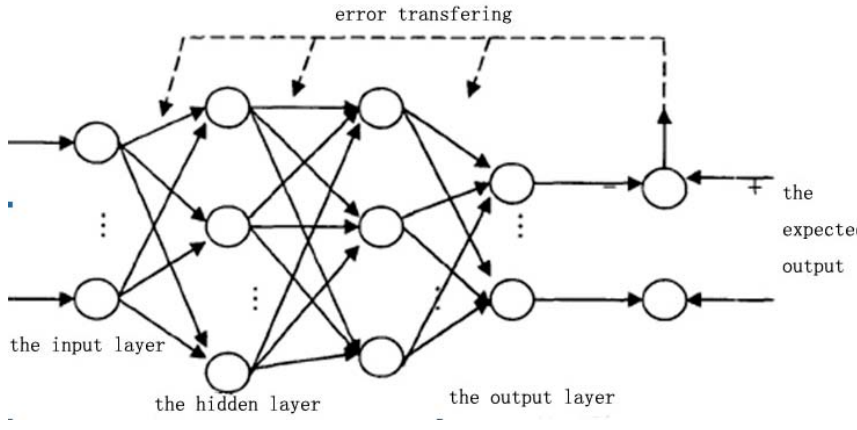


Figure 1. Model of BP neural network.

(1) Determine the training sample set and test sample set;
(2) Determine the encoded mode: encoding is to build the network along a certain way to get chromosomes of the genetic algorithm. Here, the real number method is used.
(3) Determining the population size N, generating the initial population randomly; population size affects the final result and the execution efficiency of GA;
(4) Decoding the initial population, determining the network structure and other parameters;
(5) Executing the forward propagation based on the training sample set, obtaining the output;
(6) Setting the objective function, transforming it into a fitness function; the definition of fitness is one of the key factors of getting the global extreme value. The less the error sum of squares of the input minus the output is, the better property the BP network shows. So we can define the fitness function F(X_m):

$$F(X_m) = 1/E = \frac{1}{\frac{1}{2}\sum_{i=1}^{n}(y_i - t_i)^2}$$

In the equation, y_i is the actual output; t_i is the expected output, and X_m is the sample m.
(7) Evaluating based on the fitness: if the fitness meets the requirement, assign the value to the BP network and go to (9); If it doesn't, make the genetic operations, including reproducing, crossing, and mutation, to generate the new population;
Reproducing. The proportional model is usually used to choose the superior individual. The individuals with fitness f_m(m=l,2…,N)> the average fitness $\overline{f} = \frac{\sum_{m=1}^{N} f_m}{N}$ will be chosen first to the candidate solution to generate the new population, the other part of candidate solutions will generate based on the proportion probability of individual's fitness $P = \frac{f_m}{\sum_{m=1}^{N} f_m}$, both the individual with large or small proportion have a chance to become the candidate solution in order to keep diversity and avoid to fall into local optimum.

Crossing. This operation is the main genetic operation. Choose two individuals (usually random pairing is used) to switch the gene randomly with a certain

probability P_c to generate a new individual. One-point crossover is used here. This method cut the two ancestors at the same position randomly, and then the two new individuals were produced.

Mutation. tation is changing one individual's gene randomly with a certain probability P_m. This ensures the local search capability and the diversity of the population. Uniform mutation is used here.
(8) If the model reaches the maximum genetic algebra, stop the loop and assign the value to the BP network; if it doesn't, return to step (5);
(9) Import the test sample set into the BP network, if the result meets the requirement, stop; if doesn't, train the BP network with the way of BP.

Calculation and Discussion

Introduction of predicting section and the parameter setting. The road section between the Taizhou toll station and Taixingdong toll station was chosen to be the predicting target (between station 9 and station 10 with the normal travel time of about 30 min for trucks and 22 min for passenger cars). The work for sample acquisition was conducted during July 2008. Because the formal traffic volume observations of freeway manage center were executed in the 5th, 15th, and 25th of the month, 68 samples were obtained in consideration of incident-effected and deficient data (58 for training and 10 for testing).

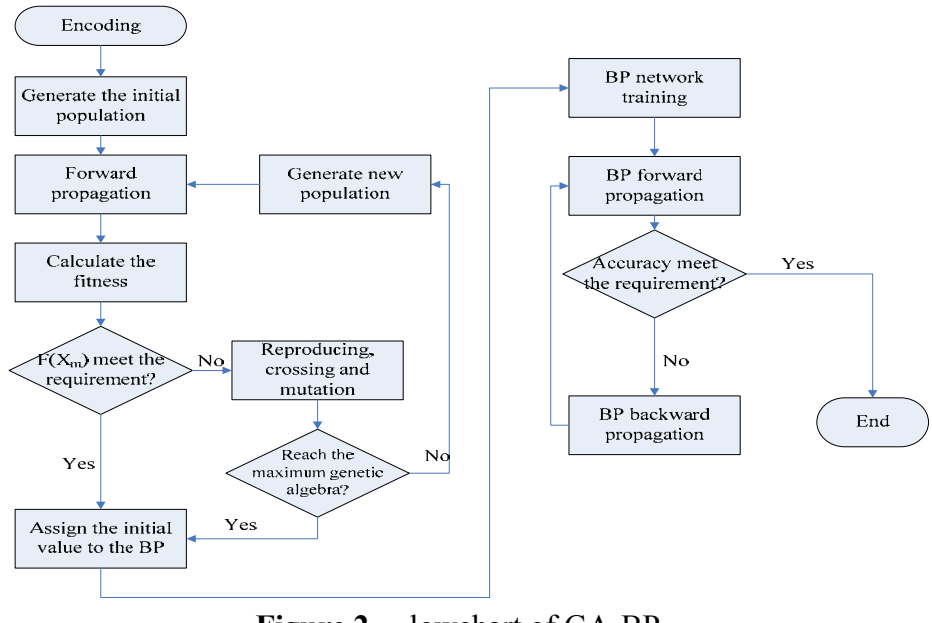


Figure 2. lowchart of GA-BP.

Normalization of sample data should be completed before the study and test of BP network to ensure the absolute value of input and output fall into the interval [0, 1]. Here the maximum genetic algebra is set to be 100, population size 50, crossing probability 0.8, mutation probability 0.005, $F(X_m)$ 0.005. The numbers of input and output parameters are 10 and 2, respectively. Usually there is no specific theory for determining the number of neuron in the hidden layer. The Kolmogorov theorem advised the researchers to test the number of 2 multiplied by the number of input parameters and then adding 1 (Hassoun 1995).

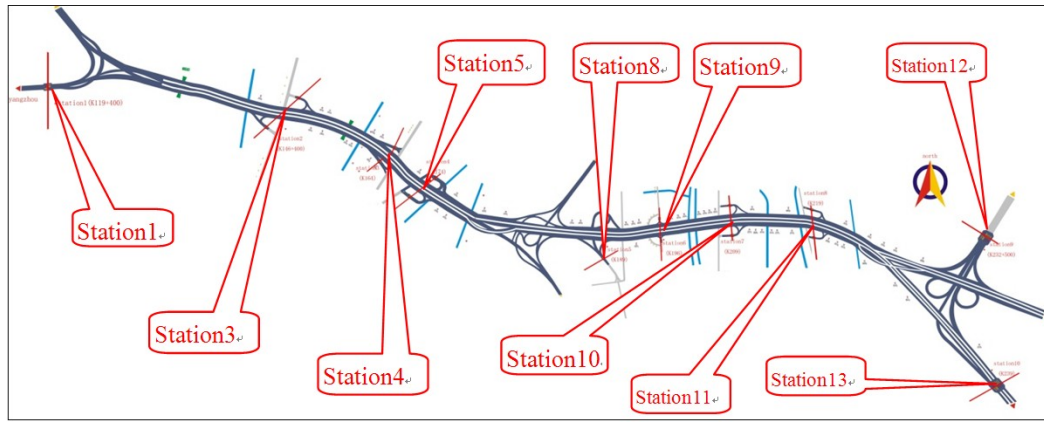


Figure 3. he Road Map of Ningtong Freeway.

So 18 was defined after a series of pilot calculations with the numbers 16-26. The initial weight was adopted randomly in the interval (0, 1). In the BP training process a gradient descent algorithm is used to train the network, the maximum training time was set to be 3000, learning rate 0.1, momentum factor 0. 90, and target error E_{bp} 0.0015.

Predicting results and analysis. Figure4 shows the error curve of normal BP network (Figure 4 (a)) and GA-BP network (Figure 4 (b)) with the training samples.

From Figure 4, it can be concluded that GA-BP only iterated 550 times within less than 2 min to the convergence while BP iterated 1035. GA improved BP network has better rate of error convergence.

Figure 5 shows the actual travel time and the forecast travel time of BP and GA-BP.

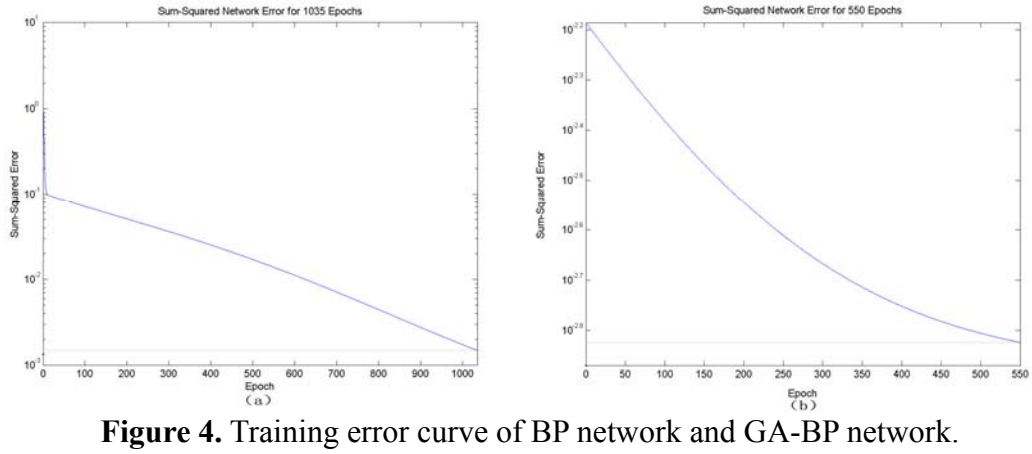


Figure 4. Training error curve of BP network and GA-BP network.

It's easily to discern that the GA-BP results approximate more closely to the actual travel time. In the Table 1, errors between predicting time and actual time were listed, and month average travel time was introduced for better comparison.

The average error and maximum error of GA-BP are obviously more positive than the two other methods. But the BP has a better minimum error of trucks, which proved that the BP is easy to fall into local convergence in a certain degree. The maximum error of GA-BP is equal to about 2.5 min in real time, which satisfied the expected accuracy.

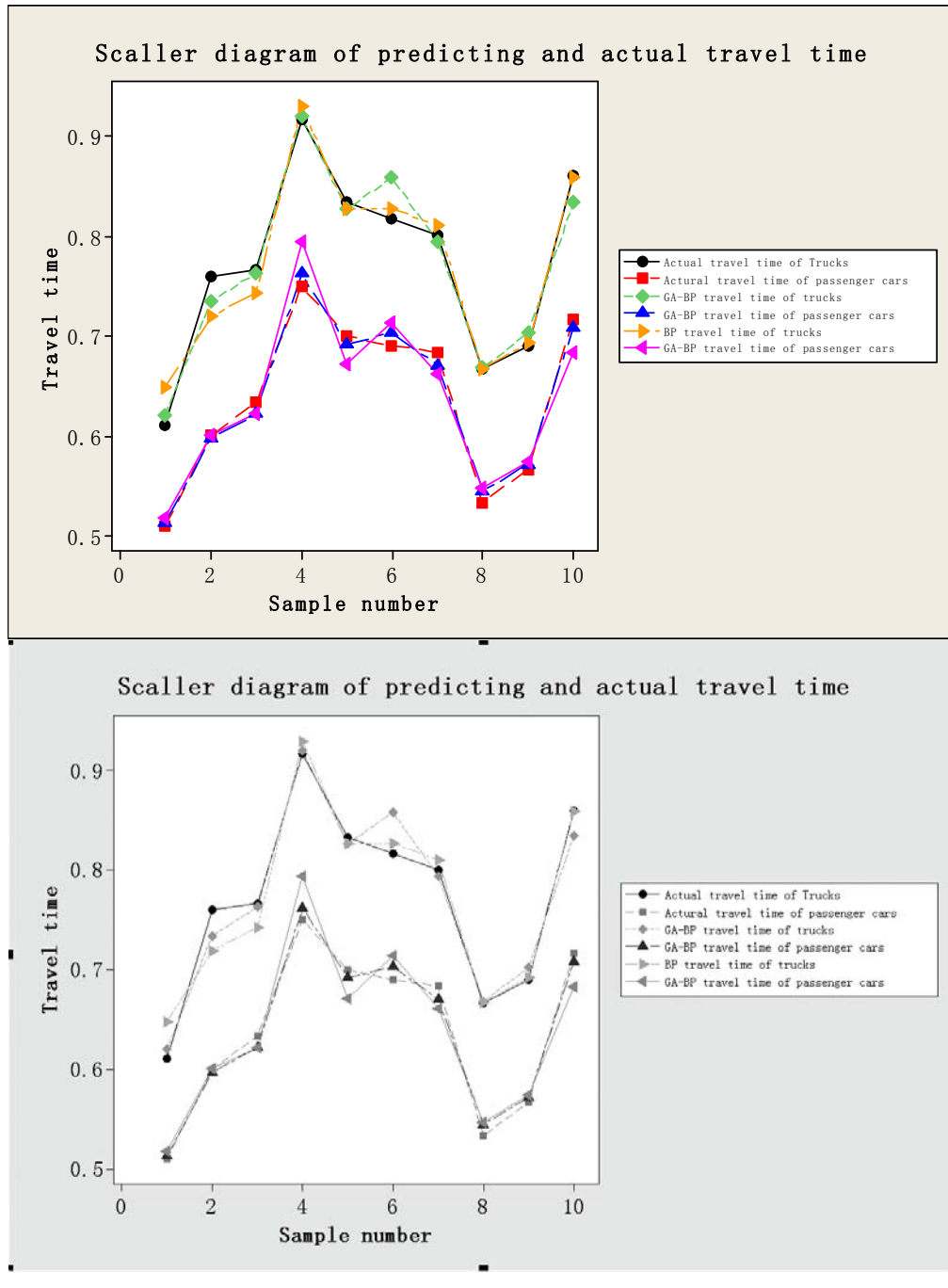


Figure 5. The comparison chart of actual travel time and the forecast travel time.

Table 1. Relative errors of different prediction model.

	Relative error of trucks			Relative error of passenger cars		
	Average	Maximum	Minimum	Average	Maximum	Minimum
GA-BP network	2.9%	4.5%	0.5%	2.1%	3.2%	0.2%
BP network	4.3%	6.1%	0.4%	3.9%	5.8%	0.3%
Month average	8.2%	17.3%	4.6%	6.9%	13.6%	2.2%

Conclusions

The predicting model of freeway section normal travel time using the GA improved BP neural network was established in the paper. The input, output parameters, basic structure, and data acquisition plan were analyzed. The calculation result indicates that the model can effectively predict the target, and the GA improved way can increase the learning rate and decrease the probability of falling into local convergence. The GA-BP is proved to have a better convergence rate and predicting accuracy.

References

Bajwa, S., and Islam, U. (2003). "A travel time prediction method based on pattern matching technique." *Proceedings - Conference of the Australian Road Research Board*, v 21, 2003, Proceedings - 21st ARRB and 11th REAAA Conference, Transport Our Highway to a Sustainable Future, 997-1010.

Boozarjomehry, R. B., and Svrcek, W. Y. (2001). "Automatic design of neural network structures." *Computers and Chemical Engineering*, 25: 1075-1088.

Chang, G. L., and Shrestha, D. (2000). "Performance evaluation of CHART – an incident management program – in 1997." Final Report. 25-30.

Chen, G., Wang, X., and, Zhuang, Z. (2001). "Genetic algorithm and its application."[M] *Beijing: Posts and Telecom Press*. 12-16.

Chen, J., Li, C., and Ding, G. (2008). "Traffic safety assessment of freeway based on BP neural network." *Journal of Tongji University (Natural Science)*, 36(7), 927-931.

Hassoun, M. H. (1995). "Fundamentals of artificial neural networks." *Cambridge: MIT Press*. 3-8.

Ayako Hiramatsu, Kazuo Nose, Kenji Tenmoku and Takeshi Morita (2009), "Prediction of travel time in urban district based on state equation." *Takeshi Source: Electronics and Communications in Japan*, 92(7), 1-11.

Lili Huang and Matthew Barth (2008). "A novel loglinear model for freeway travel time prediction." *IEEE Conference on Intelligent Transportation Systems*, Proceedings, ITSC, 210-215.

Hu, S., Yu, S., and Dai, K. (1993). "The introduction of neural network." *Beijing：National university of defense technology Press*. 72-78.

Wei-Hsun Lee, , Shian-Shyong Tseng and Sheng-Han Tsai(2009). "A knowledge based real-time travel time prediction system for urban network." *Sheng-Han Source: Expert Systems with Applications*, v 36, n 3 PART 1, 4239-4247.

Lingras, P., and Mountford, P. (2001). "Time delay neural networks designed using genetic algorithms for short- term inter-city traffic forecasting." / / *Lecture Notes in Computer Science , Engineering of Intelligent System*s. [S.l.]: Springer-Verlag GmbH. 290-299.

Liu, H., Zhou, C., Zhu, A., and Li, L. (2009). "Multi-population genetic neural network model for short-term traffic flow prediction at intersections." *Acta Geodaetica et Cartographica Sinica*. Vol. 38, No . 4, 363-368.

Liu, Y., Shen, H., and Peng, G. (2007). "Travel time model of emergency vehicle based on BP network." *Journal of Wuhan University of Technology (Transportation Science & Engineering)*, Vol. 31, No. 6, 973-975.

Nee, J., and Hallenback, M. E. (2001). "Evaluation of the service patrol program in the Puget Sound Region." *Washington State Transportation Center, Project No. T1803, Task 37*. 15-18.

Wang, W., and Guo, X. (2000). "Traffic eEngineering." *Nanjing: Southeast University Press,* 80- 85.

Yan, P., and Zhang, C. (2000). "The neural network and calculation of simulated evolution." *Beijing: Tsinghua University Press*. 28-36.

A Preliminary Test on the Interactions between Land Use and Travel Models

Shengyi Gao[1], Eric Lehmer[1], Yang Wang[1], Michael McCoy[1], Robert A. Johnston[2], Dimantha De Silva[3], John Abraham[3], John Douglas Hunt[3]

[1] Urban Land Use and Transportation Center, University of California, One Shields Avenue, Davis, CA 95616
[2] Department of Environmental Science and Policy, University of California, One Shields Avenue, Davis, CA 95616
[3] HBA Specto, Inc., Calgary, T2P 3P4, AB, Canada

Abstract

In this paper, we report the preliminary results of the sensitivity test for California Production, Exchange and Consumption Allocation System (PECAS). The sensitivity test includes three scenarios: *PECAS Only*, *Integrated*, and *High Travel Cost*. The *PECAS Only* scenario allocates production and consumption at a fixed travel cost from years 2000 to 2020. It shows that the floorspace growth is more affected by general plan than travel cost, and residential floorspace types have higher growth rates in suburb, exurb and rural areas. Additionally, nonresidential (industrial and commercial) floorspace types have higher growth rates in the inner urban areas. The *Integrated* scenario has a more compact land use pattern than the *PECAS Only* scenario. Compared with the *Integrated* scenario, the *High Travel Cost* scenario leads to more compact use. The results from all three scenarios imply that the model behaves as expected and the outputs are consistent with what has been theorized.

Introduction

In the Unites States, a major impetus of developing travel and land use models comes from the mandatory conformity to federal and state planning regulations. Required by federal and state laws, travel models are widely used by Metropolitan Planning Organizations (MPO) in making Regional Transportation Plans (RTP) and Transportation Improvement Programs (TIP), as well as for assessing the environmental impacts of transportation related projects. And while researchers have demonstrated that land use has significant impacts on travel behavior and have strongly recommended an integrated land-use/transport modeling approach (Bagly and Mokhtarian, 2002; Cao et al, 2006, 2007, 2009; Cervero, 1996; Cervero and

Kockelman, 1997; Frank and Pivo, 1995; Gao et al, 2007; Johnston et al, 2005; Handy, 1996; Hunt et al, 2001, 2005; Rodier and Johnston, 2002), the majority of MPOs do not have a land use model to forecast employers' and households' residential location choices in their long range planning.

In California, Assembly Bill 32 (AB 32) from the 2005-2006 session set up targets for California to reduce emissions of greenhouse gases. Additionally, Senate Bill 375 (SB 375) from the 2008 session requires the MPOs to address the approaches to reach these targets. The two bills have largely changed the context of modeling efforts and have led to a transition of focus from travel models to land use models. At present, the four largest MPOs in California (Southern California Association of Governments, San Diego Association of Governments, Association of Bay Area Governments, and Sacramento Area Council of Governments) and California Department of Transportation (Caltrans) are developing their land use models. After reviewing all operational land use models (Johnston and McCoy, 2006), they have all chosen an input-output framework (Abraham et al, 2005; Hunt and Abraham, 2005; Hunt et al, 2006; Zhong et al, 2007).

Model Framework

PECAS is a statewide spatial economic model which consists of Activity Allocation (AA) and Space Development (SD) modules. Its AA module has a three-level nested logit structure (see Figure 1). At the first choice level, the AA module allocates the quantities of aggregate activities (which include industry sectors and households) to each Land Use Zone (LUZ). At the second level, the module allocates the quantity of each activity in each LUZ among the available technology options. At the lowest level, the AA module allocates the quantity of each commodity produced in a given LUZ among the exchange locations (where it is sold). The total quantities of activities are exogenous inputs at the model level (Hunt and Abraham, 2005; Zhong et al, 2007).

Because of the major differences in product types, the activities in PECAS can be roughly grouped as industries, governments, households and imports/exports. Industries produce goods and services by consuming goods and services, labor, and space at elastic rates. Governments administrate financial accounts (its production and consumption functions are a part of the industry). Households consume goods, services, and space to produce labor at elastic rates as well. Imports/exports, which we categorize as a special activity, deal completely with all imports and exports. Space, as a special commodity, is sold by its owners and bought by industries and households at the market price at which the supply and demand reach equilibrium. The SD module adopts a nested structure (see Figure 2) as well. Its role is to use the market price signal generated by the AA module and other location factors to forecast the probabilities of land development events and quantity of each space type to be developed. It then feeds the quantity of each land use type back to the AA module. The market price of the space type will be determined by the AA module in the next year.

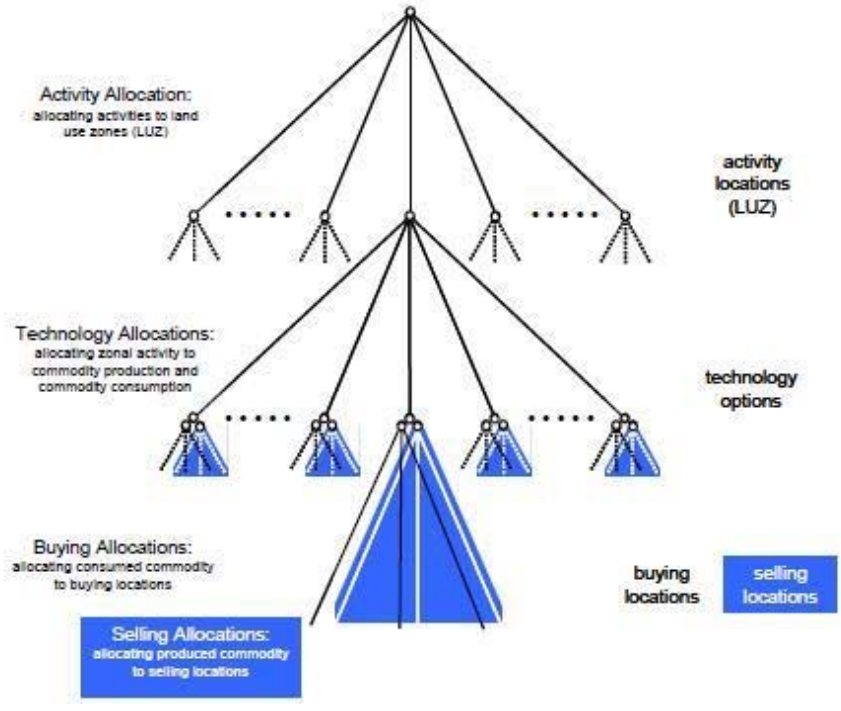


Figure 1. Nested logit structure in AA.

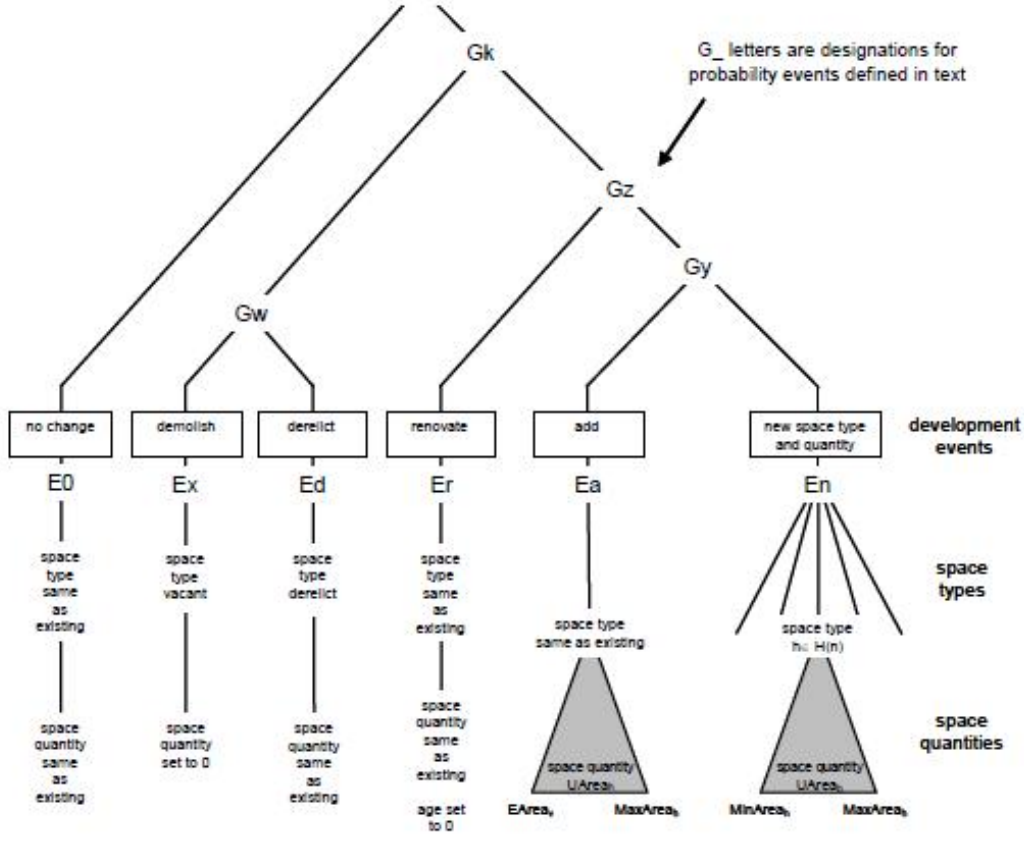


Figure 2. Nested logit structure in SD.

Travel costs computed by travel time and distance are two important components of the utility functions of buying and selling commodities. Combining with price and size term which represents agglomeration effects of economy, they determine the allocation of the quantities of commodities bought and sold in each LUZ. Due to the nested structure, the travel costs will affect the utilities of technology options (the second level of the logit structure) and activity location choice (the first level of the logit choice). Additionally, they will influence the allocation of production and consumption of goods, services, labor, and space, as well as the allocation of the quantities of activities (Figure 3).

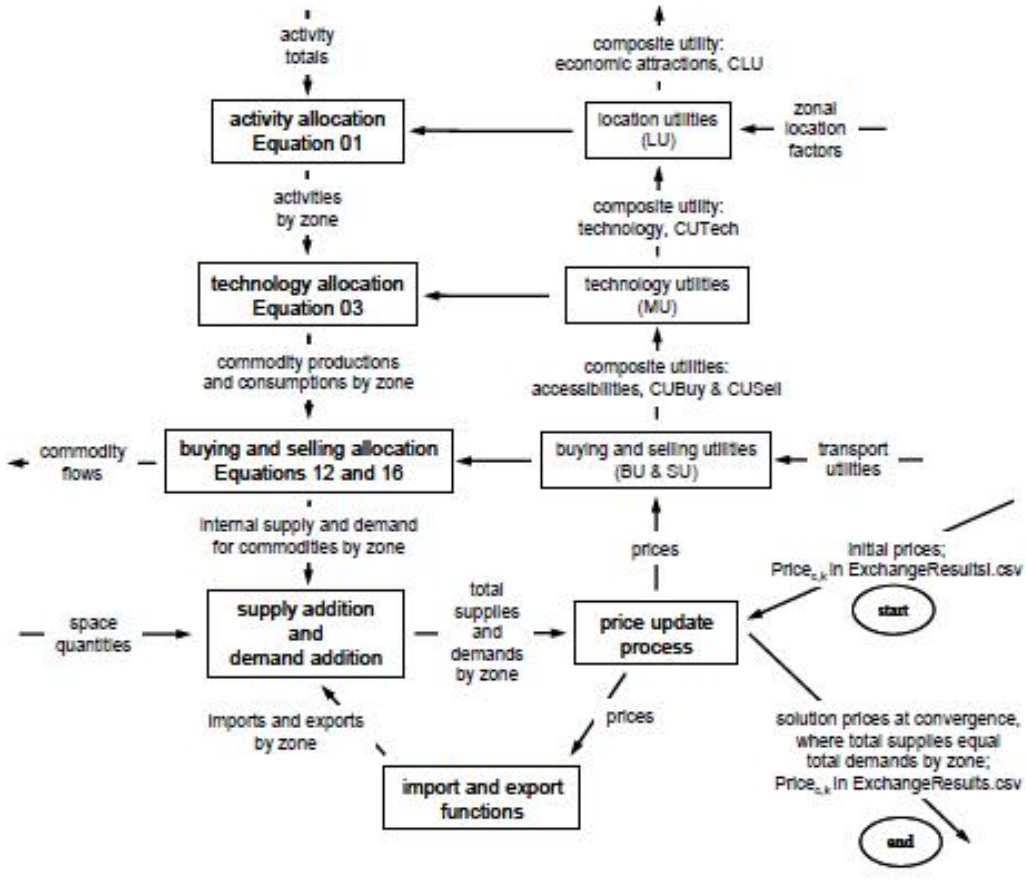


Figure 3 Allocation and price updating process

What makes PECAS different from other land use models is its mechanism of a price search process. In PECAS, each commodity has a price. The price of a commodity is determined by supply and demand of the commodity. Supply and demand are generated from their respective study area and external zones (other areas of a county and foreign counties). When supply and demand of a commodity become equal, the model reaches equilibrium. In contrast to goods and services, floorspace cannot be imported or exported: its price is completely determined by the supply and demand of the floorspace within a land use zone. Therefore, PECAS is a fully market-based

model. This feature makes PECAS a very powerful tool to analyze the impacts of policies in economy, land use and environmental protection. The algorithm of nested logit choice and price search is sketched in Figure 3.

The time step of PECAS is one year, while it is generally five years for a travel model. We integrate PECAS with the statewide travel model at 2005, 2010, and 2015. In these years, the travel model uses the number of workers and households (which are forecasted by PECAS' AA module) to calculate the person and vehicle trips, and then assigns vehicle trips to the highway networks. The new time/distance skims are used by PECAS in 2006-2010, 2011-2015, and 2016-2020 respectively to calculate the travel costs for buying and selling activities. Through the integration, a linkage between land use and travel is built (see Figure 4). Any substantial changes in land use are expected to cause some changes in the person trips between the origin and destination zones, corresponding changes in the interzonal travel times between some zone pairs, and vice versa.

Sensitivity Test Scenarios

We designed three scenarios to test the interactions between PECAS and the statewide travel model. In the *PECAS Only* scenario, PECAS is not integrated with the travel model, and we used the same skims for all years (2000-2020). This way, the travel time and distance do not change over time, and the travel costs (time-based and distance-based costs) are constant. In other words, the marginal contribution of transportation to the changes of land use pattern over time is zero. Therefore, it is used as a reference of the *Integrated* scenario.

In the *Integrated* scenario, PECAS and the travel model are integrated as shown in Figure 4. Since population and employment continue to grow over time, it is plausible to assume that land use in the year when PECAS is integrated with the travel model is different from that in previous years. These changes will alter some basic assumptions about household types and trip generation at the TAZ level. They will also change trip distribution and assignment. In the successive year, PECAS will use the new skims to calculate the transport utilities for commodity exchanges. Due to the interaction between PECAS and the travel model, we will expect to see substantial differences in land use and zone-to-zone travel time

The third scenario is a *High Travel Cost* scenario. In this scenario, the composite unit travel cost is doubled to represent a high gasoline price ($8 per gallon in year 2000 dollars). With the exception of the higher travel cost, all other inputs for this scenario were the same as those in the *Integrated* scenario. The hypothesis behind this scenario is that the high travel cost constrains people from making long commutes and other discretionary trips, and therefore leads to a more compact land use pattern. The high travel cost will strengthen the agglomeration effect of industries.

Results and Discussion

In the base scenario, the population and employment grew at 1.2% annually from 2000 to 2020. SD transition constants represent a 2.4% annual transition rate. From the perspective of the sensitivity test, we want to know if the changes in the quantity and price of floorspace and activities are consistent with those pointed by theory. Firstly, if the model behaves normally, the total floorspace will grow monotonically. The relative quantities of the floorspace in the forecast years will reflect the relative quantities in the base year. From Figure 5, we can see that this is true. An intuitive question rising up from Figure 5 is what type of space is converted into residential and industry uses. The answer is agricultural space (including vacant space). In the SD module, agricultural space is allowed to be used for non-agricultural purposes and it has a lower transition cost than other space types. In particular, agricultural lots in California are usually large and are easier for developers to purchase for large development projects. Therefore, agricultural space declines over time. This is what has happened in the real world.

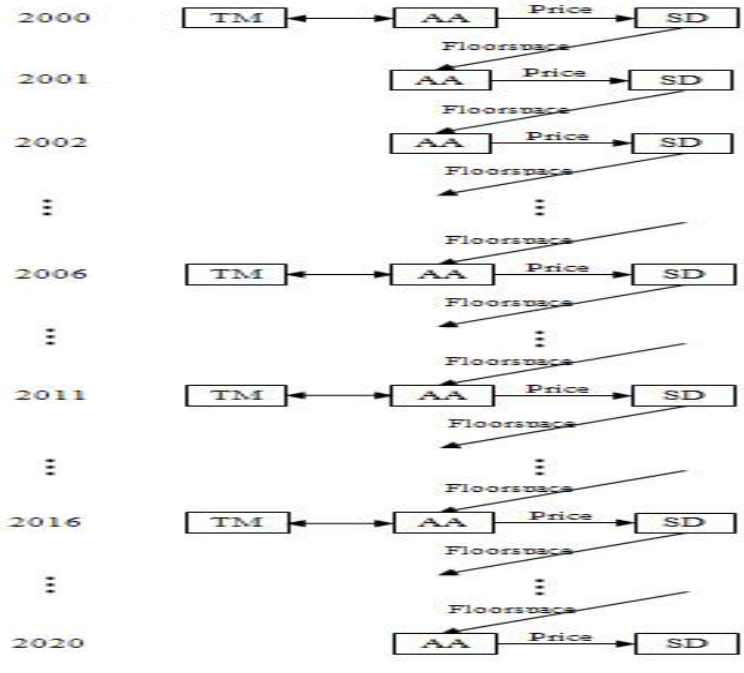


Figure 4. AA, SD and travel model integration.

Secondly, the higher growth rate of floorspace, as compared to households and employment, will generally lead to a lower floorspace price over time. In PECAS, the supply of floorspace for the AA module in year *t* is determined by the SD module in year *t-1* according to the floorspace price in year *t-1*. Thus, the floorspace in year *t* will be determined by the demand for floorspace. The lower growth rate of activities implies lower demand for floorspace. To make the supply and demand identical, we must lower the floorspace price so that activities can use more space. This way, the floorspace declines year by year. Figure 6 shows the floorspace prices of single family detached and commercial low units and confirms the assumption.

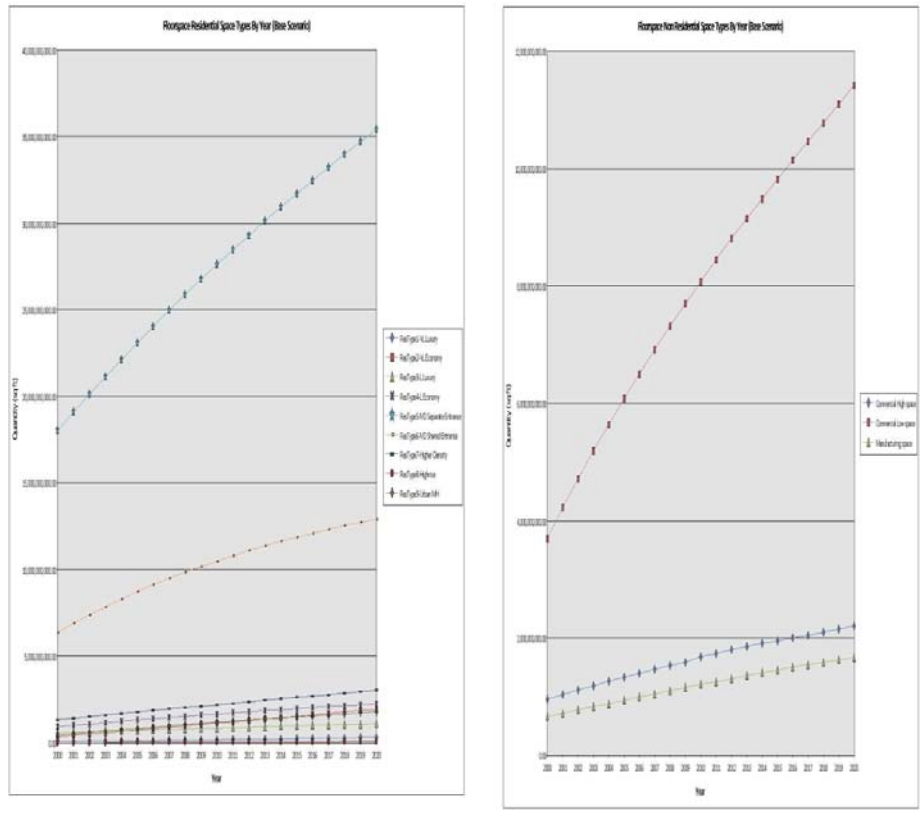


Figure 5 Trend of residential (left) and non-residential (right) floorspace over time (2000 -2020).

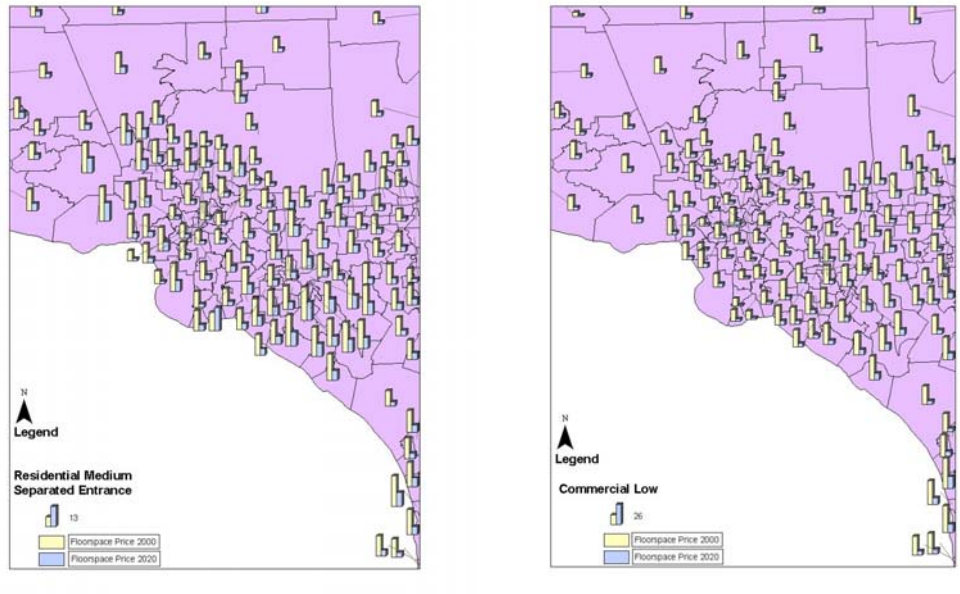


Figure 6 Floorspace Prices for Single Family Detached (left) and Commercial Low Density (right) in 2000 and 2020, Los Angeles Area.

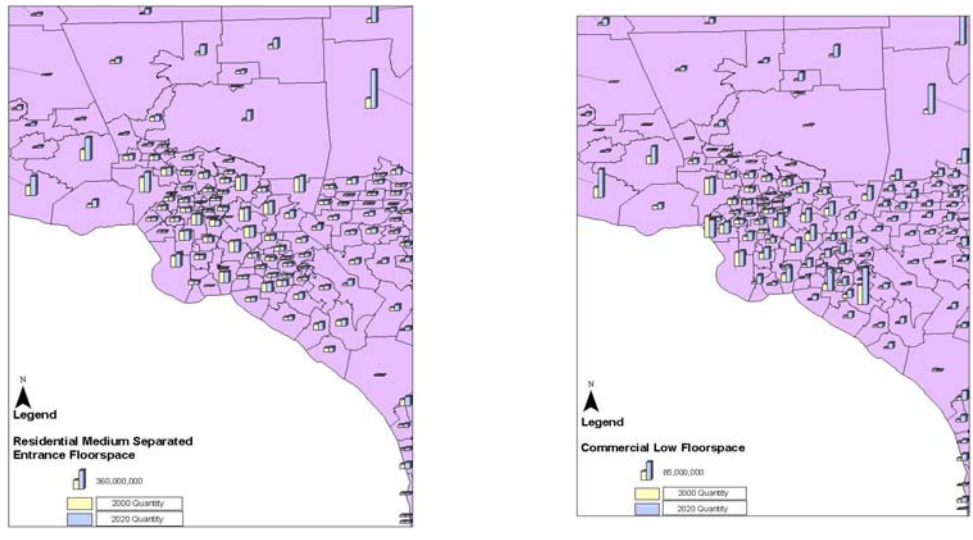


Figure 7 Quantities of floorspace of single family detached (left) and commercial low density (right) in 2000 and 2020, Los Angeles area

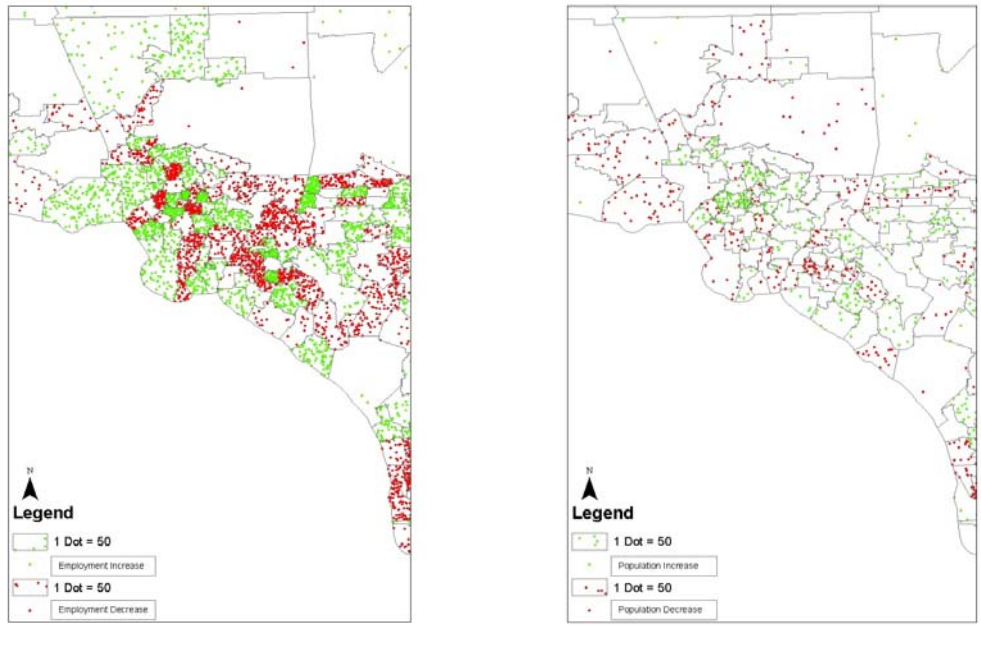


Figure 8 Difference of employment (left) and households (right) between *PECAS Only* and the *Integrated* scenarios in 2020, Los Angeles area

Thirdly, the spatial pattern of floorspace growth (which is proportional to households and employment) must reflect the dominant roles of land use policies. In California, a general land use plan is a local land use regulation and guides the land use in a relatively long period (generally 20 years). Zoning regulations detail land use in general plans to parcels under the guideline of general land use plans. Limited amendments are allowed annually. A land owner must pay additional development fees to obtain an amendment petition to change the original land use designation in the general plan. Thus, from the perspective of modeling, the development events in the SD module will be largely but not completely determined by a general land use plan. In other words, a specific potential floorspace development event will most likely happen in the areas designated for that type of floorspace. For example, in suburban and rural areas, there are more designations for residential than for commercial. The residential floorspace growth rate in those areas will likely be higher than the inner urban areas where there are more commercial designations. Generally speaking, inner urban areas have more commercial zoning while suburbs, exurbs and rural areas have more single family detached zoning. Therefore, in a twenty-year span, suburbs, exurbs, and rural areas will have increased growth of residential floorspace while inner urban areas will have more growth of commercial floorspace. Figure 7 is a snapshot of the quantities of floorspace of single family detached and commercial units in 2000 and 2020 in the Los Angeles area. The spatial patterns of two types of floorspace are the same as expected.

In the *Integrated* scenario, we first examined the interzonal travel time in 2005, 2010 and 2015. From 2000 to 2005, about 1.0% of the total zone pairs had an increase of travel time of more than 10.0%. The largest increase of travel time was 33.8%. For this time period, the cumulative growth of households is 6.0%. From 2006 to 2010 and from 2011 to 2015, similar changes in travel time between zones are observed.

The largest cumulative increase from 2000 to 2015 is as high as 227.1%. Obviously, the disproportionate increase of travel time to the gross growth rate of households can only be explained by the spatial variation in growth of households. This implies that we have observed the impacts of land use on the travel model.

On the land use side, compared with the *PECAS Only* scenario, the *Integrated* scenario has a much higher travel cost in 2020. In this case, the households which are sensitive to travel costs will choose to live closer to where they work. Industries will choose to be closer to their clients and their supplies of goods and services in order to lower total transportation costs. In other words, higher transportation costs will lead to a more compact land use pattern. As Figure 8 shows, the differences of households and employment between the *PECAS Only* and the *Integrated* scenarios (in the Los Angeles area) are substantial. Since the only difference in the model inputs is the time skims, the difference in the spatial pattern of land use between the two scenarios can be fully attributed to travel time related to travel cost. In this regard, we can see that the impacts of the travel model on land use are substantial, not only in the quantities of activities and space, but also in the spatial locations of the activities.

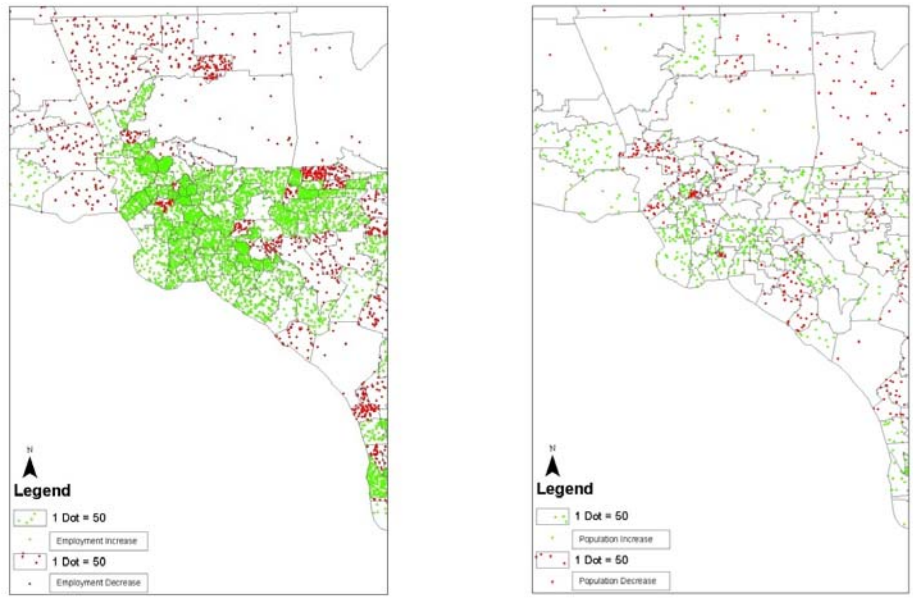


Figure 9 Difference of employment (left) and households (right) between the *Integrated* and the *High Travel Cost* scenarios in 2020, Los Angeles area.

The spatial patterns of activity distribution and space consumption in the *High Travel Cost* scenario are similar to those in the *Integrated* scenario, but tighter (see Figure 9). Both industry and households are elastic to the change of travel costs, but industry is more sensitive to the substantial increase in travel costs.

Conclusions

In this paper, we report only a small part of the outputs of the integrated land use/transportation model. This heuristic analysis demonstrates how a travel model and a land use model interact with each other, as well as the importance of integrating a travel model with a land use model to improve the forecasting accuracy of land use and travel demand. It provides more supportive evidence to the Addendum to the 2007 Regional Transportation Plan Guidelines that requires the MPOs in California to adopt a land use model in their Regional Transportation Plan analysis to address the impacts of land use on transportation, and more specifically, applying land use policy to address how to reduce vehicle miles traveled.

What we present in this paper is only a small part of the output potential of PECAS. Other outputs at the LUZ level include the quantities of activities, quantities of commodities produced and consumed by activities, quantities of commodities exchanged, commodity price, LUZ-to-LUZ commodity flows, average commodity transport distance, and welfare measures. These outputs can be utilized in policy analyses such as housing-job balance, habitat protection, agriculture land protection, equality analysis, etc.

The input-output structure of the model in technology options provides policy analysts with an opportunity for forecasting greenhouse gas emissions within a small geographic area, assessing how a carbon tax may change industries' technology options and households' lifestyles, and evaluating the impacts of a carbon tax and carbon trade on a regional economy. It should be emphasized that a carbon tax is exogenous but carbon trade is endogenous. The carbon tax and carbon trade will change not only activities' location choice, but also activities' production, consumption, exchange, and even imports and exports. The location-choice-based land use model can only deal with the former and thus has obvious structural flaws. In our next generation model, we will add a carbon tax as a commodity and test how sensitive the regional economy will be to it.

Acknowledgements

This project is funded by California Department of Transportation. The authors wish to thank the Technical Advisory Committee members from California Department of Energy, California Air Resources Board, San Diego Association of Governments, Southern California Association of Governments, Association of Bay Area Governments, and Oregon Department of Transportation for their valuable suggestions in the process of developing this model.

References

Abraham J.E., G. Garry, and Hunt J.D. (2005). "The Sacramento PECAS Model",*Proceedings of the 84th Annual Transportation Research Board Conference*, Washington DC, USA, January.

Bagley, M. N., and P. L. Mokhtarian. (2002)."The Impact of Residential Neighborhood Type on Travel Behavior: A Structural Equations Modeling Approach", *Annals of Regional Science,* 36 (2), 279–297.

Cao, X., P. L. Mokhtarian, and S. L. Handy. Neighborhood design and vehicle type choice: evidence from Northern California. *Transportation Research D*, 11(2), 133-145.

Cao, X., P. L. Mokhtarian, and S. L. Handy. (2007) "Cross-sectional and quasi-panel explorations of the connection between the built environment and auto ownership", *Environment and Planning A*, 39(4) , 830-847.

Cao, X., P. L. Mokhtarian, and S. L. Handy. The relationship between the built environment and nonwork travel: A case study of Northern California. *Transportation Research A*, 43(5), 548-559.

Cervero, R. 1996. Mixed Land-Uses and Commuting: Evidence from the American Housing Survey. *Transportation Research A,* 30(5), 361–377.

Cervero, R. and K. Kockelman. Travel demand and the 3Ds: Density, Diversity and Design. *Transportation Research Part D*, 2(3), 199-219.

Frank, L. D., and G. Pivo. Impacts of Mixed Use and Density on Utilization of Three Modes of Travel: Single-Occupant Vehicle, Transit, and Walking. *Transportation Research Record: Journal of the Transportation Research Board, No.* 1902, Transportation Research Board of the National Academics, Washington, D.C., 1995, 44–52.

Gao, S., P.L. Mokhtarian, and R. A. Johnston. Exploring the causal connections among job accessibility, employment, income, and auto ownership with structural equation modeling. *The Annals of Regional Science,* 42(2), 341-356.

Handy, S. L. Methodologies for Exploring the Link Between Urban Form and Travel Behavior. *Transportation Research D,* 1(2), 151–165.

Hunt, J.D. and J.E. Abraham. Design and implementation of PECAS: A generalized system for the allocation of economic production, exchange and consumption quantities; Chapter 11. In: Foundations of Integrated Land-Use and Transportation Models: Assumptions and New Conceptual Frameworks, Elsevier, Oxford UK, 2005, 217-238.

Hunt, J.D., J.E. Abraham, M. Zhong, and A.T. Brownlee. Using input-output tables and social accounting matrices in the development of land use transport interaction models. *Proceedings of the 85th Annual Transportation Research Board Conference*, Washington DC, USA, January 2006.

Hunt, J.D., D.S. Kriger, and E.J. Miller. Current operation urban land-use-transport modeling framework: A review. *Transport Reviews*, 25(3), 329-376.

Hunt, J.D., R.A. Johnston, J.E. Abraham, C.J. Rodier, G. Garry, S.H. Putman, and T. dela Barra. Comparison from the Sacramento Model Testbed. *Transportation Research Record: Journal of the Transportation Research Board, No.* 1902, Transportation Research Board of the National Academics, Washington, D.C., 2001, 53-63.

Johnston, R. A., S. Gao, and M. J. Clay. Modeling Long-Range Transportation and Land Use Scenarios for the Sacramento Region, Using Citizen-Generated Policies. *Transportation Research Record: Journal of the Transportation Research Board, No.* 1902, Transportation Research Board of the National Academics, Washington, D.C., 2005, 99-106

Johnston, R.A and M. McCoy. Assessment of integrated transportation/land use model: Final report. 2006. Available at: http://pecas.ultrans.ucdavis.edu/resource/23

Rodier, C. J., R. A. Johnston, and J. E. Abraham. Heuristic policy analysis of regional land use, transit, and travel pricing scenarios using two urban models. *Transportation Research Part D*, 7(4), 243-254.

Zhong, M., J.D. Hunt, and J.E. Abraham. Design and development of a statewide land use transport model for Alberta. *Journal of Transportation Systems Engineering and Information Technology*, 7(1), 79-89.

Assessment of the Congestion Impacts of Fatal Accidents using SCOOT Data

Leighton M.O. Cardwell[1], Haibo Chen[2] and Margaret C. Bell[3]

[1] Senior Consultant, MVA Consultancy UK, Second Floor, 37a Waterloo Street, Birmingham B2 5TJ; E-mail: lcardwell@mvaconsultancy.com
[2] Principal Research Fellow, Institute for Transport Studies, University of Leeds, Leeds LS2 9JT; E-mail: trahc@leeds.ac.uk
[3] Professor, School of Civil Engineering and Geosciences, Newcastle University, Newcastle NE1 7RU; E-mail: margaret.bell@ncl.ac.uk

Abstract

This study focuses on the assessment of the congestion impacts of fatal accidents. The paper addresses the current paucity of research into accident induced congestion levels outside of Motorway/Freeway environments, and using the SCOOT traffic control system, provides an assessment of these impacts in more heterogeneous urban environments. Levels of induced congestion are found to vary significantly according to the temporal and spatial settings in which they occur, with only accidents that occur in either peak periods or constricted network layouts producing significant induced congestion externalities in monetary terms. The detailed spatial and temporal resolutions required for such forecasts are subsequently discussed, as these raise a number of issues for SCOOT and ITS data more generally, if such analyses are to become more detailed, sophisticated, and ultimately automated, across a range of complex temporal and spatial settings.
Key Words: congestion, fatal accidents, SCOOT, travel time savings, ITS data.

Introduction

Both congestion and road safety are key objectives of transport policy, wherever considered. However, surprisingly little research has focused on the inter-relationship between them. Any that has dealt with motorways and trunk roads, rather than urban areas (e.g. Highways Agency, 2004, Caltrans 2007). The Highway Agency, suggests that for UK motorways and trunk roads, 25% of all congestion is caused by accidents; the other 75% is estimated to arise due to the sheer volume of traffic (65%), and roadworks (10%). Caltrans (2007) in the US, similarly state that whilst "non-recurrent congestion caused by accidents, disabled vehicles, spills, weather events, and visual distractions is speculated to account for one-half to three-fourths of the

total congestion on metropolitan freeways, there are insufficient data at present to either confirm or deny this conjecture."

The focus of this paper is therefore related to the impacts of fatal accidents in *urban* areas, where other factors such as network density, alternative routing, and spatial, and temporal variations in traffic flow are likely to be prominent. For example, speed-flow curves are known to vary by road type and area, whilst different flow levels on roads within urban areas will affect the amount of 'slack' in the road system, and its ability to recover from accident induced effects. The well documented situation of heightened delays and congestion caused by roadworks in the peak compared to the off-peak or on heavily-trafficked links compared to those that are lightly trafficked serves as a similar analogy.

Data Collection and Methodology

In order to collect real-time data regarding total levels of congestion and corresponding changes after an accident, there is a need for a specially designed detection system that detects congestion over a significant areal extent *and* that records the traffic conditions on them consecutively for long periods of time. From this data, analytical methodologies are required in order to quantify in time and space, and subsequently identify, the factors affecting the total levels of congestion caused by an accident.

In order to achieve the required spatial and temporal resolution, occupancies from a series of loop detectors were collected, as used by the demand responsive control system SCOOT (Split Cycle Offset Optimisation Technique, Hunt *et al*, 1981), to derive link-based traffic characteristics such as flow, delay, stops, and congestion measures. SCOOT mainly uses data from inductive loops embedded in the road surface, and optimises traffic signal settings to reduce vehicle delays and stops, responding to variations in traffic on a cycle-by-cycle basis. However, as SCOOT relies on good traffic data for its operation, it is also a major source of on-line traffic data. The data provided from the SCOOT area of Leicester are used as the primary data source for this paper, although at a more general level, SCOOT and other comparable systems have been demonstrated in over 200 towns and cities in over 14 countries around the world.

Data Availability

Accident information for congestion analysis was acquired through the UK STATS19 database, prior to mapping, and checking against the SCOOT database for availability of flow, delay and congestion data. This process provided 46 accidents that were potentially open to analysis that were spatially synonymous with the SCOOT system. Whilst this provides the aggregate data-set, several other data availability factors also intervened to further reduce the number of accidents that could be analysed. These include:

- The accident affected link not being in the SCOOT database due to storage issues;
- SCOOT malfunctions at time of accident;

- The SCOOT system being non-operational for an extended period of time making yearly profiling difficult (e.g. long-term repairs/ roadworks);
- Many links surrounding the accident not being in the SCOOT database/ malfunctioning; and,
- Accidents occurring at the very edge of a SCOOT region, with too many non-SCOOT links affected. No retrospective data are available for these.

These issues of data availability left a total of 6 fatal accidents that could be analysed across Leicester between 1998 -2005. As the majority of data availability issues related to long term data loss, or intermittent link malfunctions over extended periods of time, greater future emphasis within ITS systems, is required to be placed in continual and unbroken data storage and malfunction detection mechanisms, which is the only way in which sufficient real-time information for analysis and automated applications that rely on highly spatial and temporal data-sets can be provided.

'Typical' Profiling

An understanding of the variation in congestion levels is fundamental to determining whether the particular accident creates a statistically significant change. This is usually difficult when there is large variation in congestion and other indicators over time- whether within hours, between hours, over days, weeks, or years. As a result, the evaluation of a change between the pre- and post-accident states, requires the definition of 'typical' conditions in order to compare to those that occur on the accident day, in order to assess change. It is important however, that in comparing changes which may be small in comparison with huge day-to-day variations that the methodology behind the derivation of 'typical' data has to be achieved by removing outliers in a statistically robust manner. This process is achieved through the use of a computer program, SAM (Statistical Analysis Module), which is based on a previous version OULF developed by Grant and Bell (1986).
The program calculates the outliers at 95% and 99% significance, using median values as a reference. Having proved a statistically significant difference between the use of Weekday, Saturday and Sunday profiles, SAM calculates the profiles from each set of 5-minute data produced by the SCOOT M02 message, using the standard deviation, the number of observations (days), and the upper and lower confidence limits using appropriate $t_{.}$ or $z_{.}$ values for corresponding 'n' for each of the Weekday, Saturday and Sunday profiles. Such methodologies are already well established for using these ITS archives to produce information on historically-based, 'typical' conditions, informing transport users of the likely conditions of travel, especially in the UK context from SCOOT systems (see Evans 1995, Goodman 2001).

For each fatal accident, 'typical' and 'accident-day' profiles of the flow, congestion and delay data from the SCOOT M02 message are produced for every applicable link within the SCOOT Region. An example of the comparative profiles are shown in Figure 1, with the typical delay profiles across the day highlighted in blue (with confidence intervals marked), the accident profile in red. Data availability issues may also be noted by the intermittent nature of the red, accident day line across the day.

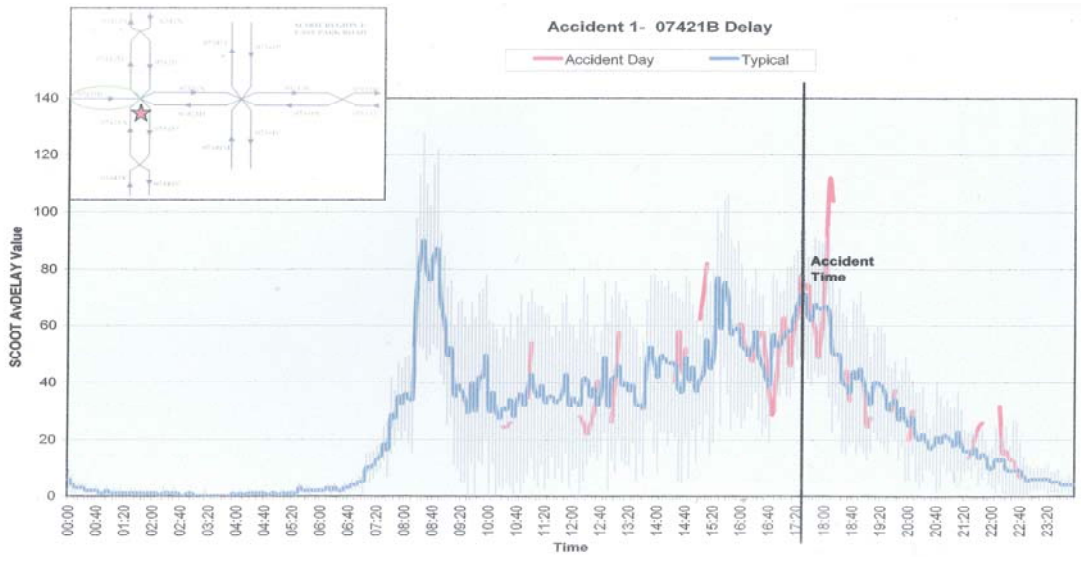


Figure 1. Typical and accident day profiling for a single link.

Spatial Specification

Clearly more than one link in each region is likely to be affected by the consequences of a fatal accident. Moreover, there is likely to be distance decay in terms of the induced congestion from accident sites. To capture the full extent of 'knock-on' effects from one region to another, as many links as possible are analysed, progressively further away from the accident site, until the limit is reached where no further statistically significant change from the typical is observed. This principle is illustrated in Figure 2. Further, where the effects of the accident reached the edge of the existing SCOOT region concerned 'knock-on' effects would then be looked for in surrounding SCOOT links and regions, until no statistically significant link results are found.

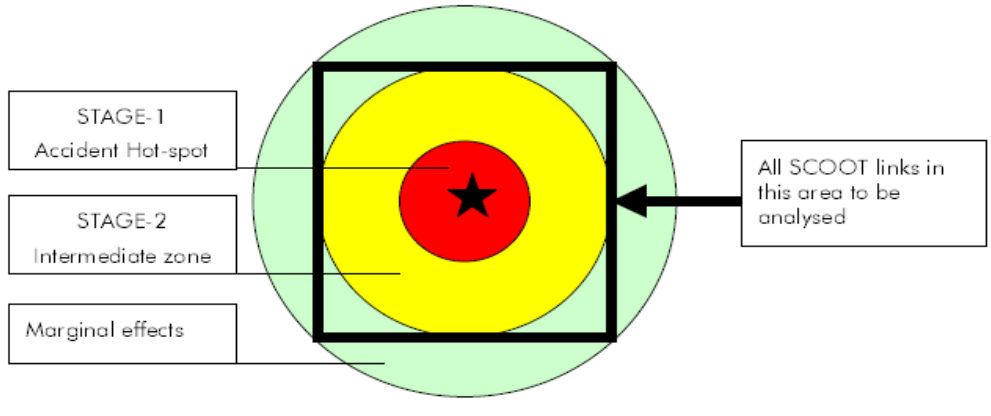


Figure 2. Spatial delineation of accident day profiling.

Temporal Specification

An objective and robust definition also needs to be produced for the temporal 'end-point' of the accident. For each parameter on every link, 'typical' and 'accident' profiles are compared, to which the standard deviation of the typical profile is also added. Differences are calculated for every 5-minute period post-accident as a raw numbers *and* in terms of standard deviations away from the typical profile in order to assess statistical significance. Assuming normal distributions, accident values above or below two standard deviations are very likely (95%) to be statistically significant.

Whilst, such calculations are useful to help identify the statistical end-point of an accident within a given confidence level, five-minute periods are *not* independent from one another, but rather cyclic, whilst natural stochastity *between* days rather than over a day- tends to produce systematic variations in data over several consecutive periods rather than true random variation.

As a result, an improved method of estimating the statistical 'end point' for the effects of an accident was developed, based upon the point where the standard deviation induced by an accident returns to zero. Thus the statistical end point can be referred to as the point where the deviations induced, cross from a positive value to a zero (or negative value), or from a negative deviation to a zero or positive value. This point is taken as the 'end' for the accident calculation purposes, provided subsequent deviations remain small, and within the range of typical variability.

Whilst this is a necessary condition of the end point of an accident, it alone is not a sufficient condition. Some attention must be paid to secondary effects of an accident-those impacts that are delayed in time from the actual accident event itself. Deviations that cannot be attributed to the accident in time (several hours after typical conditions have returned on other adjacent links), or space, (far from other affected links) beyond the initial 'cross-over' of the standard deviation pattern, are not be included within the aggregate results. This has particular advantages over fixed time period analysis, or standard deviation analysis in isolation, and represents the methodology used in each of the subsequent accident-induced congestion analyses.

Results and Analysis

The above 'typical' and 'accident-day' profiling methodology was used for analysis of each link and region statistically affected by an accident. In order to enhance the level of understanding and development of congestion after each accident, GIS-based spatial plots of the induced-congestion effects were also produced. Analysis for Accident 1 is shown in Figure 3 overleaf as an example, with links affected being colored according to the level of additional congestion induced.

Yellow shows links that experience congestion between 10-50% over and above typical levels, orange links show congestion over and above 50-100% of typical,

whilst red indicates greater than 100% of typical congestion levels. Black shows links that had no data available, either because of malfunctions, SCOOT errors, loss in data transmissions, or longer term factors that prevent a statistically 'typical' level of flow, delay, and congestion being developed. The location of the accident itself is starred. Aggregate analysis of the congestion impacts of each fatal accident is shown in Table 1.

The results show that significant amounts of congestion are only produced when an accident occurs either in a peak hour, as with Accident 1, or within a constricted spatial setting with few diversionary opportunities, Accident 4. Congestion is the first instance rises to levels 127% greater than typical over the whole area affected by the accident, whilst Accident 4 located on a major, one-way road, produced values in excess of 4000% of typical conditions, noting that absolute levels of congestion in the off-peak are significantly lower. Nevertheless, the absolute level of congestion caused is also significant, 4,826 four-second SCOOT intervals, or 253 vehicle hours of delay. Other fatal accidents analysed that are not in peak travel periods, or in constricted network locations, produce very limited levels of accident induced delay and congestion.

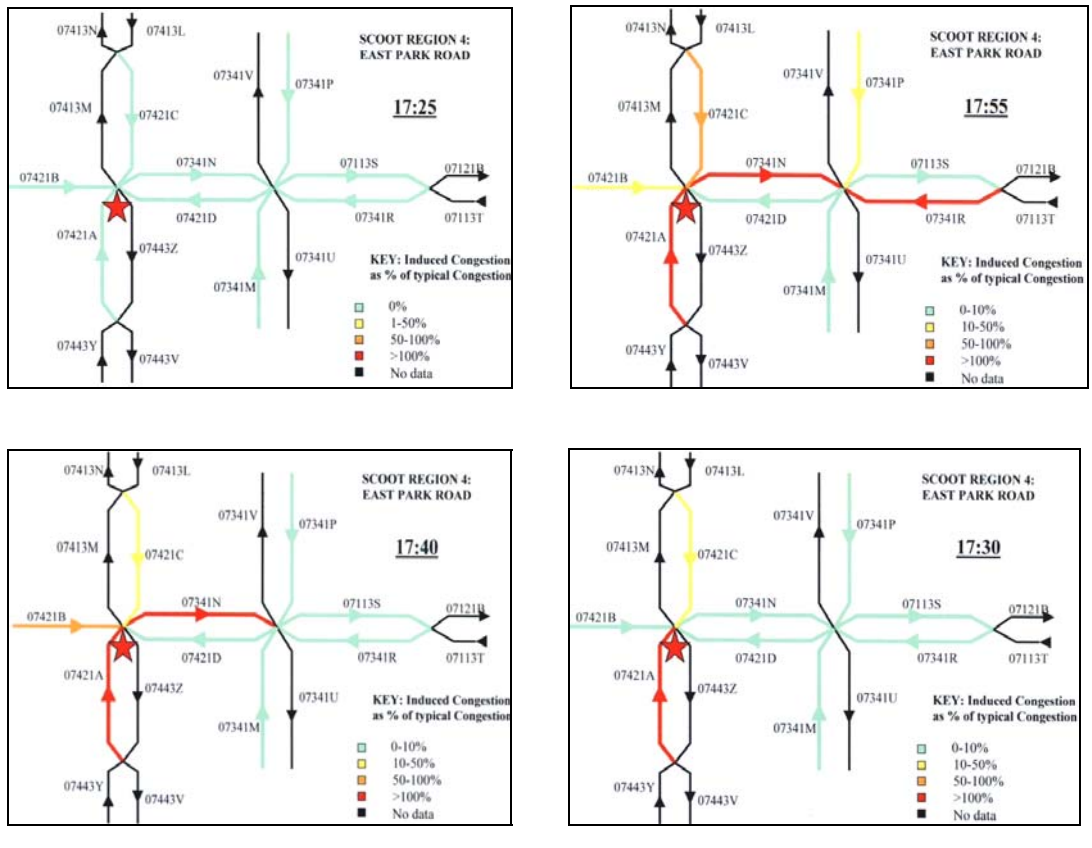


Figure 3. Congestion levels and their development compared to typical after an accident.

Table 1. Aggregate analysis of the congestion impacts of fatal accidents.

ACCIDENT	ACCIDENT DATE	ACCIDENT TIME	NETWORK	CONGESTED INDUCED[1]	AS %	DELAY INDUCED[2]	AS %
1	16/06/2003	PEAK	DIFFUSE	538.9	127	64	13
2	24/01/1999	NIGHT	DIFFUSE	0	0	3	16
3	29/05/1999	NIGHT	DIFFUSE	0	0	–0.58	–6
4	14/10/2003	OFF-PEAK	ONE-WAY	4826.1	4603	2532	327
5	19/05/2001	NIGHT	DIFFUSE	28	N/A	20.4	144
6	15/5/2004	NIGHT	DIFFUSE	9	N/A	14.4	71

1. Congestion expressed as additional number of 4-second congestion SCOOT intervals induced
2. Delay expressed as additional 1/10 vehicle hours per hour induced

Table 2. Typology and formal definitions of network setting.

Network type	Formal definition	Score
DIFFUSE	Many diversion opportunities, high network density, and easy route changes for drivers (e.g. 'Grid' street pattern).	1
LIMITED	Some opportunities for diversion, but only in one direction, or spaced apart, e.g. isolated/ peripheral links.	2
ONE-WAY/ DUAL CARRIAGEWAY	No opposing carriageway that can be used to bypass accident. Limited number of intersections, spaced apart.	3
MOTORWAY	As above, but with formal intersections only, widely spaced.	4

Aggregate Analysis

Although previously specified data and database issues meant that the original aspiration of developing a multiple regression model of congestion could not be conducted, some interpretation of the importance of each factor may be gained by more straightforward correlation and aggregation analysis. Accident time, day, network settings, and several other factors gained from the STATS19 data are used in this analysis. Time and network setting have been 'dummy' coded into '3', '2', and '1' for peak, off-peak and night respectively, with network setting defined according to the typology developed in Table 2.

The resulting correlation analysis is shown in Table 3. This shows that congestion and delay are significantly and positively inter-correlated, whilst network type also features highly. Indeed it is the only other factor with statistical significance. Accident time is the next most important factor to feature for induced congestion levels, although with the number of observations at six, there is not sufficient evidence for statistical significance.

Table 3. Correlation analysis on accident induced congestion levels.

	DELAY	CONG
DELAY	1	0.996(**)
CONG	0.996(**)	1
NUMCAS	-0.323	-0.355
ACCTIME	-0.313	-0.396
ACCDAY	-0.269	-0.312
RDTYPE	-0.300	-0.228
SURCON	-0.208	-0.228
NETWORK	-0.810	-0.838(*)

**Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Nevertheless, as these factors were also theoretically considered the most important in the formulation of the introductory hypothesis, further investigation of a number of other fatal accidents with available data sources is required in order to verify this, and in particular, to establish their comparative and relative importance of temporal and spatial setting in terms of inducing non-recurring, accident-induced congestion. This is especially the case since it is suspected from accident 4 that time of occurrence and network settings potentially work in mutually reinforcing ways.

Given the importance of network and time settings, a mathematical function is developed that works towards the expression of induced congestion levels as a function of these two variables. An 'index' of time and network setting is developed, and regressed on the results of the previous section.

This index comprises an aggregate of the coding scores applied to both time and network setting. Time periods are defined as '1' Night, '2' Off-peak and '3' as Peak, with network setting as previously defined in Table 2. In essence therefore, the aggregate score represents the 'recoverability' of the system to respond to the impact of an incident. Low scores (e.g. 2 may be caused by a night-time accident in a diffuse setting), high scores (e.g. 6 by a peak-time accident in a one-way setting).

Given the results in Table 1, the data seems to suggest that the main factors that lie behind the induced congestion can be approximately fitted to a line of the form:

$$\text{Ln}(Y) = b0 + (b1 * s)$$

Where: Ln(Y) = Ln(induced congestion)

b0 and *b1* are parameters to be estimated

s = the 'score' on the index of time + network

This produces the 'least-squares' regression line estimated for congestion is:

$$\text{Ln}(Y) = -6.384 + (3.041 * s)$$

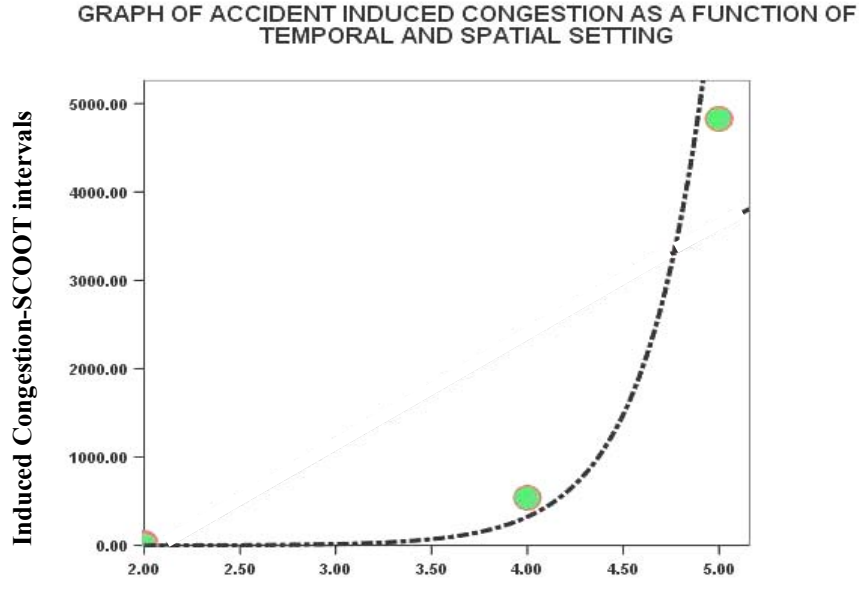


Figure. 4. Induced Congestion Levels as a Function of Time and Space 'Score'.

For delay, the approximation can be quantified through:

$$Ln(Y) = -4.059 + (2.266*s)$$

This is represented graphically in Figure 4, which closely fits an exponential function at an approximate spatial and temporal index score of '4'.

Valuing the Accident Induced Congestion Externality

Given time delays and congestion are assumed to equate to a loss of utility (and therefore money), and as the aggregate delays induced have already been calculated in 1/10 vehicle hours, it is a relatively straightforward step to monetise the externality created by accident induced congestion. Whilst the field of values of travel-time savings literature (VTTS) is numerous and complex (see, for example, Jara-Diaz, Mackie and Fowkes, 2001; Mackie, Fowkes, Wardman and Whelan 2003; Wardman, 1998, and 2000), several meta-analyses are available which show that whilst VTTS varies by almost every variable imaginable, some simple aggregate assumptions are possible

This research utilises standard values of time, as currently utilised by the DfT through it WebTAG guidance as a starting point (www.dft.gov.uk/webtag/), with the caveat that exact, retrospective analysis of journey purposes, fleet composition, occupancy levels, and associated variables can never be fully documented for the spatial and temporal settings of any of the accidents analysed. However, WebTAG, 3.5.6 does contain some useful assumptions in terms of how it would be expect that proportion of different journey purposes, occupancies and other key variables change over the course of a day, and these have been utilised in the calculations that follow.

As a result, the following valuations can be obtained for accident 4. The net delay increase is 2532 (1/10 vehicle hours), assuming the average LPUCF utilised by SCOOT in this calculation (17) is robust across the area concerned. This therefore equates to 253 actual *vehicle* hours, which requires conversion into persons affected. NTS data reproduced in WebTAG shows an average occupancy of 1.68 persons per vehicle between 10am and 4pm, which equates to 425 person hours. Using WebTAG purposed based values of time, and temporally specific assumptions regarding journey purpose splits, the induced congestion externality can be approximately valued at £2,954 when based to DfT 2003 prices.

Whilst perhaps limited in terms of monetary value, it is equally true that other accidents analysed at or below an 'index score' of 4 have very limited monetary disbenefits, and indeed are not worth valuation. Indeed, score '5' results, as per accident 4 above are somewhat marginal in the context of cost-benefit ratios for safety schemes.

However, the model can also be used estimate the significance of other types of accident. The most 'serious' impacts would occur with an accident both in the peak hour, and in a 'one-way' type setting, i.e. score '6'. Using the above model, it can be shown that the total congestion induced would be approximately 13,900 (1/10 vehicle hours), which, using peak hour NTS statistics, would provide an overall valuation of the incident in the region of £8,740. Although not substantial compared to overall values of prevention, some justification for the inclusion of these effects can be cited. Given government safety targets (e.g. DfT, 2008) and the automatic desire for increased road safety, greater 'values of prevention' for accidents increase the cost-effectiveness of road safety schemes. With higher valuations, safety schemes perform better in CBA analysis, and thus promulgate further road safety. Few would disagree this objective is not worthwhile.

Perhaps most importantly, the above valuations for the congestion externality induced are likely to be significantly under-estimated. This is because SCOOT is only able to record delays on roads in which it is installed. For the more 'severe' accidents, this can be under half the number of the actual diversionary routes used. Moreover, SCOOT regions do not always extend beyond the region affected by the accident, and it is recommended that other tools or more aggregate and traditional highway modelling techniques could be used to assist in the estimation of such effects. However, for a field that is typically conservative in the use of valuations for previously non-appraised phenomena, this represents an important point.

In terms of induced congestion valuations however, perhaps the 'dividing-line' would be better drawn between urban and motorway locations- rather than between different urban accident settings. Given Highways Agency estimates that peak-hour accidents on motorways can lead to queue length build-ups of up to one mile a minute, they would appear to warrant a separate category in terms of network setting compared to any of the accidents analysed here.

If for instance, a 'score' of 7 were to result from a peak accident in a 'new' categorisation of '4' for motorways, it can be shown that the valuation of the congestion induced is in the region £84,000. It is obviously recommended that such estimates are the study of a separate investigation, but nevertheless, findings here would give a clear rationale for such work, as well as its potential for future inclusion in the appraisal of road safety.

Conclusions

A variety of parameters and tests have been devised to assess causality, and to allow the summation of differences between 'typical' and 'accident-day' profiles across networks, and over many consecutive SCOOT time periods. It has been found that accidents do induce congestion and delay, but the amount varies substantially by accident. It is shown that the time of occurrence and the network layout produce the greatest variation in induced congestion levels. Night-time accidents typically produce negligible congestion, whilst delay induced is typically small- less than 2 vehicle hours. This increases to over 1390 vehicle hours for accidents that occur at peak times in constricted networks.

From each individual accident reconstruction, mathematical models for the levels of induced congestion and delay have been presented. These are subsequently used to assess the valuation of the delays caused by accidents for potential appraisal purposes. Only peak hour accidents, and/ or those that occur in constricted networks are found to have the potential for valuation purposes.

In comparison to previous research mainly located on trunk roads, network effects and route stochastity play a larger role in urban settings. Both of these may be thought of in terms of introducing additional 'slack' into the system, improving recoverability. However, the actual pattern of response is found to be more complex for each individual accident reconstruction. For instance, traffic may well be able to re-route in urban areas, but this is often onto roads further down the road hierarchy, and through lower capacity junctions less suited to short-duration, high-magnitude flows.

Further Requirements and Refinements for High Spatial and Temporal Resolution Modelling Using ITS Data

It is important to note that the results obtained above are inherently a function of the quantity of data available. Factors relating to data availability, SCOOT architecture, validity and the requirements of high resolution modelling all diminish the quality of the overall results- especially for large congestion events. Indeed, there is also a requirement for further investigation into the relative importance of the time and space in inducing accident congestion. The results here express a slight tendency to perceive space as the more important variable, but there is no reason *per se* that this relationship holds in other data-sets or urban areas.

Of particular importance are the problems posed by the use of non-SCOOT links as diversion routes, and the lack of estimation of the general nuisance factor associated with accident events. Both of these are likely to result in an under-estimation of the induced congestion and delay effects of fatal accidents, or other non-recurring traffic phenomena, when using SCOOT and associated ITS data at high temporal and spatial resolutions. Whilst it has been shown that such detailed modelling can be successfully achieved through post-hoc analysis- albeit for a relatively smaller number of non-recurrent congestion events given present data considerations within the SCOOT system, the challenge is to provide this level of spatial analysis in real time, *linked into* incident duration algorithms currently undergoing further development and validation.

As a result, it is emphasised that whilst the results allow a clear and potentially fruitful insight into the potential for ITS data in high spatial and temporal resolution traffic modelling, they must be placed in context, and must form the basis for further validation work and methodological refinement. The main deficiency of ITS data which this research has attempted to overcome is the current inability of current ITS systems to automatically detect the spatial and temporal dimensions of the very traffic phenomena such systems are designed to produce data in response to, and it is only once this process is complete, that the full capabilities of ITS data and its uses in high spatio-temporal resolution modelling can be realised, for a whole host of non-recurring traffic phenomena and consequential impacts.

Acknowledgements

The authors gratefully acknowledge EPSRC/DfT Research Grant EP/E002013/1 via MESSAGE (Mobile Environmental Sensing System Across Grid Environments) for support of data collection and analysis. The first-named author would like to acknowledge his financial support from the EPSRC Doctoral Training Grant.

References

Caltrans. (2007). “Development of a Tool for Measuring Non-recurrent Congestion Impacts of Accidents on Freeways”, *ATMS TESTBED PHASE III FINAL REPORT*.

Evans, R.G. (1995). “The identification and control of recurrent urban congestion using SCOOT data”, *PhD Thesis*. Department for Civil Engineering, University of Nottingham.

Goodman, P.S. (2001). “The prediction of road traffic noise in urban areas”, *PhD Thesis*, ITS, University of Leeds.

Highways Agency. (2004) “Incident Related Congestion”, *Phase 1 Project Report UPR/T/026/06* (available at http://www.ha-research.gov.uk/index.php/id=1011).

Hunt, P.B.; Robertson, D.I.; Bretherton, R.D. and Winton, R.I. (1981), "SCOOT - A traffic responsive method of co-ordinating signals", *Transport and Road Research Laboratory*, LR 1014, Crowthorne.

Mackie, P.J, Jara-Diaz, S., and Fowkes, A.S (2001). "The value of travel time savings in evaluation", Transportation Research E, 37 (2-3), 91-106.

Mackie, P.J, Fowkes, A.S, Wardman, M., and Whelan, G. (2003) "Values of Travel Time Savings in the UK- Summary and Conclusions".

Perrin, J., Martin, P.T., and Coleman, B.(2002). "Monitoring commuter congestion on surface streets in real time", *Transp. Res. Rec.* (Transportation Research Board,) 1811, 107-114.

Stathopoulos, A., and Karlaftis, M.G. (2001)."Temporal and spatial variations of real-time traffic data in urban areas", *Transp. Res. Rec. (Transportation Research Board),* 1768, 135-140

Wang W., Chen H. and Bell M.C. (2005), "A Review of Traffic Incident Duration Analysis", *Journal of Transportation systems Engineering and Information Technology*, under review.

Wardman, M (1998). "The value of travel time: a review of British evidence", *Journal of Transport.*

Economics and Policy 32 (3) 285-316.

Wardman, M. (2000). "A Review of British evidence on the valuations of time and service quality".

Working Paper 525, *ITS*, Leeds.

WebTAG (DfT, 2008). Unit 3.5.6. Available at (www.dft.gov.uk/webtag/*)*.

Wood, K and Baker, R.T.(1991). "The use of SCOOT detectors to measure traffic flows", *TRL Report*.

Radar Velocity-measuring System Design and Computation Algorithm Based on ARM Processor

Yi Zeng[1], Jianmin Xu[2] and Deng Peng[3]

[1] Doctoral Candidate, [2] Prof., School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510641, China; Email: zengyi518@163.com, aujmxu@scut.edu.cn
[3] Master, School of Civil Engineering and Transportation, South China University of Technology, Guangzhou 510641, China, Email: pengdengyjl@yahoo.cn

Abstract

The real-time monitoring of vehicles' velocity has become essential for traffic safety. Therefore, in order to improve the method of monitoring the velocity of vehicles on road, this paper described a new radar velocity-measuring system, which is based on ARM processor. This system is based on the Doppler principle and spectral estimation algorithm in its hardware design part, and it takes full use of the underlying resources of ARM processor to achieve sample, digital signal processing and peripheral control in the separate ARM core. In the software design part, it analyzes the main program flow, at the same time, since the ARM processor does not support hardware floating-point operations, it puts forward the usage of fixed-point numbers to store and calculate floating-point numbers, which makes the data loading and storage minimization. It uses ARM assembly language to write the core algorithm and takes full advantage of the 5-stage pipeline of ARM processor to achieve the right register allocation and instruction scheduling for fine-grained control, and also to avoid the pipeline interlocks. Experimental results indicate that the accuracy of this velocity measurement system can reach ±1km/h. It not only can be used in traffic enforcement agencies, but is also low cost, low power consumption advantages.

Key Words: Velocity-measuring; ARM; radar; Doppler effect; spectral estimation

Introduction

The rapid development of the transportation industry promotes the economic and social prosperity, yet the modern intelligent traffic management and the traffic safety has been pushed to an unprecedented height. This extensive research topic has become an important world issue, and how to measure the velocity of vehicles is an important part of it. Speeding has been the main cause of traffic accidents, referred to "the first killer," so the real-time monitoring of vehicles' velocity has become essential for traffic management and traffic safety.

Literature Review

At the present, there are many ways of velocity measurement, including the coil method (Hellinga 2002), image processing method (Lin et al. 2008), laser Doppler velocimetry (Liu et al. 2004), radar velocity-measuring system etc (Kleinhempel et al. 1992). The last two methods have been widely used due to their portability and high accuracy. Despite the laser Doppler velocimetry's high accuracy and effective long-distance advantages, it also has shortcomings. It is only in the stationary state to use, easily found by drivers, expensive, and so on. The radar velocity-measuring system, although its effective distance is less than the laser velocimetry, also can meet the requirements of high accuracy and can be used in stationary and motion state, moreover it has become popular due to the mature technology and affordable price. In the early 1970s, western countries began to use radar velocity-measuring systems, but these products are analog, large, and have a low accuracy. In the 1990s, a new radar velocity-measuring system appeared, which contained an integrated circuits digital signal processor, and worked in the Ka-band, compared with the analog one, it has advantages of high accuracy and small size.

With respect to the implementation technology, several generations of radar velocity-measuring systems have appeared with the development of the microcontrol and microprocessor chip. The first generation product uses a singlechip to implement. This singlechip needs to complete the control functions, and also needs to complete the core algorithm of a velocity-measuring system, but these algorithms are limited to a few simple, low-precision algorithms, such as counting cycles method and so on. The final product can achieve velocity measurement, but the accuracy is not enough, and the speed of systems operation is very slow. The second generation product uses singlechip plus DSP processor, which uses the traditional dual-core mode, that is, one singlechip is to deal with the control system and another separate DSP processor is to complete the part of digital signal processing. The accuracy and speed of these generation products are much higher than the first generation products. With the emergence and extensive application of ARM processors, its rich peripheral control module and the underlying hardware resources make it possible to achieve the second generation product, that is to say, just use of ARM single-core will be able to complete the traditional dual-core task, not only can greatly simplify the complex hardware design of the second generation product, but it will also have a low cost and low power consumption advantages. Therefore, the innovation of this paper lies

in using ARM processors to achieve the radar velocity-measuring system. It can be used in a traffic enforcement agency for measuring speeding violations, and it will be a new trend in the development of the radar velocity-measuring system.

The Composition and Related Principles of Radar Velocity-measuring System

The ARM-based radar velocity-measuring system mainly includes two parts: radar transceiver front end and radar signal processing system. As millimeter wave radar has advantages of narrow angle, high resolution, anti-interference ability, small size in light of the design of radar transceiver front end, this design uses Doppler millimeter-wave transceiver front end, which is developed by the InnoSenT company. Its operating frequency is 24 GHz, linear polarization, orthogonal field mixing, the output power is 50 mW, the antenna gain is 25 dB, and it can meet the performance indicators required for highway velocity measurement. It also takes into the power consumption. Firstly it will introduce the basic principles of radar velocity measurement and the signal processing principles in the radar signal processing system.

Doppler Principles of Radar Velocity Measurement. Measurement of target velocity by radar mainly uses the Doppler effect to carry out. The fixed-frequency electromagnetic waves, which are launched by the radar transceiver front end, will be bounced back when it encounters objects in the transmission process. If the encountered objects are stationary, then the frequency of bounced back waves will not change; If the encountered objects moved toward the direction of wave source, the waves bounced back at this time will be compressed, its frequency will increase; On the contrary, if the movement is away from the wave source, the received echo frequency will be lower than the launch frequency. Among them, the increase or decrease of frequency value is called the Doppler frequency, and is shown by the following equation:

$$f_d = \frac{2v_r}{C} f_0 \tag{1}$$

In the equation: f_d is Doppler frequency, v_r is the velocity of target vehicle, C is the speed of light, f_0 is the launch frequency of radar wave, here it is equal to 24 GHz. From Eq. 1 we can get:

$$v_r = \frac{f_d \cdot C}{2 f_0} \tag{2}$$

From Eq. 2 we can see other variables are known, as long as the Doppler frequency can be worked out, we can calculate the target vehicle velocity.

The Signal Processing Principles of the System. From the above Doppler principles we can see that through estimating the Doppler frequency of radar echo signal, we can complete the velocity estimation of target vehicle. Among them, the measurement error of frequency directly affects the measurement results of velocity.

Periodogram method is a kind of classical spectral estimation. It not only has the high accuracy of frequency estimation, but also can complete real-time signal processing. Therefore, we choose this method as the core algorithm of frequency measurement in the system. This method takes the N points observation data $x_N(n)$ of random signal $x(n)$ as a power limited signal, and directly takes the Fourier transform of $x_N(n)$, then it can get $x_N(e^{jw})$. Count the amplitude square of $x_N(e^{jw})$, which is divided by N, and make it as the estimation of real power spectrum. Use $\hat{P}_{PER}(e^{jw})$ to express the power spectrum which is estimated by Periodogram method, then we can get:

$$\hat{P}_{PER}(k) = \frac{1}{N}\left|X_N(k)\right|^2 \tag{3}$$

Since $x_N(k)$ can be calculated by FFT, $\hat{P}_{PER}(k)$ can be easily calculated too, so the frequency of the signal will be able to be identified through solving the signal's power spectrum.

The Design of Radar Signal Processing System

With the ARM architecture fortified, the radar signal processing system based on ARM is entirely feasible, and it has simple hardware design, low cost, and low power consumption advantages. Here we choose the S3C2410A Samsung ARM chip to be the main processor of the system.

The Hardware Design of Radar Signal Processing System. The system's hardware architecture includes three parts: pretreatment part of the radar signal, digital signal processing part, and peripherals control part of the system.

The Design Of Radar Signal Pretreatment Part. Radar transceiver front end will launch 24 GHz frequency electromagnetic waves, which are reflected back by the moving target, and through mixing, the radar transceiver front end will output a 100mv weak sinusoidal signal, the frequency of the signal namely is the Doppler frequency that we want to collect. However, the deferent radar signal has a variety of interference, including the low frequency and high frequency interference, which requires a band-pass filter to prevent the interference signal, pass the useful signal. And this deferent radar signal is very weak too, therefore, this paper designed an adjustable amplifier circuit, so that through the low-frequency amplified, an analog signal will be sent to the AD converter module of ARM for sampling. Since S3C2410A has a 10-bit ADC with 8-channel analog inputs and its analog input range is 0-3.3V, we need to add a 3 V limiter circuit to ensure that ARM processor does not get burned. The hardware block diagram of this part is shown in Figure 1.

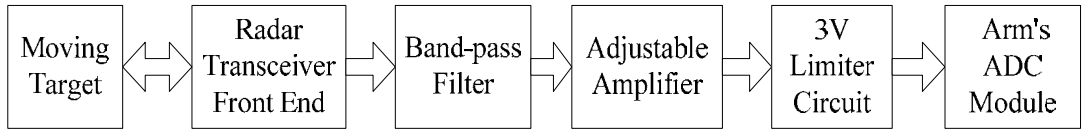


Figure 1. Hardware block diagram of the radar signal pretreatment part.

For the band-pass filter design, we use the general-purpose active filter MAX275 manufactured by the U.S. MAXIM company to achieve the fourth-order Butterworth active filter for radar signal. The transmission bands are 300Hz-16kHz. So by Eq. 2, immediately we can see the velocity measurement range is 5-250km/h, which is fully able to meet the traffic law enforcement's velocity measurement requirements. For the adjustable amplifier design, we choose the integrated dual-supply operational amplifier OP07, firstly use OP07 to constitute a voltage follower. It is mainly used to increase input impedance and decrease output impedance. Then the output of voltage follower connects another OP07 noninverting amplifier circuit, its input resistance is 2kΩ, the feedback resistance is 100kΩ adjustable resistor, so that magnification factor will be in the 0-34dB adjustable scope.

The Design of Digital Signal Processing and Peripherals Control Part. ARM processor is the core of these two parts' hardware design. First, through the radar signal pretreatment, the analog signal will be put into the ARM's ADC module for sampling, converted into a digital signal, and then use ARM to carry out the digital signal processing. It ultimately uses ARM control module to communicate with the PC, and to set the speeding threshold value in the keyboard. Finally, it displays the real-time velocity and sets the speeding violation alarm by a buzzer and other functions. The hardware block diagram of these two parts is shown in Figure 2. The digital signal processing part fully uses ARM's underlying hardware resources and is achieved by software design. In addition, S3C2410A only has a 4kB internal RAM, therefore, the usual way is to expand the dynamic RAM, as well as the NandFlash and NorFlash outside the S3C2410A. Here it is designed to: expand a 64MB SDRAM, a 64MB NandFlash, and a 2MB NorFlash. This memory expansion design not only meets the storage needs of radar system, but also provides storage resources for embedding the system to achieve a more powerful velocity measurement and additional features.

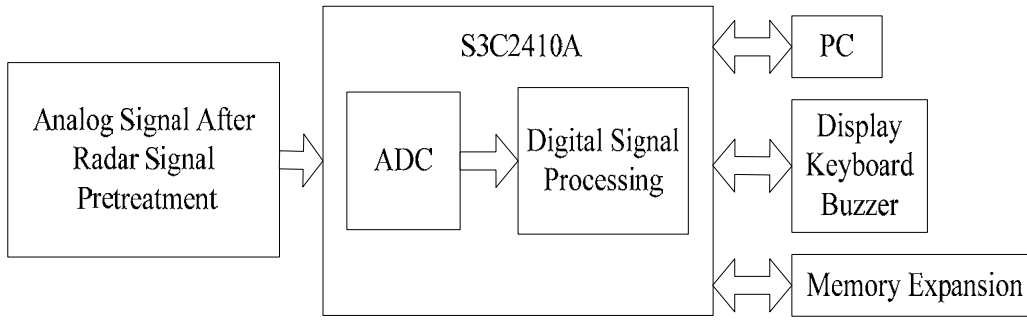


Figure 2. Hardware block diagram of these two parts.

The Software Design of Radar Signal Processing System. The main task of the radar signal processing system's software design is to carry out sampling, operation pre-treatment, spectrum operation, output, and display of the operation results for the Doppler radar signal, which passed from the radar transceiver front end. The main flowchart of its software design is shown in Figure 3. The software operation of the system is divided into two stages, the first is system initialization, including stack initialization, heap space allocation, interrupt vector table configuration, register initialization, external memory initialization, and so on. The second stage is the main program, first of all is to initialize the ARM's ADC module and the I/O port that needs to be used in the system, then starts sampling for the radar signal and enables an interruption, and stores the sampling data to the specified data buffer. If the sampling points are enough, the FFT transform will be started, and the spectrum of the sampling signal is calculated. Further analysis and follow-up treatment for the signal spectrum finally obtains the velocity. Lastly it is to realize the control function, namely real-time output and display of the velocity data, speeding alarm, and so on.

FFT algorithm needs to handle a lot of floating-point operations, furthermore it has high requirements in the multiply-accumulate and data access aspect, while the ARM processor does not support hardware floating-point operations, and there is no single instruction to achieve multiply-accumulate and parallel data access.

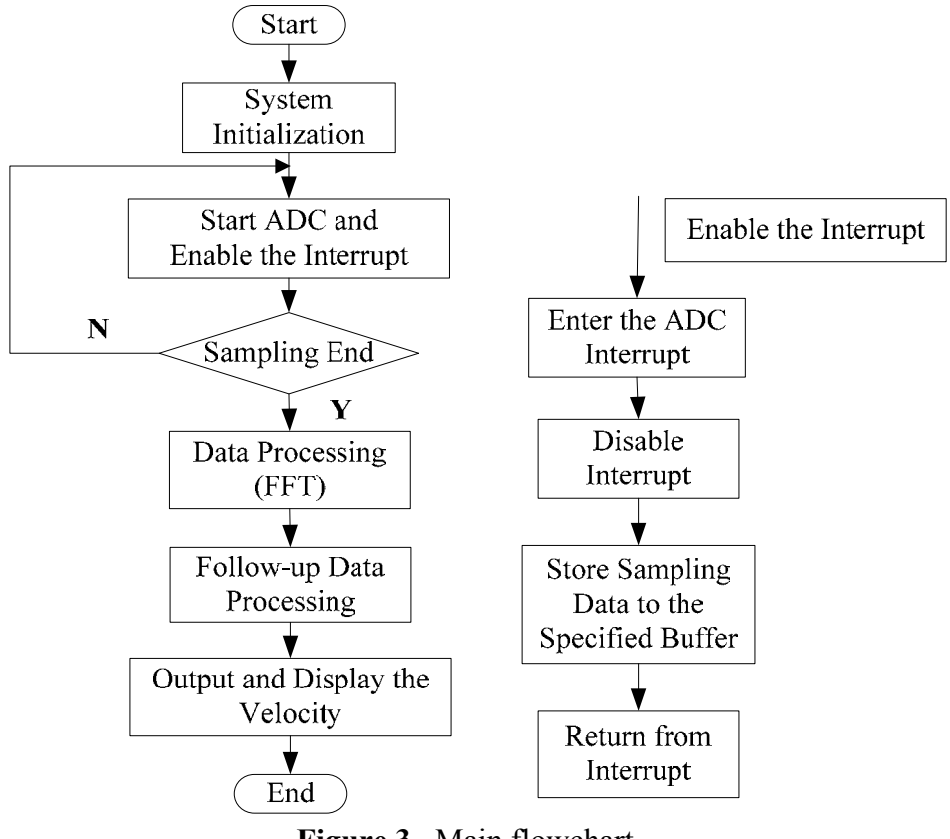


Figure 3. Main flowchart.

It is necessary to compare FFT's various implementation ways and select a good method. In the FFT's various implementation ways, we first compare the software implementation between the base radix-2 DIT and radix-4 DIT FFT algorithm. Suppose sampling points N = 4M (M is a natural number), using radix-2 algorithm needs to multiply MN times, while using Radix-4 algorithm requires for 3MN/4 times multiplications, apparently can save 25% of the multiplication, for the same reason using the radix-8 algorithm can ultimately save some multiplication steps, nevertheless, only 14 registers of S3C2410A can be used, as a result the saving cycles have been consumed by extra load and storage. So radix-4 DIT FFT algorithm is the best choice for this core algorithm, and its algorithm process is introduced as follows.

If the length of sequence $x(n)$ is $N = 4^n$, it will use the following equation to calculate its DFT (Discrete Fourier Transform):

$$x(k) = \sum_{n=0}^{N-1} x(n) W_N^{nk}, \quad k, n = 0,1,2,\cdots,N-1 \tag{4}$$

Radix-4 DIT FFT will break down $x(n)$ in the time domain:

$$x_m(r) = x(4r+m), \quad 0 \le m \le 3, \quad 0 \le r \le \frac{N}{4} - 1 \tag{5}$$

In the equation: subsequence $x_m(r)$ all are $N/4$ points sequence. Suppose the DFT of $x_m(r)$ is $X_m(r)$, then:

$$\begin{cases} X(r) = X_0(r) + W_N^r X_1(r) + W_N^{2r} X_2(r) + W_N^{3r} X_3(r) \\ X(r+\frac{N}{4}) = X_0(r) + W_N^{r+\frac{N}{4}} X_1(r) + W_N^{2r+\frac{N}{4}} X_2(r) + W_N^{3r+\frac{N}{4}} X_3(r) \\ X(r+\frac{N}{2}) = X_0(r) + W_N^{r+\frac{N}{2}} X_1(r) + W_N^{2r+\frac{N}{2}} X_2(r) + W_N^{3r+\frac{N}{2}} X_3(r) \\ X(r+\frac{3N}{4}) = X_0(r) + W_N^{r+\frac{3N}{4}} X_1(r) + W_N^{2r+\frac{3N}{4}} X_2(r) + W_N^{3r+\frac{3N}{4}} X_3(r) \end{cases} \tag{6}$$

In the equation: $0 \le r \le N/4 - 1$. From Eq. 6 we can get the radix-4 butterfly unit, because the $r = 4^{M-1}$ in $X_0(r)$、$X_1(r)$、$X_2(r)$、$X_3(r)$ can continue to break down $M-1$ times, that is, the whole radix-4 operation totally includes M levels. The above part amply analyzed FFT algorithms and program structure, but there are still many problems about how to efficiently realize the algorithm on the ARM processor. The following section will introduce some program design features, and they are the key factors for the system's implementation on the ARM processor as well.

Data Loading and Storage. ARM processor does not support floating-point, so this program uses the fixed-point data to express the floating-point data. This method is based on the requirements of precision in the calculation, which will make the floating-point data shift left, and then use a rounding algorithm to reserve the integer data part. In this way the floating-point operations in ARM instructions can be used

as part of the four arithmetic operation instructions. It does not take up the additional instruction cycle and can greatly improve the efficiency of the code. Another data storage issue is to divide each complex number into real part and imaginary part to store in a single word, in order to take full advantage of 14 ARM's 32-bit registers.

The Use of Twiddle Factor. In order to improve the computing speed of FFT, calculate the twiddle factor $W_N^r = \exp(-j\frac{2\pi}{N}r)$, which contains a large number of complex operations beforehand, then uses 14-bit left shift integer data format and divides into real and imaginary parts to store it. Therefore, in the case of the sampling points $N = 4^M$, the twiddle factor can be easily obtained using the table look-up approach, not only makes the value of the program's sampling points N optional, moreover it can speed up the computation of the code, and get a higher measurement of accuracy.

Through these innovative designs and integrations, the radix-4 DIT FFT algorithm flowchart, which finally implemented on ARM processor, will be shown in Figure 4.

Experiment and Test Result

In order to test the accuracy of the core algorithm, we have produced a sinusoidal signal generator with a five-range tunable frequency, and used this signal as a good pretreatment radar signal input to the ARM processor. Then through the serial port, the Doppler frequency, which is calculated by the program, will be sent to the PC and displayed on the screen. At the same time, send the calculation value of each of the sampling points to the PC in order to use these sample values to simulate the corresponding Doppler frequency in MATLAB.

First we use the sinusoidal signal generator to generate a sinusoidal signal, its frequency is 9.2 kHz. Then use it to carry out the simulation, and get Figure 5. The first diagram of Figure 5 is the spectrum diagram, which is used as an output sample value of the ARM processor and obtained from simulation in the MATLAB. The second diagram is the enlarged image of the spectrum maximum value. From the diagram we can see through this core algorithm, the measured signal frequency is 9182.25 Hz. Compared to the true frequency 9.2kHz, the frequency measurement accuracy of this algorithm is very high, which can reach more than 99%. The third diagram shows the final velocity measurement device of this paper. Finally we use this device to carry out the road test, and the last diagram of Figure 5 is the site where the road experiment was carried out. Its test results are shown in Table 1 (The measured frequency and velocity in Table 1 is the average of 10 times measurements, so we can see its velocity measurement accuracy can reach ±1km/h.). The velocity measurement range of this device is 5-250km/h, and from the test we know this device can detect all type of vehicles, that is to say it can measure the real-time velocity of trucks, cars, and even bicycles. In addition, the effective measurement distance for the trucks can be up to 1.5 km, for the cars can be up to 1 km.

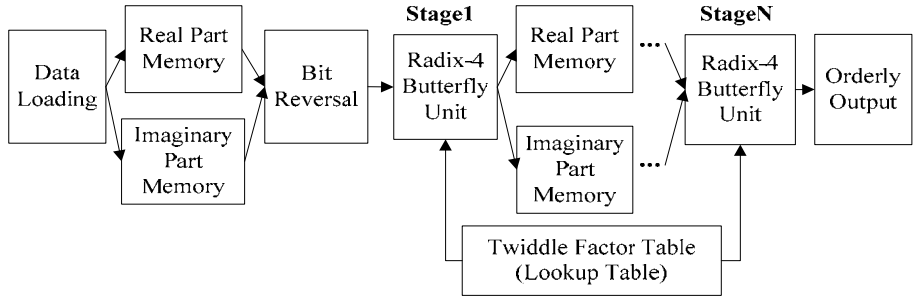


Figure 4. FFT algorithm flowchart in the ARM.

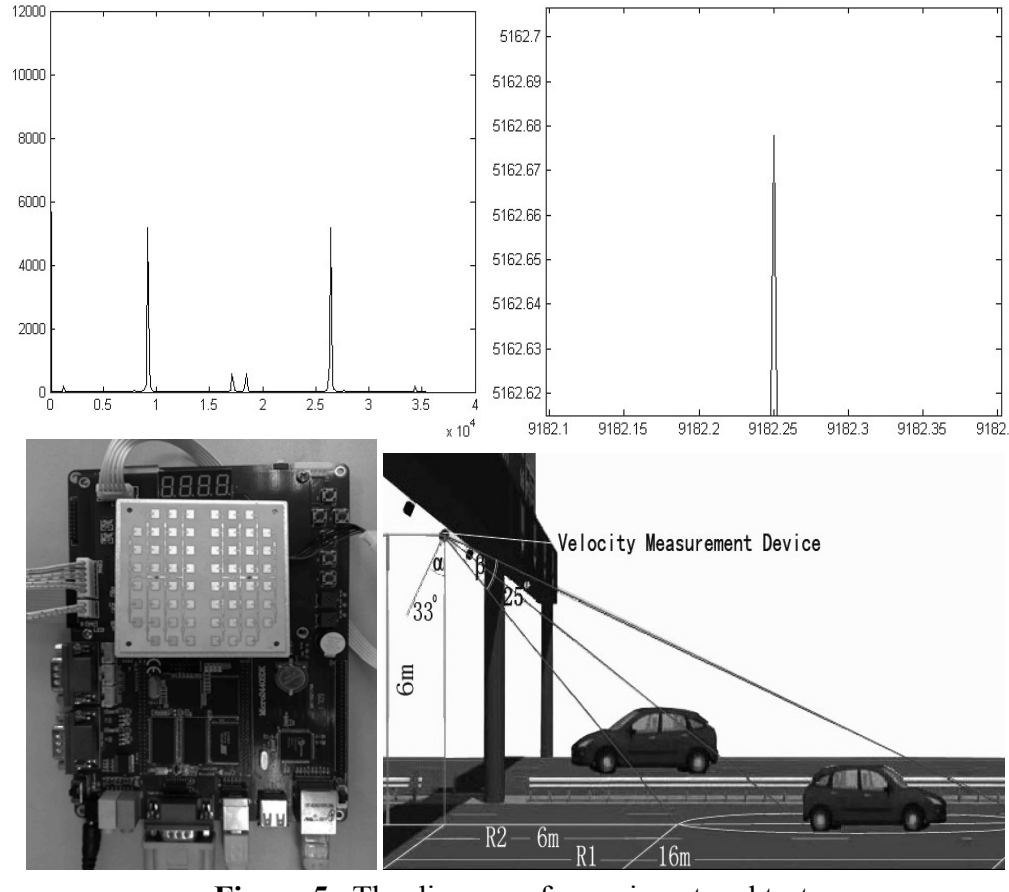


Figure 5. The diagram of experiment and test.

Table 1. Road test data.

Real velocity (km/h)	Real Frequency (kHz)	Measuring Frequency (kHz)	Measuring Velocity (km/h)	Measurement Error (km/h)
30	1.917	1.900	29.739	0.261
60	3.833	3.810	59.635	0.365
120	7.667	7.631	119.442	0.558
150	9.583	9.622	150.605	-0.605
180	11.500	11.445	179.139	0.861

Conclusions

This paper describes the radar velocity-measuring system design and algorithm research based on ARM. There are solid theoretical foundations, as well as a number of innovation points. The characteristics of this design are reflected by using a single-core ARM processor to achieve the dual-core task, which is traditionally completed by DSP plus a singlechip, and it has laid the foundation for ARM's DSP application. The test results show that the accuracy of this system can fully meet the requirements of traffic law enforcement agencies and have the advantages of a small size, low cost, low power consumption, etc. Therefore, it entirely can be applied to traffic management departments in order to enhance the traffic safety.

References

Sloss, A. N., Symes, D., and Wright, C. (2005). "ARM system developer's guide: designing and optimizing system software." *American, Elsevier PteLtd,* 1-306.

Helliga, B. R. (2002). "Improving freeway speed estimates from single-loop detectors." *Journal of Transportation Engineering,* 58-67.

Lin, H., Li, K., and Chang, C.. (2008). "Vehicle speed detection from a single motion blurred image." *Image and Vision Computing,* 1327-1337.

Liu, C., Lu. Y., Liu, J., and Lu., F.. (2004). "Spectrum-analyzing laser velocimetric system." *Opto-Electronic Engineering,* 52-55.

Kleinhempel, W., Bergmann, D., and Stammler, W. (1992). "Speed measure of vehicles with on-board Doppler radar." *IEE Conference Publication Oct 12-13,* 284-287.

Chang, Y., and Parhi, K. K.. (2003). "An efficient pipelined FFT architecture." *IEEE Transactions on Circuits and Systems II,* 322-325.

Tanno, K, Takeda, T, and Horiguchi, S. (1992). "Parallel radix 4 FFT algorithms on an eight-neighbor processor array." *Proc IEEE TENCON'92[C],* 855-859.

Li, Y. F., Chen, K. F.. (2008). "Eliminating the picket fence effect of the fast Fourier transform." *Computer Physics Communications,* 486-491.

Critical Speed Prediction at Unsignalized Intersection under Mixed Traffic Condition

Joewono Prasetijo[1], Molla Mohammad Ali[2] and Ning Wu [3]

[1] Senior Lecturer, [2]Associate Professor, School of Civil Engineering, Universiti Sains Malaysia, Engineering Campus, 14300 Nibong Tebal, Seberang Perai Selatan, P. Pinang, Malaysia, PH: 04-5996210; Email: Joewono.Prasetijo@ruhr-uni-bochum.de cejoewono@eng.usm.my,
[3] Professor, Lehrstuehle fuer Verkehrswesen, Ruhr-Universitaet Bochum, Universitaetsstrasse 150, D-44801 Bochum, Germany, PH: +49-234-32 26557; Fax: +49-234-32 14151,

Abstract

The new method of capacity analysis at unsignalized intersections was developed in the authors' study for mixed traffic condition in developing countries (e.q. Indonesia) where the drivers' behavior, traffic composition, level of roadside activities, etc. were different from that of developed countries. The new method called "conflict technique "is based on the interactions between average speed and flow of conflict streams (six streams and six conflict points) considered. All possible conflict streams were considered simultaneously and the interactions were taken into account through empirical regression models. Six conflict groups (with six subject streams) were considered for all intersections to obtain the relationship between speed and flow of conflict streams. This phenomenon would measure a critical speed of traffic streams at the intersection when the capacity is reached. This method would measure any relationship between speed and flow of each stream of traffic and speed and flow within the group of conflict. The Method could provide a new technique to determine a critical speed of unsignalized intersection under mixed traffic flow where the capacity/maximum flow is reached.
Key Words: speed; flow; mixed traffic; conflict technique; conflict streams

Background

A new method of capacity analysis at unsignalized intersections was developed in the authors 'previous study for mixed traffic condition in developing countries (e.q.

Indonesia) where the drivers' behaviour, traffic composition, level of roadside activities, etc. were different from that of developed countries. The new method called "conflict technique" is based on the interactions between average speed and flow of conflict streams (six streams and six conflict points) considered. All possible conflict streams were considered simultaneously and the interactions were taken into account through empirical regression models. Six conflict groups (with six subject streams) were considered for all intersections to obtain the relationship between speed and flow of conflict streams. Observation was conducted at ten unsignalized intersection in Indonesia. The study found that a good relationship between speed and flow of all streams. A suitable correlation between the speed of light vehicles/cars (LV) and flow of each stream of each type of vehicle of conflict group has been found at almost all of the intersections observed. Considering the level of correlation between parameters, this phenomenon would measure a critical speed/maximum possible speed of traffic streams at the intersection when the capacity is reached.

Conflict Technique Approach

The analysis is based on interactions between streams which include parameters of speed and flow. Due to these circumstances, each of the parameters of each stream was analyzed related included other streams. The model illustration consists of six (6) streams (C–A, C–B, B–C, B–A, A–C, A–B) and six (6) conflict points (1, 2, 3, 4, 5, 6). Furthermore, the model has six (6) groups of conflicts (I, II, III, IV, V, VI) which include all streams' conflicts and each group has its own subject stream. Each of them should be considered further in order to find their maximum flow. The scheme of conflict and group of conflict can be seen at Figure 1 and Table 1.

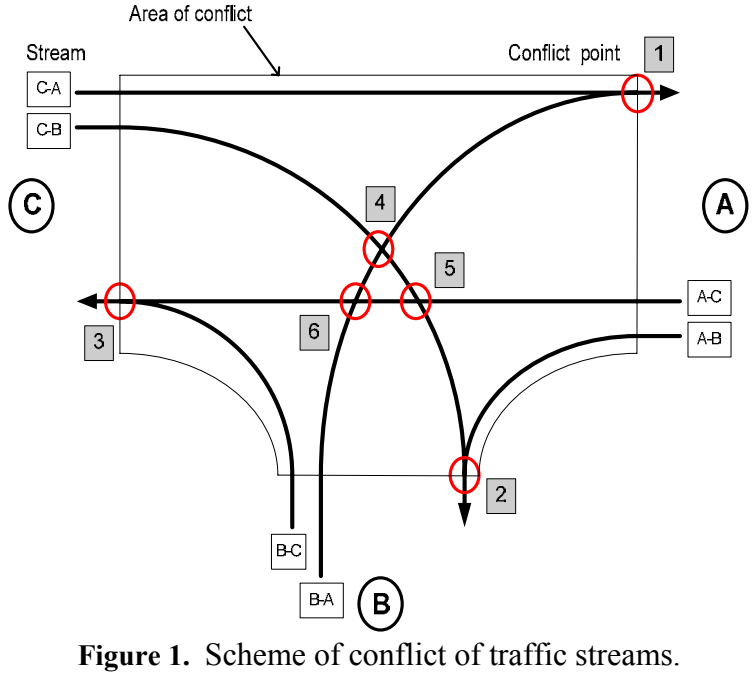


Figure 1. Scheme of conflict of traffic streams.

Table 1. Group of conflict of six streams.

Group of Conflict	Subject Stream	Conflict Point	Streams Involved
I	C – A	1	C – A, B – A
II	C – B	2,4,5	C – B, B – A, A – C, A – B
III	B – C	3	B – C, A – C
IV	B – A	1,4,6	B – A, A – C, C – B, C – A
V	A – C	3,5,6	A – C, C – B, B – A, B – C
VI	A – B	2	A – B, C – B

This method found a suitable relationship between speed and flow of each stream of traffic and speed and flow within the group of conflict. Therefore, the general form of this relationship was described as $V=a\text{-}bQ_{\text{streams involved}}$ where parameter *a* representing free-flow speed [km/h], *b* is speed reduction coefficient and *Q* is flow [pcu/h]. Speeds at each conflict point *i* (V_i, i=1,2,3,4,5,6) should be observed and measured when streams have reached their maximum flow ($Q_{j\text{-}MAX}$, with j=1,2,3,4,5,6). If one of the stream *j* has reached its maximum flow $Q_{j\text{-}MAX}$, it should be assumed that the intersection is at initial stage of congestion with speed V_i' while the other flows maintain their flow rates (maximum that has been reached).

Linear models for speed, for example subject stream C–A and flow which is defined as the total number of vehicles of streams included in conflict groups observed (e.g. C–A and B–A). Due to a large number of different parameter with regard to various types of vehicles in this study, e.g. speed and flow, therefore, the study has constructed relationships between parameters including each vehicle's performances (LT, MHV, LV, MC, UM) from each stream as $V_{\text{LV-CA}}$ (conflict group-I), $V_{\text{LV-CB}}$ (conflict group-II), $V_{\text{LV-BC}}$ (conflict group-III), $V_{\text{LV-BA}}$ (conflict group-IV), $V_{\text{LV-AC}}$ (conflict group-V), $V_{\text{LV-AB}}$ (conflict group-VI) and follow $V_{\text{LV-stream subject}}=A-\sum K_{\text{type of veh.-streams involved}}*Q_{\text{type of veh.-streams involved}}$ where *A* is constant representing free-flow speed and *K* is speed reduction coefficient.

A suitable correlation between the speed of light vehicles/cars (LV) of each stream and flow of each type of vehicle of conflict group was found at almost all of the intersections. The relationship between the flow of each stream ($Q_{C\text{-}A}$, $Q_{C\text{-}B}$, $Q_{B\text{-}C}$, $Q_{B\text{-}A}$, $Q_{A\text{-}C}$, $Q_{A\text{-}B}$) of each conflict group (I, II, III, IV, V, VI) and the speed of each stream ($V_{C\text{-}A}$, $V_{C\text{-}B}$, $V_{B\text{-}C}$, $V_{B\text{-}A}$, $V_{A\text{-}C}$, $V_{A\text{-}B}$) was also be developed. The capacity calculations were based on each stream performance of every conflict group. Therefore, it is important to have an average speed and the total flow of each stream of all intersections.

The study considered that the subject streams, e.g. b_{C-A} Q_{C-A} have also contributed to the reduction of its own speed, V_{C-A} instead of other conflict streams (e.g. b_{B-A} Q_{B-A} in group of conflict–I). The study was also found that group of conflict–II, group of conflict–IV and group of conflict–V have four streams included (C–B, B–A and A–C) while other groups (group conflict–III and VI) have only two streams (C–

A, B–C and A–B). The relationships between speed and flow of conflict streams could be described as $V_I = a_I\text{-}b_I Q_B\text{-}c_I Q_C$ and $V_{II} = a_{II}\text{-}b_{II} Q_A\text{-}c_{II} Q_C$ and so on.

Due to a large number of different parameter values with regard to various types of vehicles in this study, e.g. speed and flow, this study has constructed relationships between parameters including each vehicle's performances (LT, MHV, LV, MC, UM) from each stream as follows :
conflict group – I,

$$V_{LV\text{-}CA} = A - K_{LV\text{-}CA} \cdot Q_{LV\text{-}CA} - K_{LV-BA} \cdot Q_{LV\text{-}BA} - K_{MHVi\text{-}CA} \cdot Q_{MHVi\text{-}CA} - K_{MHVi-BA} \cdot Q_{MHVi-BA} - .... - K_{UMi\text{-}CA} \cdot Q_{UMi\text{-}CA} - K_{UMi-BA} \cdot Q_{UMi-BA} \quad (1)$$

conflict group – II,

$$V_{LV\text{-}CB} = A - K_{LV\text{-}CB} \cdot Q_{LV\text{-}CB} - K_{LV-BA} \cdot Q_{LV\text{-}BA} - K_{LV-AC} \cdot Q_{LV\text{-}AC} - K_{LV-AB} \cdot Q_{LV\text{-}AB} - K_{MHVi\text{-}CB} \cdot Q_{MHVi\text{-}CB} - K_{MHVi-BA} \cdot Q_{MHVi-BA} - K_{MHVi-AC} \cdot Q_{MHVi-AC} - K_{MHVi-AB} \cdot Q_{MHVi-AB} - .... - K_{UMi\text{-}CB} \cdot Q_{UMi\text{-}CB} - K_{UMi-BA} \cdot Q_{UMi-BA} - K_{UMi-AC} \cdot Q_{UMi-AC} - K_{UMi-AB} \cdot Q_{UMi-AB} \quad (2)$$

conflict group – III,

$$V_{LV\text{-}BC} = A - K_{LV\text{-}BC} \cdot Q_{LV\text{-}BC} - K_{LV-AC} \cdot Q_{LV\text{-}AC} - K_{MHVi\text{-}BC} \cdot Q_{MHVi\text{-}BC} - K_{MHVi-AC} \cdot Q_{MHVi-AC} - .... - K_{UMi\text{-}BC} \cdot Q_{UMi\text{-}BC} - K_{UMi-AC} \cdot Q_{UMi-AC} \quad (3)$$

conflict group – IV,

$$V_{LV\text{-}BA} = A - K_{LV\text{-}BA} \cdot Q_{LV\text{-}BA} - K_{LV-AC} \cdot Q_{LV\text{-}AC} - K_{LV-CB} \cdot Q_{LV\text{-}CB} - K_{LV-CA} \cdot Q_{LV\text{-}CA} - K_{MHVi\text{-}BA} \cdot Q_{MHVi\text{-}BA} - K_{MHVi-AC} \cdot Q_{MHVi-AC} - K_{MHVi-CB} \cdot Q_{MHVi-CB} - K_{MHVi-CA} \cdot Q_{MHVi-CA} - .... - K_{UMi\text{-}BA} \cdot Q_{UMi\text{-}BA} - K_{UMi-AC} \cdot Q_{UMi-AC} - K_{UMi-CB} \cdot Q_{UMi-CB} - K_{UMi-CA} \cdot Q_{UMi-CA} \quad (4)$$

conflict group – V,

$$V_{LV\text{-}AC} = A - K_{LV\text{-}AC} \cdot Q_{LV\text{-}AC} - K_{LV-CB} \cdot Q_{LV\text{-}CB} - K_{LV-BA} \cdot Q_{LV\text{-}BA} - K_{LV-BC} \cdot Q_{LV\text{-}BC} - K_{MHVi\text{-}AC} \cdot Q_{MHVi\text{-}AC} - K_{MHVi-CB} \cdot Q_{MHVi-CB} - K_{MHVi-BA} \cdot Q_{MHVi-BA} - K_{MHVi-BC} \cdot Q_{MHVi-BC} - .... - K_{UMi\text{-}AC} \cdot Q_{UMi\text{-}AC} - K_{UMi-CB} \cdot Q_{UMi-CB} - K_{UMi-BA} \cdot Q_{UMi-BA} - K_{UMi-BC} \cdot Q_{UMi-BC} \quad (5)$$

conflict group – VI,

$$V_{LV\text{-}AB} = A - K_{LV\text{-}AB} \cdot Q_{LV\text{-}AB} - K_{LV-CB} \cdot Q_{LV\text{-}CB} - K_{MHVi\text{-}AB} \cdot Q_{MHVi\text{-}AB} - K_{MHVi-CB} \cdot Q_{MHVi-CB} - .... - K_{UMi\text{-}AB} \cdot Q_{UMi\text{-}AB} - K_{UMi-CB} \cdot Q_{UMi-CB} \quad (6)$$

where $V_{LV\text{-}CA}$ is speed of light vehicle (LV) stream C–A [km/h], A as a constant representing free–flow speed of light vehicle [km/h] and $K_{LV\text{-}CA}$ is speed reduction effect caused by light vehicle (LV) stream C–A [-], $Q_{LV\text{-}CA}$ is traffic flow for light vehicle (LV) stream C–A [pcu/h], $K_{MHV\text{-}CA}$ is speed reduction effect caused by medium heavy vehicle (MHV) stream C–A [-] and $Q_{MHV\text{-}CA}$ is traffic flow for medium heavy vehicle (MHV) stream C–A [pcu/h].

Capacity Analysis Based on Speed and Flow Relationship

Analysis has been made in this study based on observed data at three–leg unsignalized intersections. This type of intersections contains less conflict streams

compared to four–leg unsignalized intersections. This new empirically based method relies on the average speed of subject streams and the volume of each stream to determine the capacity as the maximum possible volume at the intersection. The study has considered speeds of each traffic stream which are influenced by other streams of each group of conflict. Thus, we can also take into account the following equations:

If streams' flow $Q_{C\text{-}A}$ reaches its maximum flow at

$$V_{C-A}{}' = a_{C-A} - b_{C-A} Q_{C-A}{}^{(1)} - b_{B-A}\left[\frac{Q_{C-A}{}^{(1)}}{f_1}\right] \tag{7}$$

the maximum flow of all streams are

$$Q_{C-A}{}^{(1)} = \left\{\left[\frac{a_{C-A} - V_{C-A}{}'}{\left(b_{C-A} + \frac{b_{B-A}}{f_1}\right)}\right], 0\right\}_{MAX}$$

$$Q_{B-A}{}^{(1)} = \left\{\left[\frac{a_{C-A} - V_{C-A}{}' - b_{C-A} Q_{C-A}{}^{(1)}}{b_{B-A}}\right], 0\right\}_{MAX} = \left\{\left(\frac{Q_{C-A}{}^{(1)}}{f_1}\right), 0\right\}_{MAX}$$

$$Q_{C-B}{}^{(1)} = \left\{\left[\frac{a_{B-A} - V_{B-A} - b_{B-A} Q_{B-A}{}^{(1)} - b_{C-A} Q_{C-A}{}^{(1)}}{\left(b_{C-B} + b_{A-C} f_5\right)}\right], 0\right\}_{MAX} = \left\{\left(f_4 \cdot Q_{B-A}{}^{(1)}\right), 0\right\}_{MAX}$$

$$Q_{A-C}{}^{(1)} = \left\{\left[\frac{a_{B-A} - V_{B-A} - b_{B-A} Q_{B-A}{}^{(1)} - b_{C-A} Q_{C-A}{}^{(1)} - b_{C-B} Q_{C-B}{}^{(1)}}{b_{A-C}}\right], 0\right\}_{MAX} = \left\{\left(f_6 \cdot Q_{B-A}{}^{(1)}\right), 0\right\}_{MAX}$$

$$Q_{A-B}{}^{(1)} = \left\{\left[\frac{a_{A-B} - V_{A-B} - b_{C-B} Q_{C-B}{}^{(1)}}{b_{A-B}}\right], 0\right\}_{MAX} = \left\{\left(\frac{Q_{C-B}{}^{(1)}}{f_2}\right), 0\right\}_{MAX}$$

$$Q_{B-C}{}^{(1)} = \left\{\left[\frac{a_{B-C} - V_{B-C} - b_{A-C} Q_{A-C}{}^{(1)}}{b_{B-C}}\right], 0\right\}_{MAX} = \left\{\left(f_3 \cdot Q_{A-C}{}^{(1)}\right), 0\right\}_{MAX}$$

and the maximum flow of the intersection is

$$Q_{\text{int}-MAXIMUM}{}^{(1)} = Q_{C-A}{}^{(1)} + Q_{C-B}{}^{(1)} + Q_{B-C}{}^{(1)} + Q_{B-A}{}^{(1)} + Q_{A-C}{}^{(1)} + Q_{A-B}{}^{(1)} \tag{8}$$

If streams' flow $Q_{C\text{-}B}$ reaches its maximum flow at

$$V_{C-B}{}' = a_{C-B} - b_{C-B} Q_{C-B}{}^{(2)} - b_{B-A}\left[\frac{Q_{C-B}{}^{(2)}}{f_4}\right] - b_{A-C}\left[f_5 \cdot Q_{C-B}{}^{(2)}\right] - b_{A-B}\left[\frac{Q_{C-B}{}^{(2)}}{f_2}\right] \tag{9}$$

The maximum flows of all streams are

$$Q_{C-B}{}^{(2)} = \left\{\left[\frac{a_{C-B} - V_{C-B}{}'}{\left(b_{C-B} + \frac{b_{B-A}}{f_4} + \left(b_{A-C} \cdot f_3\right) + \frac{b_{A-B}}{f_2}\right)}\right], 0\right\}_{MAX}$$

$$Q_{A-B}^{(2)} = \left\{ \left[\frac{a_{A-B} - V_{A-B} - b_{C-B} Q_{C-B}^{(2)}}{b_{A-B}} \right], 0 \right\}_{MAX} = \left\{ \left(\frac{Q_{C-B}^{(2)}}{f_2} \right), 0 \right\}_{MAX}$$

$$Q_{A-C}^{(2)} = \left\{ \left[\frac{a_{C-B} - V_{C-B}' - a_{C-B} Q_{C-B}^{(2)} - a_{A-B} Q_{A-B}^{(2)}}{\left(b_{A-C} + \frac{b_{B-A}}{f_6} \right)} \right], 0 \right\}_{MAX} = \left\{ \left(f_5 \cdot Q_{C-B}^{(2)} \right), 0 \right\}_{MAX}$$

$$Q_{B-A}^{(2)} = \left\{ \left[\frac{a_{C-B} - V_{C-B}' - a_{C-B} Q_{C-B}^{(2)} - a_{A-B} Q_{A-B}^{(2)} - a_{A-C} Q_{A-C}^{(2)}}{b_{B-A}} \right], 0 \right\}_{MAX} = \left\{ \left(\frac{Q_{C-B}^{(2)}}{f_4} \right), 0 \right\}_{MAX}$$

$$Q_{B-C}^{(2)} = \left\{ \left[\frac{a_{B-C} - V_{B-C} - b_{A-C} Q_{A-C}^{(2)}}{b_{B-C}} \right], 0 \right\}_{MAX} = \left\{ \left(f_3 \cdot Q_{A-C}^{(2)} \right), 0 \right\}_{MAX}$$

$$Q_{C-A}^{(2)} = \left\{ \left[\frac{a_{C-A} - V_{C-A} - b_{B-A} Q_{B-A}^{(2)}}{b_{C-A}} \right], 0 \right\}_{MAX} = \left\{ \left(f_1 \cdot Q_{B-A}^{(2)} \right), 0 \right\}_{MAX}$$

And the maximum flow of the intersection is

$$Q_{\text{int}-MAXIMUM}^{(2)} = Q_{C-A}^{(2)} + Q_{C-B}^{(2)} + Q_{B-C}^{(2)} + Q_{B-A}^{(2)} + Q_{A-C}^{(2)} + Q_{A-B}^{(2)} \quad (10)$$

If streams' flow $Q_{B\text{-}C}$ reaches its maximum flow at

$$V_{B-C}' = a_{B-C} - b_{B-C} Q_{B-C}^{(4)} - b_{A-C} \left[\frac{Q_{B-C}^{(4)}}{f_3} \right] \quad (11)$$

The maximum flow of all streams,

$$Q_{B-C}^{(3)} = \left\{ \left[\frac{a_{B-C} - V_{B-C}'}{\left(b_{B-C} + \frac{b_{A-C}}{f_3} \right)} \right], 0 \right\}_{MAX}$$

$$Q_{B-C}^{(3)} = \left\{ \left[\frac{a_{B-C} - V_{B-C}'}{\left(b_{B-C} + \frac{b_{A-C}}{f_3} \right)} \right], 0 \right\}_{MAX}$$

$$Q_{A-C}^{(3)} = \left\{ \left[\frac{a_{B-C} - V_{B-C}' - b_{B-C} Q_{B-C}^{(3)}}{b_{A-C}} \right], 0 \right\}_{MAX} = \left\{ \left(\frac{Q_{B-C}^{(3)}}{f_3} \right), 0 \right\}_{MAX}$$

$$Q_{C-B}^{(3)} = \left\{ \left[\frac{a_{A-C} - V_{A-C} - b_{A-C} Q_{A-C}^{(3)} - b_{B-C} Q_{B-C}^{(3)}}{\left(b_{C-B} + \frac{b_{B-A}}{f_4} \right)} \right], 0 \right\}_{MAX} = \left\{ \left(\frac{Q_{A-C}^{(3)}}{f_5} \right), 0 \right\}_{MAX}$$

$$Q_{A-B}^{(3)} = \left\{ \left[\frac{a_{A-B} - V_{A-B} - b_{C-B} Q_{C-B}^{(3)}}{b_{A-B}} \right], 0 \right\}_{MAX} = \left\{ \left(\frac{Q_{C-B}^{(3)}}{f_2} \right), 0 \right\}_{MAX}$$

$$Q_{B-A}^{(3)} = \left\{ \left[\frac{a_{A-C} - V_{A-C} - b_{A-C} Q_{A-C}^{(3)} - b_{C-B} Q_{C-B}^{(3)} - b_{B-C} Q_{B-C}^{(3)}}{b_{B-A}} \right], 0 \right\}_{MAX} = \left\{ \left(\frac{Q_{A-C}^{(3)}}{f_6} \right), 0 \right\}_{MAX}$$

$$Q_{C-A}^{(3)} = \left\{ \left[\frac{a_{C-A} - V_{C-A} - b_{B-A} Q_{B-A}^{(3)}}{b_{C-A}} \right], 0 \right\}_{MAX} = \left\{ \left(f_1 \cdot Q_{B-A}^{(3)} \right), 0 \right\}_{MAX}$$

And the maximum flow of the intersection is

$$Q_{\text{int}-MAXIMUM}^{(3)} = Q_{C-A}^{(3)} + Q_{C-B}^{(3)} + Q_{B-C}^{(3)} + Q_{B-A}^{(3)} + Q_{A-C}^{(3)} + Q_{A-B}^{(3)} \tag{12}$$

If streams' flow $Q_{B\text{-}A}$ reaches its maximum flow at

$$V_{B-A}' = a_{B-A} - b_{B-A} Q_{B-A}^{(4)} - b_{C-A} \left(f_1 \cdot Q_{B-A}^{(4)} \right) - b_{C-B} \left(f_4 \cdot Q_{B-A}^{(4)} \right) - b_{A-C} \left(f_6 \cdot Q_{B-A}^{(4)} \right) \tag{13}$$

the maximum flows of all streams are

$$Q_{B-A}^{(4)} = \left\{ \left[\frac{a_{B-A} - V_{B-A}'}{b_{B-A} + \left(b_{C-A} \cdot f_1 \right) + \left(b_{C-B} \cdot f_4 \right) + \left(b_{A-C} \cdot f_6 \right)} \right], 0 \right\}_{MAX}$$

$$Q_{C-A}^{(4)} = \left\{ \left[\frac{a_{C-A} - V_{C-A} - b_{B-A} Q_{B-A}^{(4)}}{b_{C-A}} \right], 0 \right\}_{MAX} = \left\{ \left(f_1 \cdot Q_{B-A}^{(4)} \right), 0 \right\}_{MAX}$$

$$Q_{C-B}^{(4)} = \left\{ \left[\frac{a_{B-A} - V_{B-A}' - b_{B-A} Q_{B-A}^{(4)} - b_{C-A} Q_{C-A}^{(4)}}{b_{C-B} + \left(b_{A-C} \cdot f_5 \right)} \right], 0 \right\}_{MAX} = \left\{ \left(f_4 \cdot Q_{B-A}^{(4)} \right), 0 \right\}_{MAX}$$

$$Q_{A-C}^{(4)} = \left\{ \left[\frac{a_{B-A} - V_{B-A}' - b_{B-A} Q_{B-A}^{(4)} - b_{C-A} Q_{C-A}^{(4)} - b_{C-B} Q_{C-B}^{(4)}}{b_{A-C}} \right], 0 \right\}_{MAX} = \left\{ \left(f_6 \cdot Q_{B-A}^{(4)} \right), 0 \right\}_{MAX}$$

$$Q_{B-C}^{(4)} = \left\{ \left[\frac{a_{A-C} - V_{A-C} - b_{A-C} Q_{A-C}^{(4)} - b_{C-B} Q_{C-B}^{(4)} - b_{B-A} Q_{B-A}^{(4)}}{b_{B-C}} \right], 0 \right\}_{MAX} = \left\{ \left(f_3 \cdot Q_{A-C}^{(4)} \right), 0 \right\}_{MAX}$$

$$Q_{A-B}^{(4)} = \left\{ \left[\frac{a_{A-B} - V_{A-B} - b_{C-B} Q_{C-B}^{(4)}}{b_{A-B}} \right], 0 \right\}_{MAX} = \left\{ \left(\frac{Q_{C-B}^{(4)}}{f_2} \right), 0 \right\}_{MAX}$$

And the maximum flow of the intersection is

$$Q_{\text{int}-MAXIMUM}^{(4)} = Q_{C-A}^{(4)} + Q_{C-B}^{(4)} + Q_{B-C}^{(4)} + Q_{B-A}^{(4)} + Q_{A-C}^{(4)} + Q_{A-B}^{(4)} \tag{14}$$

If streams' flow $Q_{A\text{-}C}$ reaches its maximum flow at

$$\begin{aligned} V_{A-C}' &= a_{A-C} - b_{A-C} Q_{A-C}^{(5)} - b_{C-B} Q_{C-B} - b_{B-A} Q_{B-A} - b_{B-C} Q_{B-C} \\ &= a_{A-C} - b_{A-C} Q_{A-C}^{(5)} - b_{C-B} \left(\frac{Q_{A-C}^{(5)}}{f_5} \right) - b_{B-A} \left(\frac{Q_{A-C}^{(5)}}{f_6} \right) - b_{B-C} \left(f_3 \cdot Q_{A-C}^{(5)} \right) \end{aligned} \tag{15}$$

the maximum flow of all streams are

$$Q_{A-C}^{(5)} = \left\{ \left[\frac{a_{A-C} - V_{A-C}^{'}}{b_{A-C} + \frac{b_{C-B}}{f_5} + \frac{b_{B-A}}{f_6} + \left(b_{B-C} \cdot f_3\right)} \right], 0 \right\}_{MAX}$$

$$Q_{B-C}^{(5)} = \left\{ \left[\frac{a_{B-C} - V_{B-C} - b_{A-C} Q_{A-C}^{(5)}}{b_{B-C}} \right], 0 \right\}_{MAX} = \left\{ \left(f_3 \cdot Q_{A-C}^{(5)}\right), 0 \right\}_{MAX}$$

$$Q_{C-B}^{(5)} = \left\{ \left[\frac{a_{A-C} - V_{A-C}^{'} - b_{A-C} Q_{A-C}^{(5)} - b_{B-C} Q_{B-C}^{(5)}}{\left(b_{C-B} + \frac{b_{B-A}}{f_4}\right)} \right], 0 \right\}_{MAX} = \left\{ \left(\frac{Q_{A-C}^{(5)}}{f_5} \right), 0 \right\}_{MAX}$$

$$Q_{B-A}^{(5)} = \left\{ \left[\frac{a_{A-C} - V_{A-C}^{'} - b_{A-C} Q_{A-C}^{(5)} - b_{C-B} Q_{C-B}^{(5)} - b_{B-C} Q_{B-C}^{(5)}}{b_{B-A}} \right], 0 \right\}_{MAX} = \left\{ \left(\frac{Q_{A-C}^{(5)}}{f_6} \right), 0 \right\}_{MAX}$$

$$Q_{A-B}^{(5)} = \left\{ \left[\frac{a_{A-B} - V_{A-B} - b_{C-B} Q_{C-B}^{(5)}}{b_{A-B}} \right], 0 \right\}_{MAX} = \left\{ \left(\frac{Q_{C-B}^{(5)}}{f_2} \right), 0 \right\}_{MAX}$$

$$Q_{C-A}^{(5)} = \left\{ \left[\frac{a_{C-A} - V_{C-A} - b_{B-A} Q_{B-A}^{(5)}}{b_{C-A}} \right], 0 \right\}_{MAX} = \left\{ \left(f_1 \cdot Q_{B-A}^{(5)}\right), 0 \right\}_{MAX}$$

And the maximum flow of the intersection is

$$Q_{\text{int}-MAXIMUM}^{(5)} = Q_{C-A}^{(5)} + Q_{C-B}^{(5)} + Q_{B-C}^{(5)} + Q_{B-A}^{(5)} + Q_{A-C}^{(5)} + Q_{A-B}^{(5)} \quad (16)$$

If streams' flow $Q_{A\text{-}B}$ reaches its maximum flow

$$V_{A-B}^{'} = a_{A-B} - b_{A-B} Q_{A-B}^{(6)} - b_{C-B}\left(f_2 \cdot Q_{A-B}^{(6)}\right) \quad (17)$$

the maximum flows of all streams are

$$Q_{A-B}^{(6)} = \left\{ \left[\frac{a_{A-B} - V_{A-B}^{'}}{b_{A-B} + \left(b_{C-B} \cdot f_2\right)} \right], 0 \right\}_{MAX}$$

$$Q_{C-B}^{(6)} = \left\{ \left[\frac{a_{A-B} - V_{A-B}^{'} - b_{A-B} Q_{A-B}^{(6)}}{b_{C-B}} \right], 0 \right\}_{MAX} = \left\{ \left(f_2 \cdot Q_{A-B}^{(6)}\right), 0 \right\}_{MAX}$$

$$Q_{A-C}^{(6)} = \left\{ \left[\frac{a_{C-B} - V_{C-B} - b_{C-B} Q_{C-B}^{(6)} - b_{A-B} Q_{A-B}^{(6)}}{\left(b_{A-C} + \frac{b_{B-A}}{f_6}\right)} \right], 0 \right\}_{MAX} = \left\{ \left(f_5 \cdot Q_{C-B}^{(6)}\right), 0 \right\}_{MAX}$$

$$Q_{B-A}^{(6)} = \left\{ \left[\frac{a_{C-B} - V_{C-B} - b_{C-B} Q_{C-B}^{(6)} - b_{A-B} Q_{A-B}^{(6)} - b_{A-C} Q_{A-C}^{(6)}}{b_{B-A}} \right], 0 \right\}_{MAX} = \left\{ \left(\frac{Q_{C-B}^{(6)}}{f_4} \right), 0 \right\}_{MAX}$$

$$Q_{C-A}^{(6)} = \left\{ \left[\frac{a_{B-A} - V_{B-A} - b_{B-A} Q_{B-A}^{(6)} - b_{C-B} Q_{C-B}^{(6)} - b_{A-C} Q_{A-C}^{(6)}}{b_{C-A}} \right], 0 \right\}_{MAX} = \left\{ \left(f_1 \cdot Q_{B-A}^{(6)}\right), 0 \right\}_{MAX}$$

$$Q_{B-C}^{(6)} = \left\{ \left[\frac{a_{B-C} - V_{B-C} - b_{A-C} Q_{A-C}^{(6)}}{b_{B-C}} \right], 0 \right\}_{MAX} = \left\{ \left(f_3 \cdot Q_{A-C}^{(6)}\right), 0 \right\}_{MAX}$$

And the maximum flow of the intersection,

$$Q_{\text{int}-MAXIMUM}^{(6)} = Q_{C-A}^{(6)} + Q_{C-B}^{(6)} + Q_{B-C}^{(6)} + Q_{B-A}^{(6)} + Q_{A-C}^{(6)} + Q_{A-B}^{(6)} \quad (18)$$

From equation 7 to equation 18, it can be concluded that the total maximum flow of the intersection is the least maximum flow from all possible maximum flows, $Q_{int-MAXIMUM}$

$$C_{\text{int}.} = \left(Q_{\text{int}-MAXIMUM}^{(1)}, Q_{\text{int}-MAXIMUM}^{(2)}, Q_{\text{int}-MAXIMUM}^{(3)}, Q_{\text{int}-MAXIMUM}^{(4)}, Q_{\text{int}-MAXIMUM}^{(5)}, Q_{\text{int}-MAXIMUM}^{(6)}\right)_{MIN} \quad (19)$$

Developed figure representing speed and flow relationship which a pre-defined speed V' is considered was performed. Each stream in a group of conflict would have the same average speed. If one conflict group reach their maximum flow/capacity, Q_j' other conflicts would produce a negative value, $Q_j'' \rightarrow Q<0$, this illustration is shown in Figure 2. In this case, it is required to have the maximum flow/capacity for all streams where flow is higher or equal than zero, $Q_j'' \geq 0$.

The maximum flow was found from the aggregated short–based 1–minute and 5–minutes of speed and flow data in the flat/normal conditions. Each sample in this database represents the average speed and flow value for all observed 1–minute and 5–minute intervals. The predetermined flows fall in the range of 438.4 pcu/h – 2209.4 pcu/h for all sites. Evidence shows that a relationship between speed and flow was found in every case of mixed traffic at a certain segment of road.

It is obvious that the correlation between speed and flow of streams could be used to develop such an approach of conflict streams analysis. The total maximum flow of intersection is assumed to be reached by one of the traffic streams while the traffic flow at an intersection is congested, but traffic movements are still possible. When one of the stream i reaches its maximum flow (capacity) and the subject stream reach its maximum flow j, $C_i^{\,j}$,

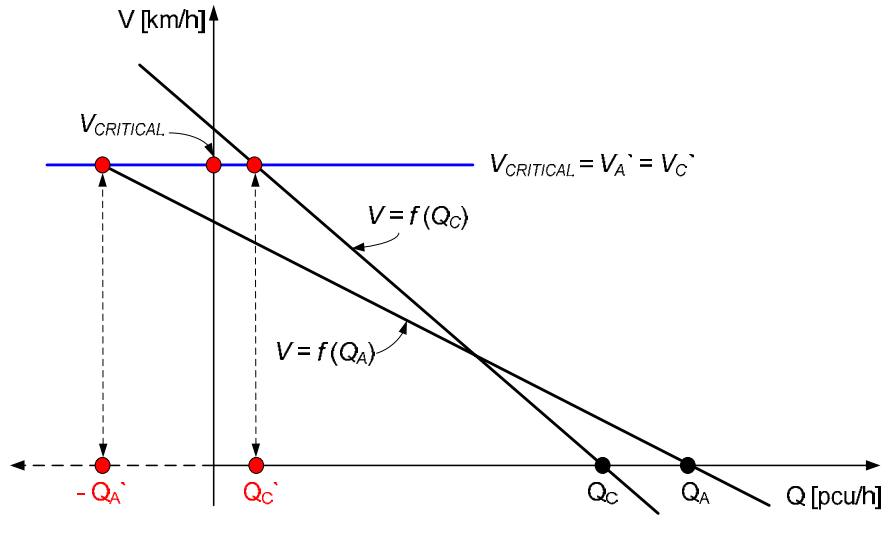


Figure 2. Critical speed for a group of conflict (stream A and stream C).

Table 2. Matrix of the maximum flow of intersection

Speed at maximum flows subject stream	Maximum flows subject stream	Maximum Flows Stream						Total maximum flow at intersection
		$Q_{C\text{-}A}$	$Q_{C\text{-}B}$	$Q_{B\text{-}C}$	$Q_{B\text{-}A}$	$Q_{A\text{-}C}$	$Q_{A\text{-}B}$	
$V_{C\text{-}A}'$	$Q_{C\text{-}A}^{(1)}$	$C_{C\text{-}A}$	$Q_{C\text{-}B}^{(1)}$	$Q_{B\text{-}C}^{(1)}$	$Q_{B\text{-}A}^{(1)}$	$Q_{A\text{-}C}^{(1)}$	$Q_{A\text{-}B}^{(1)}$	$\Sigma\, Q_i^{(1)}$
$V_{C\text{-}B}'$	$Q_{C\text{-}B}^{(2)}$	$Q_{C\text{-}A}^{(2)}$	$C_{C\text{-}B}$	$Q_{B\text{-}C}^{(2)}$	$Q_{B\text{-}A}^{(2)}$	$Q_{A\text{-}C}^{(2)}$	$Q_{A\text{-}B}^{(2)}$	$\Sigma\, Q_i^{(2)}$
$V_{B\text{-}C}'$	$Q_{B\text{-}C}^{(3)}$	$Q_{C\text{-}A}^{(3)}$	$Q_{C\text{-}B}^{(3)}$	$C_{B\text{-}C}$	$Q_{B\text{-}A}^{(3)}$	$Q_{A\text{-}C}^{(3)}$	$Q_{A\text{-}B}^{(3)}$	$\Sigma\, Q_i^{(3)}$
$V_{B\text{-}A}'$	$Q_{B\text{-}A}^{(4)}$	$Q_{C\text{-}A}^{(4)}$	$Q_{C\text{-}B}^{(4)}$	$Q_{B\text{-}C}^{(4)}$	$C_{B\text{-}A}$	$Q_{A\text{-}C}^{(4)}$	$Q_{A\text{-}B}^{(4)}$	$\Sigma\, Q_i^{(4)}$
$V_{A\text{-}C}'$	$Q_{A\text{-}C}^{(5)}$	$Q_{C\text{-}A}^{(5)}$	$Q_{C\text{-}B}^{(5)}$	$Q_{B\text{-}C}^{(5)}$	$Q_{B\text{-}A}^{(5)}$	$C_{A\text{-}C}$	$Q_{A\text{-}B}^{(5)}$	$\Sigma\, Q_i^{(5)}$
$V_{A\text{-}B}'$	$Q_{A\text{-}B}^{(6)}$	$Q_{C\text{-}A}^{(6)}$	$Q_{C\text{-}B}^{(6)}$	$Q_{B\text{-}C}^{(6)}$	$Q_{B\text{-}A}^{(6)}$	$Q_{A\text{-}C}^{(6)}$	$C_{A\text{-}B}$	$\Sigma\, Q_i^{(6)}$
		Maximum Flow/Capacity of Intersection=						$\Sigma\, Q_i^{(j)}{}_{MIN}$

the total capacity of the intersection is the sum of the maximum flow (C_i) of all streams with the subject stream of j, follows $C_{total}=\sum C_i^j$. When stream i reaches its maximum flow (capacity) j, the total capacity of the intersection is the sum of the maximum flow C_i of all streams is described as,

$$C_{total} = \sum_{j=1}^{6} C_i^j \tag{20}$$

where C_{total} is the total maximum stream volumes at the intersection [pcu/h], C_i^j is the maximum flow of stream i when subject stream j reaches its maximum flow [pcu/h] and i is movement of stream i. In order to complete the measurement above and to simplify the calculation of speed and flow of each stream, the following table of matrix of maximum flow is presented.

Results and Discussions

Each intersection would create their own equations of maximum flow with measurements from the field. However, it was not always possible to have a speed at the maximum streams' flow since field observation covered only two hours for each intersection. Therefore, the critical speed at the maximum flow was estimated for a further calculation where speeds varied from 10 km/h to 15 km/h by considering previous studies at the current intersections. An example of capacity analysis on intersections performed the capacity with speed values of 10 km/h, 13 km/h, 14 km/h, 15 km/h and the result shows the capacities are close to the results from the manual with speeds of 12 km/h and 13 km/h as can be seen in Table 3.

Table 3. Approximate speed to the maximum flow.

Type of analysis	Intersection/approximate speed to the maximum flow [km/h]									
	1	2	3	4	5	6	7	8	9	10
Data	22	26	10	11	16	10	17	6	13	11
Model	12.6	13.1	10.3	13.6	10.6	12.8	10.6	10.5	13.2	11.3

Table 4. Standard deviation of capacity for different speed.

Speed, V [km/h]	10	11	12	13	14	15
Capacity, C [pcu/h]	4619.53	3905.10	3208.23	2514.86	1860.50	1331.94
Standard deviation, σ [pcu/h]	1024.39	875.94	718.35	558.78	390.78	208.72

The model developed is applied with different speeds (v=10 km/h, v=11 km/h, v=12 km/h, v=13 km/h, v=14 km/h, v=15 km/h). The results show that the capacity based on the Indonesia Highway Capacity Manual is between results of the model with speeds' range 10 km/h–14 km/h and is close to the results from Indonesia Highway Capacity Manual (1997) at the average speed of 11.86 km/h. The study also shows that every intersection remains to have different speeds at a maximum flow close to the results from the manual. Table 4 shows that analysis from data performed very close to the results from the manual with suitable standard deviation. The reasons might be an over–estimation in speed. In such a case, it might be necessary to consider more accurate field observations and measurements conducted with a special apparatus/camera (speed and flow detector).

The developed model with different speeds is applied (v=10 km/h, v=11 km/h, v=12 km/h, v=13 km/h, v=14 km/h, v=15 km/h), see Table 4. A standard deviation of the resulting capacities for each speed is presented in Table 4. In a general, all intersections would reach their maximum flow with the speeds' range 10 km/h – 13 km/h at the maximum flow which is very close to the capacity resulted from the manual (Indonesian Highway Capacity Manual). This result is shown in Figure 3.

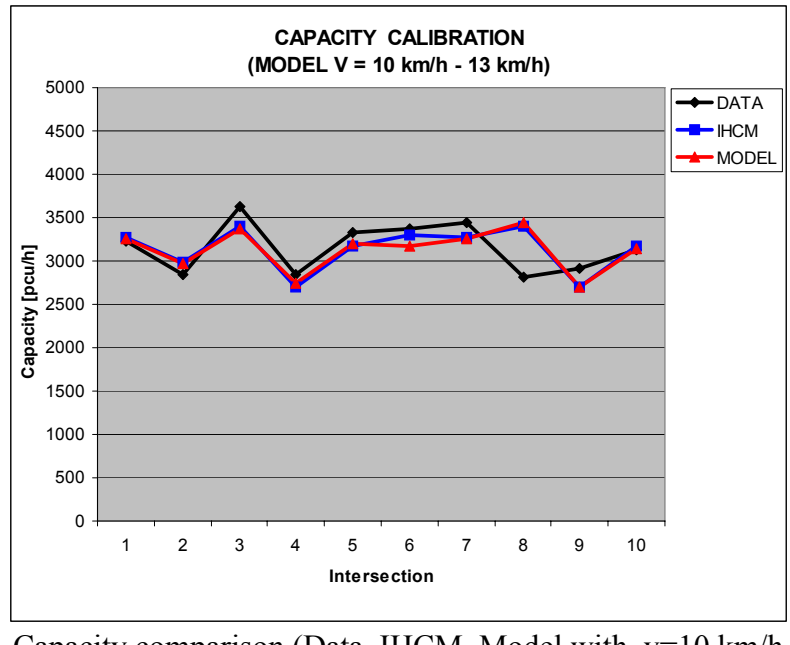


Figure 3. Capacity comparison (Data–IHCM–Model with v=10 km/h–13 km/h).

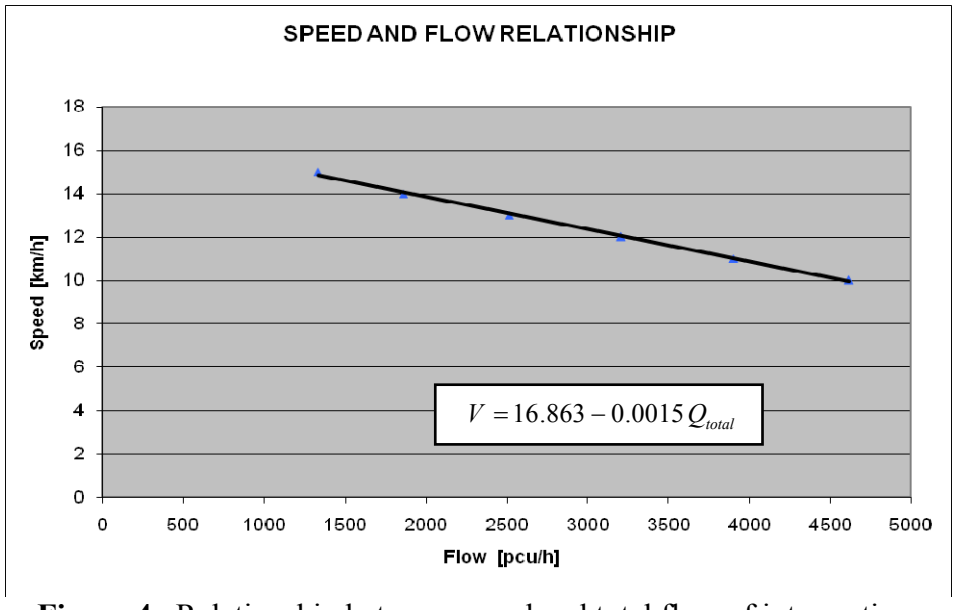


Figure 4. Relationship between speed and total flow of intersection.

Results of maximum flows (capacity) at several intersections with various speed levels have been presented in Table 4. The average speed of vehicles at intersections have a significant impact on total (maximum) flow where small differences of speed would indicate large differences in maximum flow (ΔQ=575.70 pcu/h, σ=293.60). While the model was developed by the portion of streams' flow, it is necessary to create a model which is suitable for total flow of the intersection, Q_{total} and average speed of the intersection, V. Based on data flow and speed at 10 km/h, 11 km/h, 12 km/h, 13 km/h, 14 km/h and 15 km/h, were plotted to show the relationship between speed and flow of unsignalized intersections and the following form V=16.863–$0.0015Q_{\text{total}}$ was found. Based on the relationship, it can be concluded that the free–flow speed at intersection is 16.863 km/h and the speed decreasing by 1.50 km/h for every 1000 pcu/h. The Method could provide a new technique to determine a critical speed of unsignalized intersection under mixed traffic flow where the capacity/maximum flow is reached. The result is shown in Figure 4.

Conclusions and Recommendations

The model of capacity which is defined as the maximum possible flow of the intersection was developed corresponding to the relationship of speed and flow of streams at each group of conflict. Based on traffic flow measurements and speed prediction at the maximum flow of a stream, the total capacity of an intersection can be calculated and the maximum flow of each stream can also be measured. Since there were no data of speed at the maximum flow (capacity) at intersections, the study has made an assumption for the speed which also refers to the secondary data (previous study). All possible maximum flows at intersections have been calculated by using the model (equation) and use the matrix for simplification. The real maximum flow as the capacity of the intersection is the least of all possible maximum flows.

Each of the movements at an intersection has been observed related to their speed and flow. By those two parameters, investigations have been made further at any correlation between conflict streams (six streams; C–A, C–B, B–C, B–A, A–C, A–B) as defined by group of conflicts (I, II, III, IV, V, VI). Results of the relation between parameters show that a suitable correlation between speed and flow of conflict groups could be developed (corresponds to the R^2) even if there was only a small correlation at some groups, e.g. conflict group–IV, stream B–A (which is thought to be an impact of the lack of lane discipline that drivers tend to use other lane paths passing through the intersection with the consequence of higher speed). It was also found that a good estimation of the relation between volumes of streams and intersection occupancy could be obtained, especially with 5–minute intervals of observation.

Further analysis has taken into account the speed at conflict points or the speed of each movement through the intersection. The volume of each movement is the most important parameter to calculate the maximum flow (capacity) based on the conflict streams. Maximum flows of each stream were found to correspond to the speed and flow of other streams at a group of conflict. There are six (6) alternatives of maximum flows at the intersection because the maximum flow of each stream has to be counted.

Each stream was observed related to its speed and flow at its own group of conflict. Theory of conflict was then adopted in analysis. It was assumed that each stream has reached its maximum flow, Q_i^j (stream i and alternative j) at the smallest speed, V_i'. When one stream reaches its maximum flow, e.g. $Q_{C\text{-}A}^{(1)}$, $V_{C\text{-}A}'$ means that other streams would not meet their (real) maximum flow ($Q_{C\text{-}B}^{(2)}$, $Q_{B\text{-}C}^{(3)}$, $Q_{B\text{-}A}^{(4)}$, $Q_{A\text{-}C}^{(5)}$, $Q_{A\text{-}B}^{(6)}$) and their (real) speed ($V_{C\text{-}B}'$, $V_{B\text{-}C}'$, $V_{B\text{-}A}'$, $V_{A\text{-}C}'$, $V_{A\text{-}B}'$). By using the value of maximum flow, e.g. $Q_{C\text{-}A}^{(1)}$, the speed, $V_{C\text{-}A}'$ and the streams' speeds $V_{C\text{-}B}$, $V_{B\text{-}C}$, $V_{B\text{-}A}$, $V_{A\text{-}C}$, $V_{A\text{-}B}$, other streams' flow, $Q_{C\text{-}B}^{(1)}$, $Q_{B\text{-}C}^{(1)}$, $Q_{B\text{-}A}^{(1)}$, $Q_{A\text{-}C}^{(1)}$, $Q_{A\text{-}B}^{(1)}$ can easily be calculated from the regression equations.

References

Brilon, W (Editor). (1988). "Intersections without traffic signals." *Proceedings of an International Workshop*, Bochum, West Germany.

Brilon, W (Editor). (1991). "Intersections without traffic signals II." *Proceedings of an International Workshop*, Bochum, West Germany.

Brilon, W., R. Koenig., and R. Troutbeck. (1997). "Useful Estimation Procedures for Critical Gaps." *Proceedings of the Third International Symposium on Intersections Without Traffic Signals*, Portland, Oregon USA, 71-87.

Brilon, W., R. Troutbeck., and M.Tracz. (1997). "Review of international practices used to evaluate unsignalized intersections." *Transportation Research Circular*, 468, Transportation Research Board, National Research Council, Washington, D.C.

Brilon, W., and N. Wu. (2001). *Capacity at TWSC Derived by Conflict Technique*, Lehrstuhl für Verkehrswesen, Ruhr–Universität Bochum, Bisher Unveröffentlicht.

BÄNG, K.-L., Carlsson., A., and Palgunadi. (1995). "Development of speed – flow relationships for indonesian rural roads using empirical data and simulation." *Transportation Research Record 1484*, TRB, National Research Council, Washington D.C., 24-32.

BÄNG, K–L.and Heshen AI. (2000). "Development of Capacity Guidelines for Road Links and Intersections for Henan and Hebei Provinces", PRC, *Transportation Research Circular E–C018 : 4th International Symposium on Highway Capacity*, TRB, National Research Council, Washington, D.C.

Budhy Tjahjati S. Soegijoko.; Sharif I. Horthy. (1991). "Role of Nonmotorized Transport Modes in Indonesia Cities", *Transportation Research Record 1294*, TRB, National Research Council, Washington D.C., 16-25.
Committee On Highway Capacity and Quality of Service. (1997). "Review of International Practices Used To Evaluate Unsignalized Intersections", *Transportation Research Circular, Number 468, Transportation Research Board*, National Research Council, Washington, D.C.

Chandra, S.; Sikdar, P. K. (2000). "Factors Affecting PCU in Mixed Traffic Situations in Urban Roads"**,** *Road and Transport Research*, 9(3), 40-50.

Directorate General of Highways, Ministry of Public Works. (1997). *Indonesian Highway Capacity Manual*, Jakarta.

Gaca, S..and J. Chodur. (1998). "Research on Capacity and Traffic Conditions at Unsignalized Urban Intersections", *Third International Symposium on Highway Capacity*, 1, Road Directorate, Copenhagen, Denmark.

Gleue, A. W. (1972). "Vereinfachtes verfahren zur Berechnung signalgeregelter Knotenpunkte", *Forschung Strassenbau und Strassenverkehrstechnik*, 136, Bonn.

Harders, J. (1968). "Die Leistungsfähikeit nicht signalgeregelter staädtischer Verkehrsknoten", Forschungberichte *Strassenbau und Strassenverkehrstechnik*, Bonn.

Huber, Matthew J. (1982). "Estimation of Passenger–Car Equivalents of Trucks in Traffic Stream", *Transportation Research Record 869*, TRB, National Research Council, Washington D.C., 60-70.

TRB. *Highway Capacity Manual 2000*, National Research Council, Washington, D.C.

Kockelke, Wilhelm. (1991). "Untersuchungen zum Fahrerverhalten bei Rechts–vor–Links Regelung in Tempo 30 – Zonen", *Bundesanstalt für Straßenwesen*, Heft 236, Bergisch Gladbach.

Kyte, M.; W. Kittelson.; Z. Tian; B. Robinson; M. Vandehey.(1996). "Analysis of Operations at All–Way Stop–Controlled Intersections by Simulation", *Transportation Research Record 1555*, TRB, National Research Council, Washington, D.C.

Kyte,M.Z.Tian,Z.Mir.,Z.Hameedmansoor.,W.Kittelson.,M.Vandehey.,B.Robinson. W.Brilon.,L.Bondzio.,N.Wu.and R.Troutbeck.(1996). "Capacity and Level of Service at Unsignalized Intersections", National Cooperative Highway Research Program (NCHRP) 3-46.

Kyte, M., and Joseph M. (1990). "Estimating Capacity and Delay at a Single–Lane Approach, All–Way Stop–Controlled Intersection", *Transportation Research Record 1225*, TRB, National Research Council, Washington, D.C.

First Order Phase Transition in a New Cellular Automaton Model within the Framework of Kerner's Three-Phase Traffic Theory

Junfang Tian, Bin Jia [1], Xingang Li and Fangfei Zhang

[1] MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing, 100044, People's Republic of China, Email: jiabin@jtys.bjtu.edu.cn

Abstract

The cellular automata model proposed by Tian and Jia, was established in the framework of Kerner's three-phase traffic theory and able to reproduce the three-phase traffic flow. But the first order transition from free flow to synchronized flow, which has been observed in the real traffic, could not reproduced in Tian and Jia model. One can see that the synchronized flow begin to emerge in Tian and Jia model, when the density is very low, which is obviously not in consistent with the real traffic. Therefore, we improved their model by the following main aspects. (i) The randomisation probability function and the deceleration extent in the new model have considered more situations according to the real traffic than that of Tian and Jia model. (ii) The slow-to-start rule in the new model is different with that of Tian and Jia model, since the slow-to-start rule in Tian and Jia model is not insistent with the real traffic. (iii) According to the new model, all vehicles will go through the randomisation step first according to its updated rules. Through these improvements, the new model can successfully reproduce the first order transition from free flow to synchronized flow, which is clearly shown on the fundamental diagram. Moreover, one can see the coexistence of the free flow and synchronized flow on the space-time plots, which is strong evidence of this transition. Then the 1-min average flux density diagram, the correlation functions are presented to identify the synchronized flow state. All results have shown that the new model is well consistent with the empirical findings.

Key Words: phase transition; synchronized flow; traffic flow

Introduction

Traffic phenomena have attracted the interest of more and more physicists (e.g., Chowdhury 2000; Helbing 2000,2001; Kerner 2004; Nagatani 2002; Jia 2007). Many traffic flow models have been proposed, including car-following models(e.g., Gazis 1961; Bando 1995; Helbing 1998; Jiang 2001), cellular automation models (e.g., Nagel 1992; Fukui 1996; Li 2001; Kerner 2002; Gao 2007; Knospe 2000; Jiang and Wu 2003; Lee 2004; Jia and Li 2009; Tian and Jia 2009), and the hydrodynamics models (e.g., Whitham 1955; Richard 1956; Payne 1971). Among these models, cellular automation models (CA) models form one type of most widely used models. CA has the advantage of flexible evolution rules and high computation efficiency. Although so many CA models had been proposed, just a few of them could reproduce the synchronized flow.

Based on the empirical traffic data, Kerner pointed out that there exists synchronized flow in the real traffic (Kerner 2004). Then congested traffic flow has been classified into synchronized flow and wide moving jams. Later, Kerner et al. proposed the three-phase traffic theory. The three phases are free flow, synchronized flow and wide moving jams. In synchronized flow, the speed of vehicles is considerably lower but the density of vehicles is considerably higher than the corresponding values in the free flow at the same flux. In the three-phase traffic theory, traffic breakdown is a phase transition from free flow to synchronized flow (F—>S transition). Wide moving jams can occur spontaneously in synchronized flow only (S—>J transition), i.e. due to a sequence F—> S —>J transitions. In real traffic, a fluctuation, whose amplitude exceeds the critical amplitude, occurring in the vicinity of the bottlenecks in the free traffic flow, will lead to the transition from free flow to synchronized flow (F —> S transition). Jams emerge in the synchronized flow, i.e. narrow moving jams, which spontaneously emerge in the synchronized flow, move and grow in the upstream direction. Finally, these narrow moving jams (or a part of them) transform into wide moving jams (S —> J transition). And all the transitions among the three phases are the first order transitions.

In this paper, we simplified and improved the cellular automaton model proposed by Tian & Jia (Tian and Jia 2009), which is established within the framework of Kerner's three-phase traffic theory and able to reproduce the three-phase traffic flow. But in Tian & Jia model, there exist three main drawbacks: (i) The update rules are very complicated. (ii) In the relatively low density (such as density $K \geq 0.05$), the synchronized flow begin to emerge, which is not in consistent with the real traffic. (iii) The discontinuous transition from free flow to synchronized flow could not reproduce in their model. In fact, some of the existing models which could reproduce synchronized flow also have the problems that too complicated or could not reproduce the first order phase transition. Such as, the update rules of the model proposed by Lee which could exhibits some features of synchronized patterns and general patterns is too complicated (Lee 2004). The model produced by Jiang & Wu faces the problem that could not reproduce the first order phase transition (Jiang and

Wu 2003). Although Jiang & Wu solved this problem later by adding a new brake rule, this new brake rule is not in accordance with the real traffic (Jiang and Wu 2005).

The paper is organized as follows. In the following section, we will introduce the new model. In section III, results are analyzed in detail. The fundamental diagram and the spatiotemporal patterns are presented, which can exhibit synchronized flow and metastable states. At last, the conclusion is given in section IV.

The New Model

The parallel update rules of our model are as follows.

Step 1. Determining the randomisation probability $p_n(t+1)$ and the deceleration extent Δv.

$$p_n(t+1)=\begin{cases} if\,(v_n(t)>v_c)\ then \begin{cases} p_a,\ if\,(b_{n+1}(t)=1\ and\ t_h<t_s) \\ p_b,\ else \end{cases} \\ if\,(v_n(t)\le v_c)\ then\ p_c \end{cases}$$

$$\Delta \mathrm{v}=\begin{cases} if\,(v_n(t)>v_c)\ then \begin{cases} d_1,\ if\,(b_{n+1}(t)=1\ and\ t_h<t_s) \\ d_2,\ else \end{cases} \\ if\,(v_n(t)\le v_c)\ then\ d_3 \end{cases}$$

Step 2. Randomisation with probability $p_n(t+1)$:

$$v_n(t+1)=\max(v_n(t)-\Delta \mathrm{v},0)$$

Step 3. Acceleration:

$$if\,(b_n(t)=0\ or\ t_h\ge t_s)\ then\ v_n(t+1)=\min(v_n(t+1)+1,v_{\max})$$

Step 4. Braking:

$$v_n(t+1)=\min(v_n(t+1),d_n^{eff})$$

Step 5. Slow-to-start:

$$if\,(v_n(t)=0\ or\ rand()<p_d)\ then\ v_n(t+1)=\max(v_n(t+1)-1,0)$$

Step 6. The determination of $b_n(t+1)$:

$$if\,(v_n(t)\le v_n(t+1))\ then\ b_n(t+1)=0$$

$$if\,(v_n(t)>v_n(t+1))\ then\ b_n(t+1)=1$$

Step 7. Car motion

$$x_n(t+1)=x_n(t)+v_n(t+1)$$

Here $x_n(t)$ is the position of vehicle n at time t (vehicle $n+1$ precedes vehicle n). $v_n(t)$ is the velocity of vehicle n at time t . $b_n(t)$ is the status of the brake light($b_n(t)=1(0)$ means the brake lights is on(off)). $t_h=(d_n(t)/v_n(t))$ is the time gap that the following vehicle reaches the position of the rear of leading car and $d_n(t)=x_{n+1}(t)-x_n(t)-l$ is the gap between vehicle n and its preceding vehicle $n+1$. l is the length of the vehicle. $t_s=\min(v_n(t),h)$ is the safe time gap, h determines the range of interaction with the brake light. $d_n^{eff}(=d_n(t)+\min(v_{anti}-gap_{safety},0)$ is the effective distance. $v_{anti}=\min(v_{n+1}(t),d_{n+1}(t))$ is the expected velocity of the

preceding vehicle in the next time step and gap_{safety} controls the effectiveness of the anticipation. $rand()$ is a random number between 0 and 1. v_c is the critical velocity to determine the randomisation probability and the deceleration extent. Step 5 is the slow-to-start rule in order to reproduce jams.

In the new model, one can see there exist three main different aspects, compared with Tian & Jia model: (1) The randomisation probability function and the deceleration extent of the new model have consider the fact that when the velocity of the new model is small, the randomisation probability and the deceleration extent is also small, because a bigger randomisation probability and deceleration extent may make the vehicle stop, which seldom happens in the real traffic. (2) The slow-to-start rule is different from that of Jiang and Wu (2003), since Li's research (Ling X. G. 2009) have found that this slow-to-start rule is more insistent with the real traffic than that of Jiang and Wu. (3) In order to reproduce the first order transition from free flow to synchronized flow, we let randomisation step (step 1) go first. We will explain that why this could reproduce the synchronized flow in the future. Moreover, one could notice that there exist two steps of randomisation in the new model, i.e. step 1 and step 5. It should be said that step 5 is only used to reproduce wide moving jams and has no relation with the reproduction of the synchronized flow.

In the next section, the simulations are carried out. In the simulations, $p_a = 0.95, p_b = 0.1, p_c = 0.05, p_d = 0.5, d_1 = 2, d_2 = 1, d_3 = 1, v_c = 2, v_{\max} = 20, gap_{safety} = 7$, $h = 4$. Each cell corresponds to 1.5 m and a vehicle has a length of five cells, $l = 5$ .One time step corresponds to1s, the system size is set to L = 10000.

Simulations

Firstly, we carry out the simulations of the new model under the periodic boundary condition. Figure 1 shows the fundamental diagram, in which three traffic phases and two first order transitions are clearly distinguished.

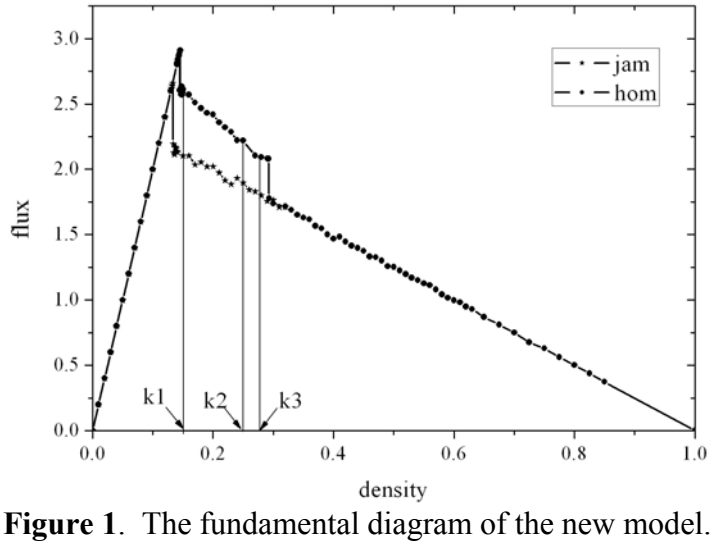


Figure 1. The fundamental diagram of the new model.

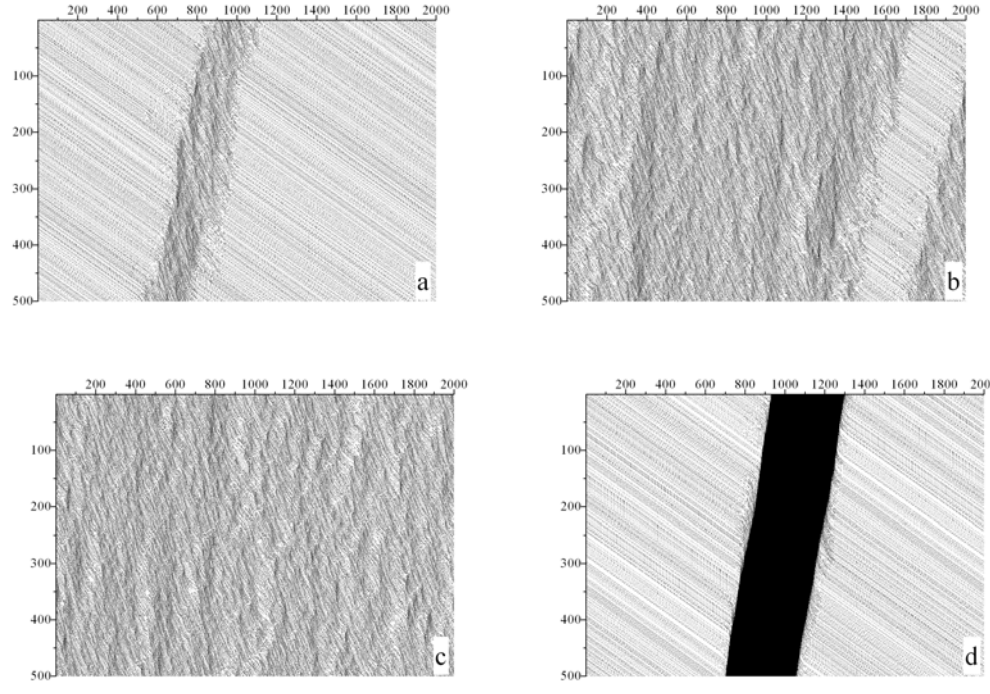


Figure 2. The spatiotemporal diagram of the new model.(a) k1=0.15, (b) k2=0.25, (c) k3=0.275, (d) k3=0.275. In (a), (b) and (c) the traffic starts from a homogeneous initial state. In (d) the traffic starts from a megajam. The horizontal direction (from left to right) is space and the vertical direction (from top to bottom) is time.

The upper branch is obtained from homogeneous initial states, in which all the vehicles have the same speed and gap. And the lower branch is obtained from megajam. Next, we will choose three density points (k1<k2<k3) to show the synchronized flow situations on the road. From figure.2(a), one can see that when the density k=k1, there exist two kind of traffic flow, i.e., free flow and synchronized flow, which is different with Tian & Jia model that just synchronized flow exists (see Tian and Jia 2009, figure.5(b)). When the density increases, the region of the synchronized flow gradually expands, figure.2 (b). With the further increase of density, the free flow disappears, figure.2 (c).

In order to identify the synchronized flow, time series data of flow rate J, velocity v and local density k are analysed. The data are obtained by a virtual detector. J is the number of vehicles passing the detector per time step within 1 min and v is the average velocity of the vehicles passing the detector within 1 min. Then k is calculated by k= J /v. Figure. 3 shows the 1 min average flux against density of the new model. The data is widely scattered, which is consistent with observed features of synchronised flow. Furthermore, the autocorrelation function:

$$\alpha_x(\tau) = \frac{< x(t)x(t+\tau) > - < x(t) >^2}{< x(t)^2 > - < x(t) >^2}$$

The autocorrelations of flux, average velocity, and local density, are shown in Fig. 4(a). Although there are some fluctuations in the values of the autocorrelations, they are close to zero.
Moreover considering the cross-correlation function:

$$c_{xy}(\tau) = \frac{< x(t)y(t+\tau) > - < x(t) >< y(t) >}{\sqrt{< x(t)^2 > - < x(t) >^2}\sqrt{< y(t)^2 > - < y(t) >^2}}$$

The cross-correlation between density and flux is shown in Fig. 4(b). One can see that it vanishes in large time scale. Such results indicate that the new model could capture the dynamics of synchronized flow.

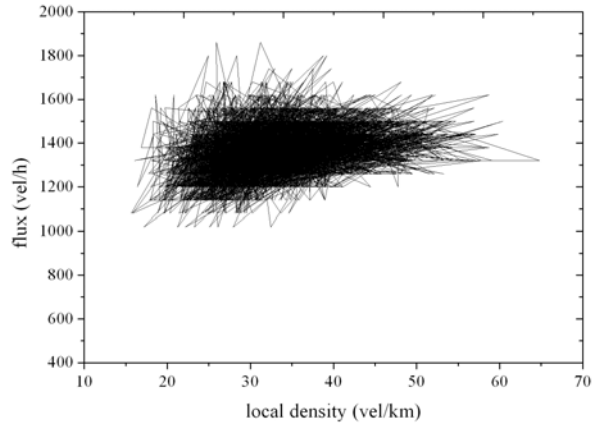


Figure 3. The 1 min average flux against density of the new model. The global density k=0.275.

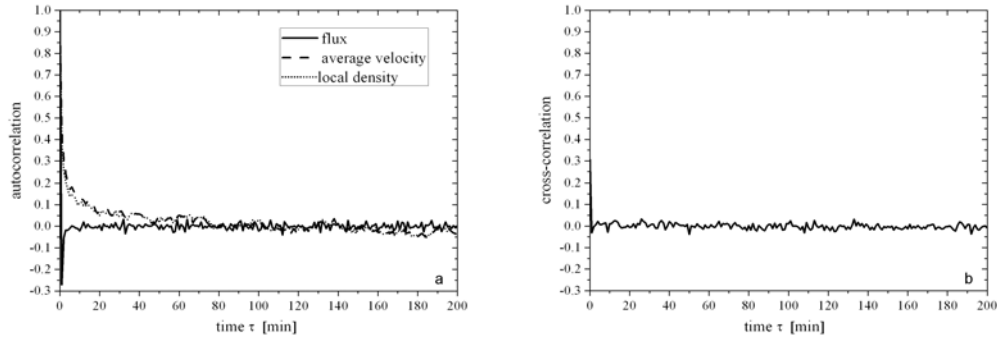


Figure 4. (a) The autocorrelation functions of flux, average velocity, local density. (b) The cross-correlation function between local density and flux. The global density k=0.275

Conclusions

In this article, a new CA model is proposed to capture the properties of traffic flow in the framework of Kerner's three-phase traffic theory. Although it is based on the

Tian & Jia model, there still exist many differences between two models, such as the acceleration and randomization steps, the slow-to-start rule etc. The most significant is that, the new model successfully reproduces the first order transition from free flow to synchronized flow which could not appear in Tian & Jia model.

The fundamental diagram is analysed. The space-time plots of different phases are shown. The 1-min average flux density diagram, the correlation functions are presented. All results have shown that the new model is well consistent with the empirical findings.

Acknowledgements

This work is financially supported by 973 Program (2006CB705500), the National Natural Science Foundation of China (Nos. 70631001,70501004 and 70701004), Program for New Century Excellent Talents in University (NCET-07-0057) and the Natural Science Foundation of Beijing (9093020).

References

Bando, M., Hasebe, K., Nakayama, A., Shibata, A. and Sugiyama, Y. (1995). "Dynamical model of traffic congestion and numerical simulation." *Phys.Rev.E*, 51, 1035-1042.

Chowdhury, D., Santen, L., and Schadschneider, A. (2000). "Statistical physics of vehicular traffic and some related systems." *Phys. Rep*, 329,199-329.

Fukui, M., Ishibashi, Y. (1996). "Traffic Flow in 1D Cellular Automaton Model Including Cars Moving with High Speed." *J. Phys. Soc*. Jpn, 65,1868-1870.

Kerner, B. S. (2004) *The physics of traffic*, Springer, Heidelberg.

Kerner, B. S., Klenov, S. L., and Wolf, D. E. (2002). "Cellular automata approach to three-phase traffic theory." *J. Phys. A*, 35,9971-10013.

Gao, K., Jiang, R., Hu, S. X., Wang, B. H. and Wu, Q. S. (2007). "Cellular-automaton model with velocity adaptation in the framework of Kerner's three-phase traffic theory." *Phys. Rev.E*, 76,026105.

Gazis, D. C., Herman, R., and Potts, R. B. (1961). "Nonliner follow-the lead models of traffic flow." *Oper.Res*, 9,545-567.

Helbing, D. (2000) *Traffic and Granular Flow'99*, Springer, Heidelberg.

Helbing, D. (2001). "Traffic and related self-driven many-particle systems." *Rev. Mod. Phys*, 73,1067-1141.

Helbing, D., Tilch, B. (1998). "Generalized force model of traffic dynamics." *Phys.Rev.E*, 58,133-138.

Jia, B., Gao, Z. Y., Li, K. P. and Li, X. G. (2007). "Models and Simulations of Traffic System Based on the Theory of Cellular Automaton", *Science*, Beijing.

Jia, B., Li, X. G., Chen, T., Jiang, R. and Gao, Z. Y. (2009). "Cellular automaton model with time gap dependent randomisation under Kerner's three-phase traffic theory." *Transportmetrica*, 99999,1.

Jiang, R.,Wu, Q. S. and Zhu, Z. J. (2001). "Full velocity difference model for a car-following theory." *Phys.Rev.E*, 64,017101.

Jiang, R., Wu, Q. S. (2003). "Cellular automata models for synchronized traffic flow." *J. Phys. A: Math.* Gen, 36,381-390.

Jiang, R., Wu, Q. S. (2005). "First order phase transition from free flow to synchronized flow in a cellular automata model." *The European Physical Journal B*, 46,581.

Knospe, W., Santen, L., Schadschneider, A. and Schreckenberg, M. (2000). "Towards a realistic microscopic description of highway traffic." *J. Phys. A*, 33,L477-L485.

Lee, H. K., Barlovic, R., Schreckenberg, M., and Kim, D. (2004). "Mechanical Restriction versus Human Overreaction Triggering Congested Traffic States." *Phys.Rev.Lett*, 92,238702.

Li, X. B., Wu, Q. S. and Jiang, R. (2001). "Cellular automaton model considering the velocity effect of a car on the successive car." *Phys. Rev. E,* 64,066128.

Li, X. G. (2009). "Microscopic Modeling and Properties Analyzing of Transportation System based on Cellular Automata Model." *Ph.D. Dissertation*, School of Traffic and Transportation, Beijing Jiaotong university.

Nagatani, T. (2002). "The physics of traffic jams." *Rep. Prog. Phys*, 65,1331-1386.

Nagel, K., and Schreckenberg, M. (1992). "A cellular automaton model for freeway traffic." *J. Phys. I France 2*, 12,2221-2229.

Payne, H. J. (1971). "Models of free traffic and control, In: Bekey G.A.(ed.)." *Mathematical Methods of Public Systems*, 1,51-61.

Richard, P. I. (1956). "Shock waves on the high way." *Operations Research*, 4,42-51.

Tian, J. F., Jia B., Li, X. G., Jiang, R., Zhao, X. M., and Gao, Z. Y. (2009). "Synchronized traffic flow simulating with cellular automata model." *Physica A*, 388, 4827-4837.

Whitham, M. J., Whitham, G. B. (1955). "On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads." *Proc. Roy. Soc.* London Ser.A, 229(1178),317-345.

Traffic Prediction In Metropolitan Freeways

Carlo Liberto, Roberto Ragona and Gaetano Valenti

[1] Environment New technologies and Energy Agency (ENEA),'Casaccia' Research Centre,via Anguillarese 301- 00060, Rome, Italy; Email: carlo.liberto@enea.it

Abstract

The successful wide scale deployment of Advanced Traveler Information Systems (ATIS) and Advanced Traffic Management Systems (ATMS) relies significantly on the capability to perform accurate short-term predictions of traffic parameters over the entire road network. Most of previous studies and applications of traffic prediction have developed statistical or heuristic methods based on historical or real-time traffic data collected by fixed sensor networks. Meanwhile the effectiveness and wide applicability of such methods is strongly affected by the high cost of deploying and maintaining fixed sensor networks able to provide large spatial coverage. Recently, mobile sensors or probe vehicles appeared as a complementary solution to fixed sensors in order to increase coverage areas and prediction accuracy without requiring expensive infrastructure investments. The focus of this paper is to present findings of a study on short-term traffic prediction based on data collected by a probe vehicle system operating on the Italian national scale and consisting of a large fleet of privately-owned vehicles. Particularly the paper proposes and compares two models, respectively based on Artificial Neural Networks and Pattern-Matching techniques, designed to on-line perform short-term (15 to 60 minutes) predictions of link travel speeds. The Rome Ring Road (GRA - Grande Raccordo Anulare) is used to assess the feasibility of the two models. Testing results showed that the proposed approaches for short-term predictions are very promising and effective.

Key Words: short-term traffic forecasting; real-time floating car data; artificial neural network; pattern matching

Introduction

The successful wide scale deployment of Advanced Traveler Information Systems (ATIS) and Advanced Traffic Management Systems (ATMS) relies significantly on the capability to perform accurate short-term predictions of traffic parameters such as traffic volumes, travel times and travel speeds (as a proxy for travel times) [7], [3], [8], [10].

Most of the studies and applications conducted so far have used statistical or heuristic techniques to predict traffic conditions starting from historical and real-time traffic data collected by fixed sensor networks [11], [13]. Among these traffic sensors may be mentioned loop detectors, that provide flows and speeds through a road section, and automatic vehicle identification systems, that keep track of the vehicles passing through a particular installed device/checkpoint and compute travel times from the difference between arrival times at consecutive checkpoints. However these traffic information collection techniques have limitations in terms of spatial coverage due to the high cost of installation and maintenance.

Recently, mobile sensors or probe vehicles appeared as a complementary solution to fixed sensors for increasing coverage areas and prediction accuracy without requiring expensive infrastructure investments. Unlike other existing traffic data collection techniques (Automated Vehicle Identification systems, video cameras, inductive loops, radar based sensors, etc.), probe vehicles act as moving sensors traveling in a traffic stream and do not require instrumentation to be set up on the roadway (Yang, 2003). Key and innovative aspects of probe vehicle technology are that vehicle travel time is measured directly and the quality of data is fixed by the percent of vehicles monitored. Moreover travel speed is inferred from travel time and speed estimates are space-mean speeds as opposed to point speeds measured by fixed-point traffic sensors. The space-mean speed is the average speed of vehicles traveling a given segment of roadway during a specified period of time and is calculated using the average travel time and length of the road segment.

The focus of this paper is to present the results of a study on short-term traffic prediction based on information provided by a probe vehicle system (PVS) operating on the Italian national scale. Unlike previously proposed probe-based traffic monitoring systems (mostly using data from taxis or bus fleets), this system, operated by OCTOTelematics (www.octotelematics.com), exploits data from a large fleet of privately owned vehicles, to deliver real-time traffic speed information throughout the Italian motorway network (more than 6000 Km) and along some important arterial streets located in major Italian metropolitan areas. Real-time traffic speed estimates are currently deduced from GPS traces wirelessly transmitted from a large and still growing number (about 800.000 units) of privately owned vehicles corresponding to approximately 2% of the national fleet. Each vehicle is equipped with a specific device including a GPS receiver and a GSM/GPRS transmitter and

covering a range of insurance-related applications. Link travel speed estimates are regularly updated at intervals of 3 minutes 24 hours a day and delivered to Infomobility service providers and motorway/roadway operators.

In the context of this data source the paper proposes and compares two models, respectively based on Artificial Neural Networks and Pattern-Matching techniques, designed to on-line perform link travel speeds predictions for a time horizon of 15 minutes up to one hour.

The paper is organized as follows: Section 2 provides an overview of the PVS operated by OCTOTelematics; Section 3 presents a brief description of the testing case study: the Rome ring road; Section 4 outlines the basic definitions and properties of the two proposed models; Section 5 offers a comparison of the results obtained from the application of the two proposed approaches and Section 6 concludes this paper.

Outline of The "Octotelematics" Probe-Vehicle System

The target of the PVS operated by OCTOTelematics is basically insurance profiling. This PVS is based on the wireless exchange of information between a large fleet of privately-owned probe vehicles traveling across Italy and a Data Processing Centre (DPC) which performs a variety of functions such as statistics computation on driver behavior, mileage, accident detection and reconstruction, real-time traffic condition estimation, antitheft satellite tracking and fleet management.

Each vehicle is equipped with an on-board unit (OBU) that integrates the following components: a GPS receiver, a GPRS transmitter, a 3-axis accelerometer sensor, a battery pack, a mass memory, a processor and RAM. The OBU stores GPS measurements (position, heading, speed, quality) and periodically transmits (on request or automatically) the recent accumulated measurements to the DPC. Transmission occurs every 100 km traveled or every 12 minutes when the OBU equipped vehicle is traveling along the Italian motorway network or along freeways in major Italian metropolitan areas.

As aforementioned, a primary function performed in the DPC is to collect and process the received location and time information from the probe vehicles in order to provide real-time estimates on traffic conditions, in terms of average link speeds, every 3 minutes, 24 hours a day, 7 days a week.

The method to infer traffic conditions from probe vehicles' position and time information involves 4 sequential steps, as follows: 1.Map-matching; 2.Path identification; 3.Travel time allocation 4.Travel speed aggregation. The first step determines the most likely position of the vehicle on the road network given the reported locations. The second step identifies the most likely path taken by the vehicle to travel between two consecutively matched positions. The third step allocates the path travel time to the individual links traversed by the vehicle. The last

step combines and transforms link travel time from individual probe vehicles into aggregate estimates of the average link travel speed for all vehicles. In the travel speed aggregation procedure only current and near-past link travel time values are considered and an exponential weighting function is applied giving greatest weight to the most recent link travel time values.

Field tests were performed by comparing measured values coming from the Automatic Vehicle Identification systems installed on the Italian motorway network. The results indicated that the accuracy of the link travel speed estimates performed by this PVS is over 90% on the entire road network.

The Rome Ring Road Case Study

The case study selected for the development and evaluation of the proposed prediction methods is the Rome Ring Road (GRA - Grande Raccordo Anulare).
The GRA is a toll-free motorway (68.2 km in circumference) that encircles Rome (Figure 1). It is a dual carriageway with three lanes in both the clockwise (inner carriageway) and anticlockwise (outer carriageway) directions.

The GRA has 33 numbered entry / exit junctions (starting from "Aurelia" junction and proceeding in the clockwise direction) and is a major city traffic artery distributing traffic on radial routes and handling circumferential traffic in the city.

The GRA is a very challenging testing bed because it is located in one of the most congested metropolis in the country. Traffic is heavy for most of the day and frequent delays and traffic-jams are experienced, due to accidents or queue spillbacks from the exit ramps or the adjacent radial arterial streets leading into the city centre. The most severely congested links are in the east quadrant specifically from junction 10 "A1 – Roma Firenze" to junction 23 "Appia" (from km 21.0 to km 44.6, clockwise direction) and in the south quadrant from junction 30 "Fiumicino" to junction 18 "Casilina" (from km 7.6 to km 30.2, anticlockwise direction).

In a working day, about 15000 probe vehicles monitored by OCTOTelematics pass through the GRA. The average distances traveled by a floating car on the GRA is about 10 km. During the peak period, an average of more than 2000 floating cars per hour travel on the GRA.

The link travel speed time series used in this study are organized at 3-minute periods (480 values per link and per day) and cover the whole period ranging from April 2008 to November 2009. Currently, the penetration level of equipped private cars in Rome (about 2.5%) is significantly higher than the national average.

Link Travel Speed Prediction Models

Two modeling approaches have been proposed and tested in order to on-line perform short-term (15 to 60 minutes) predictions of link travel speeds. The proposed models,

based on Pattern Matching (PM) and Artificial Neural Networks (NN) techniques, have been developed to directly use current and near-past average link speeds estimated for the target link and the correlated upstream/downstream links.

The two models have been chosen due to their ability to account for spatial and temporal average speed information simultaneously. This feature is essential since near future conditions for a given link rely both on its previous status and on what happened in its correlated upstream/downstream links. That is, an increasing congestion in an upstream/downstream link may be a useful indicator of a forthcoming travel speed change for the target link.



Figure 1. The Rome Ring Road.

A preliminary analysis of historical time series of link travel speeds was undertaken to investigate their dynamic spatio-temporal evolution and indentify complete and appropriate time series for the learning and testing process of the models.

From this analysis, the month of November 2009 resulted particularly suitable for testing the prediction capability of the proposed models. Actually this month was characterized by the occurrence of a high number of traffic congestion events, due both to bad weather conditions and accidents. Time series from 6:00 am to 9:00 pm related to only working days have been taken into account, because traffic conditions are quite stable during the other days and the night time.

As an example Figure 2 outlines the extent and intensity of congestion observed in the GRA inner carriageway during the working days of November 2009. Most of the links presents a mean travel speed lower than 50 km/h for the 30% of the time interval taken into consideration. The highest congestion levels are from link “10” to “15” and from link “18” to “23”. Furthermore, strong variations of the daily speed

profiles with respect to the daily mean profile have been identified in this month. To stress this discrepancy, Figure 3 compares the daily speed profiles of the 5 Mondays of November 2009 with the "Monday" mean profile (calculated on a one year period excluding Summer and holidays) related to link "12" (Nomentana - Centrale del Latte).Note that November's speed profiles have often different shapes from the average one and the greatest differences occur in the most congested time ranges (i.e. from 8:00 am to 10:30 am and from 4:30 pm to 8:00 pm).

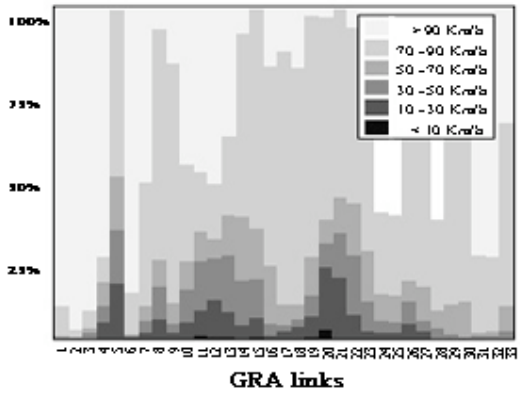


Figure 2. Time percentages for link travel speed range. Data refer to November 2009, working days only, from 6:00 am to 9:00 pm

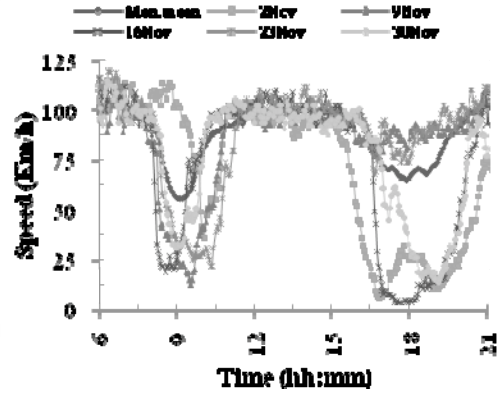


Figure 3. Link 12 mean velocity as a function of time for the November's Mondays. The red line represents the "Monday" mean profile.

Neural Networks. Artificial Neural Networks are quantitative models which learn to associate input and output patterns adaptively with the use of learning algorithms without understanding the fundamental or physical relationships between them [4], [6], [12]. In general, feedforward NNs consist of processing elements (neurons), or units, arranged in layers and relations between neighboring layers' units, whereas units in the same layer are not connected to each other. Feedforward NNs comprise an input layer, one or more hidden layers and an output layer, furthermore each layer contains different number of units.

As a rule, the type and number of units in the input layer, as well as the appropriate number of units and hidden layers are chosen through a preliminary analysis of the data or by empirically comparing the results from different NN architectures having different input variables and units in the hidden layers.

The relations between neighboring layers' units are defined by the weights given to the connections during the training process. The output of each unit is given by a transfer function fed up with the weighted sum of the incoming units values and then transmitted to all of the units in the next layer. Log-sigmoid or tan-sigmoid are commonly used as transfer functions.

In the GRA case study, four feedforward NN models for each link with one hidden layer (including at most 40 units) were proven to perform better link travel speed predictions in the learning-testing process.
The four NN models are aimed at predicting the link travel speed at 5, 10, 15 and 20 steps into the future corresponding respectively to 15, 30, 45 and 60 minutes. They incorporated as input the current and the near-past travel speed time-series of the target link and of the immediate neighboring upstream and downstream links, to reflect the correlation degree (higher than 0.60) between neighboring links found in a previous correlation analysis.

The NN learning-testing process was conducted on the PVS link travel speeds stored from 6:00 am to 9:00 pm. In particular, working days from September 2008 to December 2008 and during November 2009 were considered for the learning and testing process, respectively.

NN models were developed, implemented and tested using MatLab with extensive use of its programming language and Neural Network Toolbox functions. The adopted training was based on the Levenberg-Marquardt learning algorithm [5].

Pattern Matching. Pattern Matching is a very flexible method of forecasting that reports interesting results as the available time series data become sufficiently large. It can be considered as a model-free method in the sense that no structured modelling and consequent (model) parameter fitting is implied.

The basic idea is that traffic patterns are recurrent and that the evolution of historical data trend can be used to evaluate the future traffic scenario. This approach can be also seen as an application of the so-called lazy learning modelling [2], implying that past examples are considered the only representation of the system being modelled. For this reason, available historical database is searched for examples that, according to some figure of merit, are considered most similar to the actual one and extracted to fulfil the prediction requirement generally by means of local techniques.

In developing the model, a pattern of speed data was constructed for a specific link by considering the actual and the nearest past speed values of the target and of the spatial correlated upstream/downstream links (see Figure 4). That's a dynamical fragment represented on both temporal and spatial scales, whose time depth depends on the time forecasting horizon and on the involved links length. A pattern so structured constitutes the reference profile of the target link at the current time. The scanning of the historical database produces a quantity of profiles referred to preceding days with an identical structure, to be compared with the reference one in terms of an assumed criterion and degree of similarity.

The searching method for finding the possible profiles follows a computationallyreduced procedure, based on the scanning of all previous days in the historical database within a time frame of $\pm$ x minutes from current time [1]. Pattern

candidates are then evaluated in terms of their similarity to the reference profile and chosen (*matched profiles*) or discarded for the subsequent steps.

The matching process is performed through a measure of the Euclidean distance E_n between the target profile and the candidate profiles: only the past speed profiles having E_n within fixed limits () are admitted to the subsequent process of estimation. To assure a non-empty and sufficient populated ensemble of matched profiles, a number of the nearest profiles are in any case accepted.
Finally a speed forecasting is produced as a weighted average of the matched profiles time evolution (the more distant, the less weighted).

Figure 4 shows a simplified Pattern Matching procedure for link 12 (*Nomentana – Centrale del Latte*). Reference time is fixed at 11/11/09 17:00 and time forecasting horizon at 60 minutes. The pattern profile on the left of the "17:06" vertical line is formed by speed data of: three downstream links (number 15, 14 and 13), one upstream link (11) and the target link itself (12). The time evolution of the matched profiles is reported on the right of the "17:06" line (simple lines), whose weighted average describes the prediction trend (semicircle-shaped symbols). For a comparison the real trend is reported with triangle-shaped symbols.

Testing Results

Model testing has be done including all the 33 links related to the GRA inner carriageway. The analyzed period comprises all the 21 working days of November 2009 from 6:00 am to 9:00 pm. The *root mean square error* (RMSE) defined by equation (1) and the *mean absolute percentage error* (MAPE) defined by equation (2), have been calculated to investigate the accuracy of the models:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2} \quad (1)$$

$$MAPE = \frac{1}{N}\left(\sum_{i=1}^{N}\left|\frac{x_i - \hat{x}_i}{x_i}\right|\right)\cdot 100 \quad (2)$$

where N is the total number of testing cases (20 time steps/hour · 15 hours/day · 21 working days), x_i the measured link speed and $\hat{x}_i$ the predicted link speed.

The main goal was to investigate the stability of the models at different time prediction horizons. To this purpose, the analysis pipeline was first applied to a single link to better highlight the forecasting time dependencies and discover the daily profile of the predictions errors.

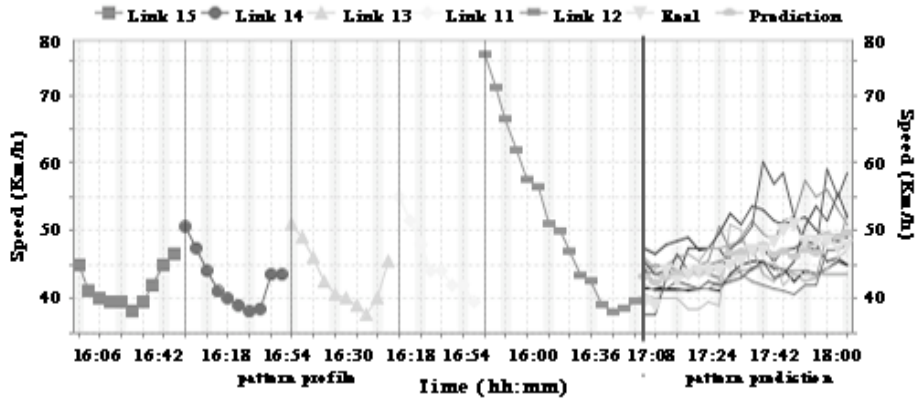


Figure 4. A pattern matching application example.

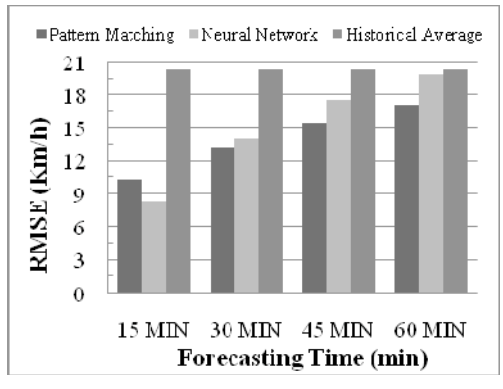


Figure 5. Speed RMSE at several time forecasting horizons for the link 12.

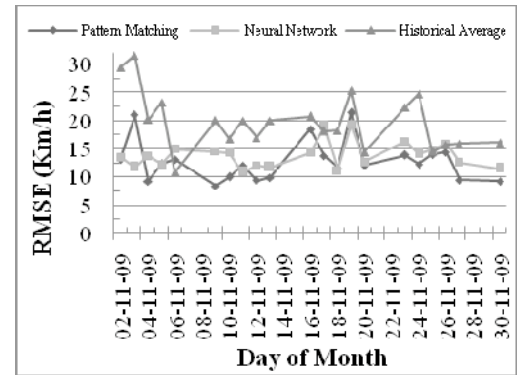


Figure 6. Speed RMSE at different working days of Nov. 2009 for the link 12. Time horizon is fixed at 30 mins.

Figure 5 reports the RMSE profiles for the link 12 (Nomentana – Centrale del Latte) versus the time forecasting horizons. It shows that, even if both modeling RMSE errors increase with time horizons, PM forecasting presents a lower growing rate.

For the sake of comparison, the historical mean speed, obtained averaging the data from September 2008 to June 2009 (only working days), and its associated errors have been also calculated with the aim to outline a reference model: in fact, the mean daily profile can be considered as the simplest model of forecasting.

Furthermore, focusing on a single time forecasting threshold of 30 minutes, it was possible to point out the RMSE daily fluctuations: it results to be very similar for both the investigated models apart from particular days (e.g. 03/11/2009) that are supposed to be associated with anomalous traffic situations (see Figure 6).

Subsequently, the analysis has been extended to the overall inner carriageway of the Rome ring road, achieving a confirmation of the single link study.

NN models show in fact better performances at very short time forecasting horizon (15 minutes), whereas PM is more suitable at longer prediction time (45 and 60 minutes).

Figure 7 reports the RMSE values at 15 minute forecasting horizon for each GRA link, showing that NN models give the most confident results except for links 9, 30 and 31. Errors range from 5.9 Km/h to 9.1 Km/h with a mean value of 7.4 Km/h for NN and from 5.6 Km/h to 10.7 Km/h with a mean value of 8.5 Km/h for PM. To emphasize the different performances of the methods at this fixed time forecasting horizon, the differences between PM and NN RMSE values have been calculated (Figure 8). The mean value of these differences is 1.1 Km/h while the maximum value can reach 2.8 Km/h.

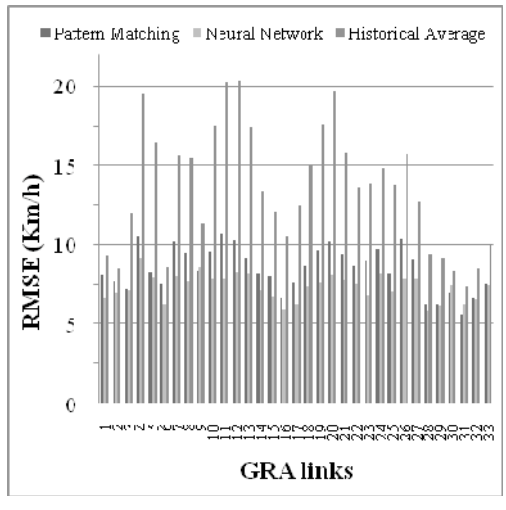


Figure 7. Speed RMSE at 15 minutes forecasting horizon for each GRA link.

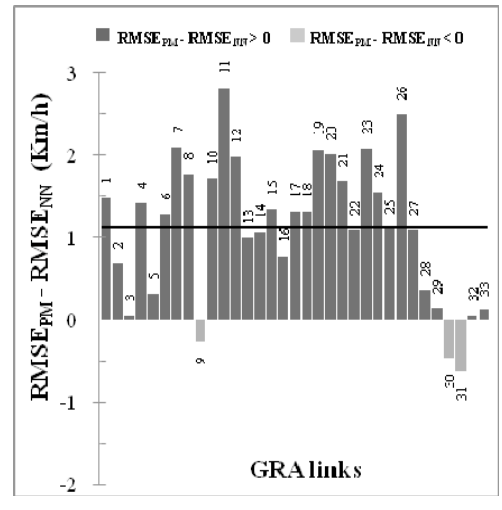


Figure 8. Differences between PM and NN speed RMSE differences at 15 minutes forecasting horizon.

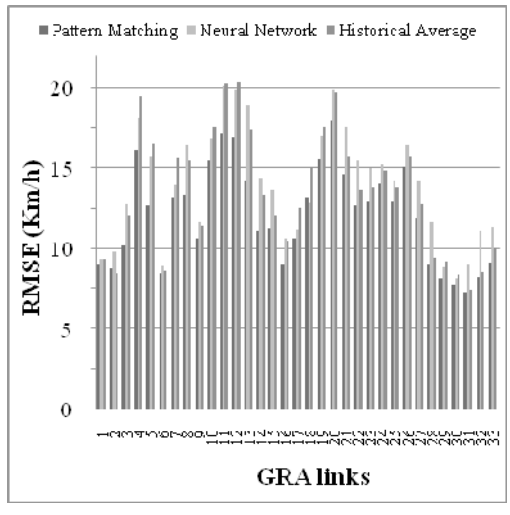


Figure 9. Speed RMSE at 60 minutes forecasting horizon for each GRA link.

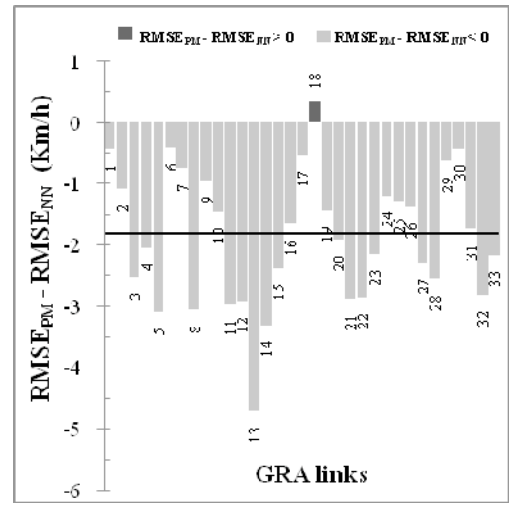


Figure 10. Differences between PM and NN speed RMSE differences at 60 minutes forecasting horizon.

Increasing the time forecasting threshold (Figure 9 and Figure 10), PM gives more stable results with respect to NN. Errors range from 8.1 Km/h to 20.1 Km/h with a mean value of 14.0 Km/h for NN and from 7.3 Km/h to 17.9 Km/h with a mean value of 12.1 Km/h for PM. The absolute maximum value of RMSE differences is 4.7 Km/h. The absolute mean value is 1.87 Km/h, about twice as much the value obtained for 15 the mins forecasting horizon.

Table 1. RMSE and MAPE at 15 min forecasting horizon for most congested links.

				PM	NN	HA		PM	NN	HA
Link num		Len. (Km)		RMSE (Km/h)				MAPE %		
10		0.7		9.6	7.9	17.6		12.6	9.1	27.5
11		4.9		10.7	7.9	20.3		16.9	10.4	39.0
12		1.7		10.3	8.3	20.4		15.0	10.9	32.4
13		1.6		9.2	8.2	17.4		12.9	10.4	25.3
14		1.6		8.2	7.1	13.4		11.5	8.8	20.1
15		1.3		8.1	6.7	12.1		11.5	9.0	19.6
18		0.5		8.7	7.4	15.0		9.7	7.4	21.1
19		1.3		9.7	7.6	17.6		13.8	9.7	35.4
20		2.3		10.2	8.2	19.8		17.4	12.9	44.3
21		0.4		9.4	7.7	15.7		14.5	11.5	25.6
22		2.4		8.6	7.5	13.7		12.2	10.2	19.5
23		3.6		9.0	6.9	13.8		10.9	7.9	17.6

Table 2. RMSE and MAPE at 30 min forecasting horizon for most congested links.

				PM	NN	HA		PM	NN	HA
Link num		Len. (Km)		RMSE (Km/h)				MAPE %		
10		0.7		12.3	12.7	17.6		16.3	15.0	27.5
11		4.9		13.5	13.0	20.3		21.6	18.5	39.0
12		1.7		13.3	14.0	20.4		20.0	16.7	32.4
13		1.6		11.4	12.7	17.4		16.3	16.1	25.3
14		1.6		9.5	10.0	13.4		13.8	13.6	20.1
15		1.3		9.6	9.9	12.1		14.2	13.4	19.6
18		0.5		10.9	10.8	15.0		12.8	10.9	21.1
19		1.3		12.6	12.0	17.6		19.1	16.3	35.4
20		2.3		13.8	15.1	19.8		24.6	24.3	44.3
21		0.4		11.9	11.9	15.7		18.6	17.5	25.6
22		2.4		10.4	10.6	13.7		14.8	14.4	19.5
23		3.6		10.8	10.4	13.8		13.3	11.5	17.6

Better performances at very short time forecasting horizon (15 minutes), whereas PM is more suitable at longer prediction time (45 and 60 minutes). Furthermore testing results show that RMSE values are peaked in correspondence to the most congested link groups, which deviated the most from the mean traffic conditions.

Future works will mainly address the problem to achieve reliable traffic predictions in case of congestions by the development of smoothing functions to filter out the

Table 3. RMSE and MAPE at 45 min forecasting horizon for most congested links.

				PM	NN	HA		PM	NN	HA
Link num		Len. (Km)		RMSE (Km/h)				MAPE %		
10		0.7		14.2	14.5	17.6		19.7	17.2	27.5
11		4.9		15.6	16.8	20.3		25.8	24.9	39.0
12		1.7		15.4	17.5	20.4		24.4	25.8	32.4
13		1.6		13.0	16.9	17.4		19.3	24.7	25.3
14		1.6		10.5	11.7	13.4		15.5	17.5	20.1
15		1.3		10.6	13.0	12.1		16.2	19.6	19.6
18		0.5		12.3	12.6	15.0		15.1	14.2	21.1
19		1.3		14.4	15.4	17.6		23.1	22.0	35.4
20		2.3		16.3	18.6	19.8		30.6	31.3	44.3
21		0.4		13.5	16.2	15.7		21.6	23.6	25.6
22		2.4		11.7	13.2	13.7		16.6	17.6	19.5
23		3.6		11.9	12.3	13.8		15.0	14.3	17.6

Table 4. RMSE and MAPE at 60 min forecasting horizon for most congested links

				PM	NN	HA		PM	NN	HA
Link num		Len. (Km)		RMSE (Km/h)				MAPE %		
10		0.7		15.4	16.9	17.6		22.3	20.9	27.5
11		4.9		17.2	20.1	20.3		29.5	31.8	39.0
12		1.7		17.0	19.9	20.4		28.1	32.2	32.4
13		1.6		14.2	18.9	17.4		21.9	30.4	25.3
14		1.6		11.1	14.4	13.4		16.9	22.9	20.1
15		1.3		11.2	13.6	12.1		17.9	22.1	19.6
18		0.5		13.2	12.9	15.0		16.7	16.2	21.1
19		1.3		15.6	17.0	17.6		26.3	25.4	35.4
20		2.3		17.9	19.9	19.8		35.6	40.1	44.3
21		0.4		14.6	17.5	15.7		23.9	27.1	25.6
22		2.4		12.7	15.6	13.7		18.3	20.2	19.5
23		3.6		12.9	15.0	13.8		16.3	17.7	17.6

As expected the RMSE plots peak in correspondence to the most congested link groups of Figure 2 (e.g. 10-15 and 18-23), which deviated the most from the mean traffic conditions.

Detailed results, including MAPE errors, have been reported in Tables from 1 to 4. Values refer to the most congested link groups for each analyzed time forecasting horizon.

Conclusions and Future Perspectives

Short-term traffic prediction is of ultimate importance for advanced traffic management systems and traveler information services applications.

In this study two approaches, based on Pattern Matching and Neural Networks techniques, were developed and tested to on-line derive short-term link travel speed predictions (15 to 60 minutes) from data collected by a probe vehicle system operating on the Italian national scale and consisting of a large fleet of privately-owned vehicles.

The two techniques were chosen to take into account the spatio-temporal dependency of traffic flow by directly using current and near-past link travel speeds estimated for the target link and the correlated upstream/downstream links.

A disadvantage of the NN models with respect to PM is that they need to be recalibrated regularly. Conversely, the main issue in employing PM model is its

heavy computational burden in searching similarity in the past. To address this issue a suitable searching method for real-time application was implemented in JAVA programming language.

Model testing included the overall inner carriageway of the Rome Ring Road (33 links, 68.2 Km).

Both approaches exhibit robust prediction capabilities and their utilization into a real-time ATIS and ATMS application results proven. Particularly NN models show data noise and the investigation of different criteria to better identify the shape similarity of historical trends for the PM analysis.

References

Bajwa S.I., Chung E., Kuwahara M., (2005). "Performance evaluation of an adaptive travel time prediction model." In *Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems, Vienna, Austria*.

Bontempi G., Birattari M., Bersini H., (1999). "Lazy learning for local modeling and control design.", *International Journal of Control* ,72, 643-658.

Chung E. et al., (2003). "Travel time prediction: issues and benefits.". *In Proceedings of the 10th World Conference on Transport Research*.

Dayhoff J.E., (1990). "Neural network architectures: an introduction.". *Van Nostrand Reinhold Co., New York, NY, USA*.

Demuth H., Beale M., (2002). "The neural network toolbox for MATLAB.". *Mathworks Inc., Nattick, MA, USA*.

Duda R., Stork D., Hart P., (2000). "Pattern classification.". *Wiley-Interscience*.
Federal Highway Administration, (2001). "A roadmap for the research, development and deployment of traffic estimation and prediction systems for real-time and off-line applications". *Operations Research & Development*.

Gordon R., (2009). "Intelligent freeway transportation systems.". *Springer*.

Hellinga B., Izadpanah P., Takada Y., Fu L., (2008). "Decomposing travel times measured by probe-based traffic monitoring systems to individual road segments.". *Transportation Research* , 16, 768-782.

Lin H.E. et al., (2004). "Short-term arterial travel time prediction for advanced traveller information systems.". *Journal of Intelligent Transportation Systems*, 8, 143-154.

Lin H.E., Zito R., (2005). “A review of travel-time prediction in transport and logistics.”. *In Proceedings of the Eastern Asia Society for transportation studies,* 5, 1433-1448.

Ripley B.D., (1996). “Pattern recognition and neural networks.”. *Cambridge University Press.*

Turner S. M. et al., (1998). ”Travel time data collection handbook.”. *Federal Highway Administration, Texas Transportation Institute (Report Fhwa-PL- 035).*

Yang Z., Gong B., Lin C., (2009). ”Travel time estimate based on floating car.". *2th Conf. on Intelligent Computation Technology* ,3, 868-871.

Urban Short-Term Traffic Flow Forecasting Based on the Semi-Variable Cell Transmission Model

Xiaojian Hu[1], Wei Wang[2] and Jian Lu[3]

[1] Post Doctor, Mobiles Station for the Post-doctors, Southeast University, Si Pailou 2, Nanjing, Jiangsu, China 210096; Email: huxiaojian@seu.edu.cn
[2] Prof., [3]Prof., Transportation College, Southeast University, Si Pailou 2, Nanjing, Jiangsu, China 210096; Email: Wangwei@seu.edu.cn, lujian_1972@seu.edu.cn.

Abstract

The accuracy of the urban short-term traffic flow prediction directly affects the real-time and reliability of the urban traffic control and management. To improve the accuracy and usefulness of the urban short-term traffic flow prediction, the Semi-variable Cell Transmission Model (SCTM) is proposed based on the cell transmission model. According to the traffic flow conservation, the cell length and the cell density are introduced into the SCTM, which overcomes the limit that the cell length must be uniform, in order to ensure the applicability and usefulness of the urban short-term traffic flow prediction based on the SCTM. Furthermore, the SCTM limits the scope of the cell length and sets the time interval to meet the requirement of the uniformity of the traffic distribution in the cells to improve the accuracy of the flow prediction. Finally, the SCTM is used to predict the short-term traffic flow in the urban road, and the predicted data are compared with the actual data. The analysis results indicate that the short-term traffic flow forecasting by the SCTM is satisfactory.
Key Words: traffic management; traffic flow; traffic flow conservation; cell density; cell length

Introduction

The accuracy of urban short-term traffic flow forecasting directly affects the real-time and reliability of urban traffic control and management. The various methodologies and techniques of the urban short-term traffic forecasting have been reported in the literature, including time series models (William, 1999), kalman filter theory (Okutani, 1984), nonparametric regression model (Smith, 1996), chaotic

theory (Hu, 2003), markov chain model (Yu, 2003), discrete state propagation model (Guin, 2004), neural network model (Zheng 2006), support vector machine model (Wang 2004), gaussian maximum likelihood model (Lin, 2001), etc. These methods are generally classified into seven categories (Guo, 2005), one of which is "traffic flow theory based methods". Traffic flow theory that describes the law of traffic dynamics is natural to provide a forecasting method, including the simple first order continuum theory (Lighthil, 1955), the higher order continuum theory (Whitman, 1974), car-following model (Pipes, 1967), cellular automaton (Nagel, 1992), etc. Traffic flow theory based methods can be further divided into macroscopic models and microscopic models. Macroscopic models take the continuum traffic as the object of study, and traffic is usually described in terms of aggregate quantities such as density, flow, and space-mean speed, wherein individual vehicle behavior in car following is not modeled explicitly. Macroscopic models are good at predicting continuum traffic flow, but the result of predicting intermittent traffic flow is poor. In contrast, Microscopic models are distinguished by their specification of the dynamics of individual vehicles. Microscopic models are mainly built upon dynamics of individual vehicles and interactions between vehicles and their environment, wherein the motion of a vehicle depends on the proximity to the vehicle immediately downstream, reflecting the fact that drivers will usually act to maintain a safe spacing, and therefore microscopic models are better at predicting intermittent traffic flow. However, microscopic models are not suitable to predict lots of traffic flow at the large network, because the models need lots of operation and are difficult to timely correct the models' parameters.

To overcome these defects of macroscopic models and microscopic models, the cell transmission model (CTM) becomes a good candidate for constructing forecasting methods. The main idea of the CTM is to approximate the LWR model in a discrete-time, discrete-space fashion with a piecewise-liner fundamental (Daganzo, 1994, 1995). The whole network is represented by a collection of small road segments, called cells, each with homogeneous traffic characteristics. In general, each vehicle proceeds to the next downstream cell in each time interval when the cell is not congested. Although the CTM belongs to macroscopic models, the vehicle individuality has been well described by the cell. Consequently, the CTM has its inherent advantages of predicting continuum and intermittent traffic flow. The CTM has been used widely in dynamic traffic simulations (Laura 2004; Shang et al., 2007), signal control (Lan 2003; Zen, 2006), and dynamic traffic assignment (Lo and Szeto 2002; Lian et al., 2007).

However, the CTM has the own defect that the cell length must be uniform, which hinders the utility of the CTM from predicting traffic flow in the roads of different lengths. For improving the accuracy and usefulness of the urban short-term traffic flow forecasting, the Semi-variable Cell Transmission Model (SCTM) is proposed based on the CTM.

Method

Overview of the Cell Transmission Model. The CTM, proposed by Daganzo (1994), simplifies spatially and temporally the solution scheme of the LWR model. The road is divided into N cells of which the length is uniform, numbered in increasing order beginning at the upstream boundary. The number of vehicles in cell i in the t^{th} time interval, $n_i(t)$, evolves according to a conservation equation,

$$n_i(t+1) = n_i(t) + y_i(t) - y_{i+1}(t) \tag{1}$$

where $y_i(t)$ is the number of vehicles that flow from cell $i-1$ to i when the time index advances from t to $t+1$, (veh).

For all cells, the number of vehicles from cell $i-1$ to cell i is the minimum of three quantities in time index t:

$$y_i(t) = \min\left(n_{i-1}(t), Q_{\max}, (w/v)\left(N_i(t) - n_i(t)\right)\right) \tag{2}$$

where $n_i(t)$ is the number of vehicles that will exit cell i in free-flow conditions, (veh); $Q_{\max}$ is the maximum inflow flow during the time interval, (veh); w is the speed of the backward shock wave, (m/sec); v is the free-flow speed, (m/sec); $N_i(t)$ is the maximum number of vehicles that can be present in cell i, (veh).

The proposed model According to the principles of the CTM, the cell length must be uniform and not less than the distance traveled by a typical vehicle in the free stream velocity during the time interval. However, the length of the urban road is not uniform, and the urban road includes different type sections such as the ordinary section and the approach section (Figure. 1). A vehicle can change lanes in the ordinary section, but cannot change lanes in the approach section. Therefore, it is difficult to build the uniform length cells in the urban road.

The SCTM is proposed to overcome the limit that the cell length must be uniform, in order to ensure the applicability and usefulness of the SCTM in the urban road.

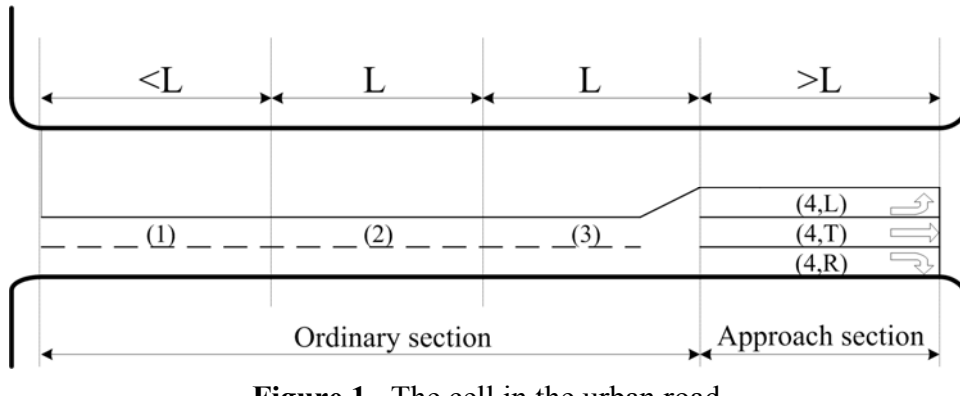


Figure.1. The cell in the urban road.

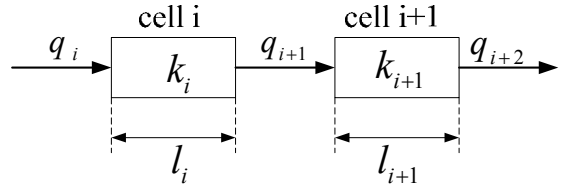


Figure.2. Cells of the SCTM.

In the SCTM, the length of the cell is not uniform, as shown in Figure.2.

Wherein, the traffic flow in the cell must satisfy the traffic flow conservation law, that is, the inflow volume minus the outflow volume is equal to the change in the number of vehicles in the cell (Wang, 2000).

$$\left[q_i(t) - q_{i+1}(t)\right] dT = \left[k_i(t+1) - k_i(t)\right] l_i \tag{3}$$

where, $q_i(t)$ is the traffic flow rate from cell $i-1$ to i when the time index advances from t to $t+1$, (veh/sec). $k_i(t)$ is the traffic flow density in cell i during the t^{th} time interval, (veh/m). dT is the time interval, (sec). l_i is the length of cell i and not less than the distance traveled by a typical vehicle in the free stream velocity during the time interval, (m).

Simplify Eq. (3):

$$k_i\left(t+1\right) = k_i\left(t\right) + \frac{dT}{l_i}\left[q_i(t) - q_{i+1}(t)\right] \tag{4}$$

Eq. (4) shows the traffic flow characteristics in the cell. Compare with Eq. (1), Eq. (4) is more complicated, but Eq. (4) introduces the cell length and the traffic flow density into the SCTM, which contributes to the application of the model in the urban road.

In the SCTM, the number of vehicles from cell $i-1$ to cell i is the minimum of two factors including the number of the outflow from cell $i-1$ and the number of the inflow to cell i.

Shown as Figure. 3, according to the principles of the CTM, during the time interval, the number of the outflow from cell $i-1$ of the SCTM is determined by the vehicle

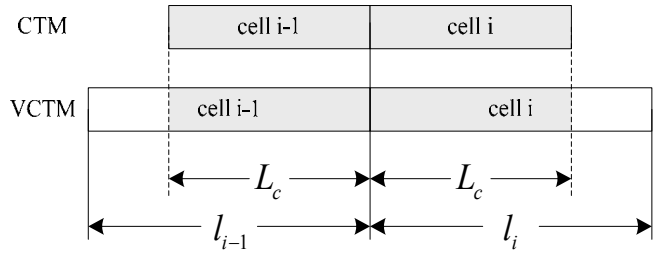


Figure.3. The cell of the CTM and the SCTM.

number in the gray section of cell $i-1$ and the maximum number of the traffic flow from cell $i-1$ to i. Therefore, the number of the outflow from cell $i-1$ of the SCTM ($S_{i-1}(t)$) is expressed as:

$$S_{i-1}(t) = \min\left\{\frac{L_c}{l_{i-1}} \cdot n_{i-1}(t), Q_i(t)\right\} \tag{5}$$

where, L_c is the distance traveled by a typical vehicle in the free stream velocity during the time interval, (m). $Q_{i-1}(t)$ is the maximum number of the traffic flow from cell $i-1$ to i during the t^{th} time interval, (veh).

The number of the inflow to cell i of the SCTM is determined by the vehicle number in the gray section of cell i and the maximum number of the traffic flow from cell $i-1$ to i during the t^{th} time interval. Therefore, the number of the inflow to cell i of the SCTM ($R_i(t)$) is expressed as:

$$R_i(t) = \min\left\{Q_i(t), \frac{\delta \cdot L_c}{l_i}\left[N_i(t) - n_i(t)\right]\right\} \tag{6}$$

where δ is equal to w/v.

According to Eqs. (5) and (6), the vehicle number of the traffic flow from cell $i-1$ to cell i is expressed as:

$$y_i(t) = \min\left\{\frac{L_c}{l_{i-1}} n_{i-1}(t), Q_i(t), \frac{\delta \cdot L_c}{l_i}\left[N_i(t) - n_i(t)\right]\right\} \tag{7}$$

where $y_i(t)$ is the vehicle number of the traffic flow from cell $i-1$ to cell i during the t^{th} time interval, (veh).

Divide both sides of Eq. (7) by the time interval dT:

$$\frac{y_i(t)}{dT} = \min\left\{\frac{L_c}{dT \cdot l_{i-1}} n_{i-1}(t), \frac{Q_i(t)}{dT}, \frac{\delta L_c}{dT \cdot l_i}\left[N_i(t) - n_i(t)\right]\right\} \tag{8}$$

Simplify Eq. (8):

$$q_i(t) = \min\left\{v\frac{n_{i-1}(t)}{l_{i-1}}, q_i^{\max}(t), \delta \cdot v\frac{N_i(t) - n_i(t)}{l_i}\right\} \tag{9}$$

where, v is the free stream velocity in the road, (m/sec). $q_i^{\max}(t)$ is the maximum traffic flow rate from cell $i-1$ to i when the time index advances from t to $t+1$, (veh/sec).

If the vehicle distribution is uniform in every cell, Eq. (9) is further simplified:

$$q_i(t) = \min\left\{v \cdot k_{i-1}(t), q_i^{\max}(t), \delta v\left[k_i^{jam}(t) - k_i(t)\right]\right\} \tag{10}$$

where, $k_i^{jam}(t)$ is the maximum density of the traffic flow in cell i during the t^{th} time interval, (veh/m).

Eqs. (4) and (10) of the SCTM describe the traffic transmission, queuing and evacuation process between cells. The SCTM doesn't only consider the relationship among traffic flow, velocity and density, but also satisfies the variability of the cell length. But the precondition of the reliability of the SCTM is that the vehicle distribution must be uniform in every cell.

The length of the cell is variability, but the change range of the length is limited. Firstly, the lower limit of the length is the distance traveled by a typical vehicle in the free stream velocity during the time interval. Secondly, the higher limit of the length is not unlimited, because if the length of the cell is excessively long, it is hardly ensured that the density of the cell is uniform, which directly affects the reliability of the SCTM. Thus, the length of the cell must satisfy Eq. (11):

$$L_c \le l_i \le 1.5L_c \tag{11}$$

If the length doesn't satisfy Eq. (11), the time interval dT will be adjusted in order to change the L_c. In additional, the adjustment of the time interval dT must consider the traffic signal in the upstream intersection. If the traffic signals changes, the traffic flow will be obviously changes in the road, which will affect the uniform density of the cell. Therefore, the time interval is the integral multiple of the smallest common divisor of the green time and the red time in the upstream intersection as far as possible.

According to Eq. (10), the traffic outflow rate $s_{i-1}(t)$ (veh/sec) and the traffic inflow rate $r_i\ (t)$ (veh/sec) of the cell are shown as Eqs. (12) and (13), respectively.

$$s_{i-1}(t) = \min\left\{vk_{i-1}\left(t\right), q_i^{\max}(t)\right\} \tag{12}$$

$$r_i\ (t) = \min\left\{q_i^{\max}(t), \delta v\left[k_i^{jam}\left(t\right) - k_i\left(t\right)\right]\right\} \tag{13}$$

In the merge connection, traffic flow rates from cell $\left(i-1,U\right)$ and cell $\left(i-1,D\right)$ to cell $\left(i\right)$ are shown as follows:

$$q_{i,U}(t) = \begin{cases} mid\left\{s_{i-1,U}(t), r_i\left(t\right) - s_{i-1,D}(t), p_{i-1,U}\left(t\right) r_i\left(t\right)\right\} & r_i(t) \le s_{i-1,U}\left(t\right) + s_{i-1,D}\left(t\right) \\ s_{i-1,U}(t) & r_i(t) > s_{i-1,U}\left(t\right) + s_{i-1,D}\left(t\right) \end{cases} \tag{14}$$

$$q_{i,D}(t) = \begin{cases} mid\left\{s_{i-1,D}(t), r_i\left(t\right) - s_{i-1,U}(t), p_{i-1,D}\left(t\right) r_i\left(t\right)\right\} & r_i(t) \le s_{i-1,U}\left(t\right) + s_{i-1,D}\left(t\right) \\ s_{i-1,D}(t) & r_i(t) > s_{i-1,U}\left(t\right) + s_{i-1,D}\left(t\right) \end{cases} \tag{15}$$

where, $q_{i,U}(t)$ (veh/sec) and $p_{i-1,U}$ are the traffic flow rate and the proportion of vehicles from cell $\left(i-1,U\right)$ to cell $\left(i\right)$,respectively. $q_{i,D}(t)$ (veh/sec) and $p_{i-1,D}$ are the traffic flow rate and the proportion of vehicles from cell $\left(i-1,D\right)$ to cell $\left(i\right)$,

respectively. $p_{i-1,U} + p_{i-1,D} = 1$. *mid* means to take the medium value of three quantities.

In the diverge connection, traffic flow rates from cell $i-1$ to cell (i,U) and cell (i,D) are shown as Eqs. (17) and (18):

$$q_{i-1}(t) = \min\left\{s_{i-1}(t), r_{i,U}(t)/\beta_{i,U}, r_{i,D}(t)/\beta_{i,D}\right\} \quad (16)$$

$$q'_{i,U}(t) = \beta_{i,U} q_{i-1}(t) \quad (17)$$

$$q'_{i,D}(t) = \beta_{i,D} q_{i-1}(t) \quad (18)$$

where, $q'_{i,U}(t)$ (veh/sec) and $\beta_{i,U}$ are the traffic flow rate and the proportion of vehicles from cell $(i-1)$ to cell (i,U), respectively. $q'_{i,D}(t)$ (veh/sec) and $\beta_{i,D}$ are the traffic flow rate and the proportion of vehicles from cell $(i-1)$ to cell (i,D), respectively. $\beta_{i,U} + \beta_{i,D} = 1$.

Computational simulation

Shown as Figure. 4, the length of the simulated road is 464m, including the ordinary section (356m) and the approach section (108m). The ordinary road from west to east direction has two lanes, and the approach section has three lanes. The ratio of left-turn vehicles at the downstream intersection is 0.3. The green time and the red time of the west phase in the intersections are 36sec and 30sec, respectively. Based on the principle of the time interval to identify, the time interval sets 6sec. The speed limit of the road is 50 km/h, so the lower limit of the length is 83m. Therefore, the road is divided 6 cells: cell (1), cell (2), cell (3) and cell (4) compose the ordinary section; cell (5,L) and cell (5,T) compose the approach section. The maximum density and the maximum flow rate of every cells except cell (5,L) are 0.4 veh/m and 1.2 veh/sec, respectively. The maximum density and the maximum flow rate of cell (5,L) are 0.2 veh/m and 0.6 veh/sec, respectively.

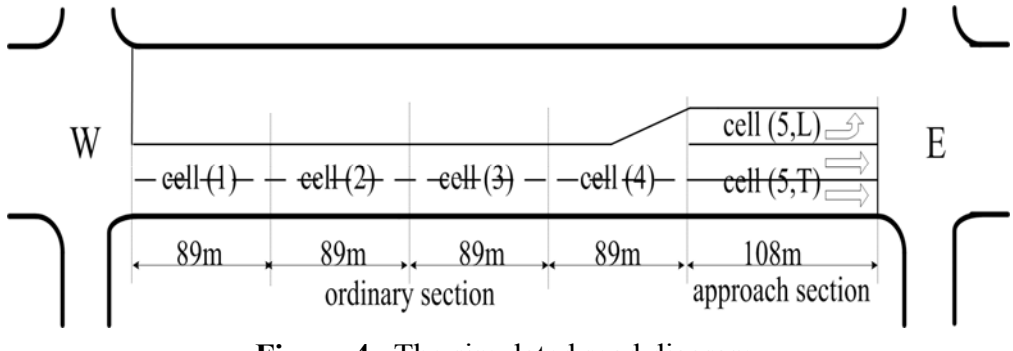


Figure.4. The simulated road diagram.

Table.1. The performances of the short-term traffic flow forecasting by the SCTM.

Traffic flow	600pcu/h			1000pcu/h			1400pcu/h		
Cell	(4)	(5,L)	(5,T)	(4)	(5,L)	(5,T)	(4)	(5,L)	(5,T)
Mean number of the actual vehicles	1.14	1.13	3.84	1.94	3.24	4.72	3.95	6.24	12.65
Mean number of the predicted vehicles	1.15	1.68	2.72	1.79	2.58	3.78	3.50	5.09	9.58
Mean error	0.01	0.55	-1.12	-0.15	-0.76	-0.94	-0.45	-1.16	-3.07

The traffic inflow rate of the road at the upstream intersection is simulated as 600veh/h, 1000veh/h and 1400veh/h, respectively. The SCTM forecasts the short-term traffic flow at every cell according to the mean traffic inflow rate during the time interval. Finally, compare the predicted data with the actual data to analyse the performances of the short-term traffic flow forecasting by the SCTM, shown in Table.1.

The SCTM predicts the downstream traffic flow by the traffic inflow rate of the road at the upstream intersection. The prediction time is the integral multiple of the time interval dT . In the computational simulation, the prediction time sets 18sec that means the traffic flow of cell (4), cell (5,L) and cell (5,T) is predicted by the SCTM.

Shown in Tab 1, the error of the traffic flow forecasting is relatively small, and the mean errors progressively larger with the increase in the traffic flow at the upstream intersection. Since the increase in the traffic flow induces the increase of the traffic

saturation which leads to the instability of the traffic flow in the road. Furthermore, the errors of the predicted numbers in the ordinary section (cell (4)) is generally less then the errors of the predicted numbers in the approach section (cell (5,L) and cell (5,T)). The main reasons are: firstly, the choice of the direction of the vehicle is randomness at the downstream intersection, so the ratio doesn't strictly meet 0.3 in fact during each time interval, but the traffic flow forecasting strictly satisfies the condition that the ratio is 0.3. Secondly, in the approach section, the vehicle is forbidden to change lanes, which is followed by the SCTM, but lane change actually happens fairly often. Thirdly, the influence of the control signal in the downstream intersection to cell (4) is less then the influence to cell (5,L) and cell (5,T).

On the whole, with the increase in the traffic flow at the upstream intersection, the accuracy of the traffic flow predicted at the downstream intersection decreases slightly, but the mean errors of the cells are generally small, which indicates the results of the short-term flow forecasting by the SCTM are satisfactory.

Conclusions

(1) Based on the traffic flow conservation law, the study introduces the cell length and the cell density into the SCTM to ensure that the length of all cells need not uniform and building cell is flexible. The short-time traffic flow forecasting by the

SCTM is applicable of kinds of urban roads, and helpful for decreasing the number of the cell to improve operational efficiency.

(2) The precondition of the SCTM is that the vehicle distribution must be uniform in every cell. The SCTM ensure the uniformities of the vehicle distribution in the cell by limiting the scope of the cell length and setting the time interval to improve the accuracy of the flow forecasting.

(3) The results of the simulation indicate that the SCTM can build the cell of different lengths for a complex urban traffic network structure, which improves the feasibility of the urban short-term traffic flow forecasting by the SCTM. The degree of fitness between the predicted data with the actual data is good to meet the request of the traffic flow forecasting in the urban road.

Acknowledgements

This work is partially supported by National Eleventh Five-Year Science and Technology Supporting Plan (2009BAG13A06), National Natural Science Foundation (50708019) , The Ministry of Education for New Century Excellent Talent Support Program(NCET-08-0115), Blue project in Jiangsu Province and Post-doctoral Research Funding Schemes of Jiangsu Province (No. 0902004B).

References

Daganzo, C. F. (1994). "The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory." *Transportation Research Part B*, 28(4), 269-287.

Daganzo, C. F. (1995). "The cell transmission model, part ii: network traffic." *Transportation Research Part B*, 29(2), 79-93.

Guin, A. (2004). "An incident detection algorithm based on a discrete state propagation model of traffic flow." *Ph.D. dissertation, Georgia Institute of Technology Atlanta, Georgia.*

Guo J. H. (2005). "Adaptive Estimation and Prediction of Vehicular Traffic Condition Series," *Ph.D. dissertation, North Carolina State University, Carolina.*

Hu, J., Zong C., and Song, J. etc. (2003). "An applicable short-term traffic flow forecasting method based on chaotic theory." *Proceedings of the 2003 IEEE Intelligent Transportation Systems,* 1, 608-613.

Lan C. J. (2004). "Developments of multi-criteria signal timing strategies for critical intersections[D]." *University of Miami*, 2003.

Laura M. M. (2004). "Macroscopic modeling and identification of freeway traffic flow." *University of California*.

Lian AP, Gao ZY, Long JC. (2007). "A dynamic user optimal assignment problem based on link variables of the cell transmission model." *Acta Automatica Sin*, 33(8), 852–859.

Lighthill, M. J., and Whitman, G. B., (1955). "On kinematic waves. ii. a theory of traffic flow on long crowded roads." *Proceedings of the Royal Society of London, Series A, Mathematical* and *Physical Sciences,* 229(1178), 317-345.

Lin, W. H. (2001). "A gaussian maximum likelihood formulation for short-term forecasting of traffic flow." *IEEE Intelligent Transportation Systems Proceeding, Oakland, California,* 150-155.

Lo HK, Szeto WY. (2002). "A cell-based variational inequality formulation of the dynamic useroptimal assignment problem," *Transp Res Part B,* 36(5):421–443.

Nagel, K., and Schreckenberg, M. (1992). "A cellular automaton model for freeway traffic," *J. Phys*, 2221–2229.

Okutani, I., and Stephanedes, Y. J. (1984). "Dynamic prediction of traffic volume through kalman filtering theory." *Transp. Res. B,* 18, 1–11.

Pipes, L. A. (1967). "Car-following models and the fundamental diagram of road." *Traffic Transportation Research*, 1(1): 21-29.

Shang H. Y., Huang H. J. Gao Z. Y. (2007). "Locating the variable message signs by cell transmission model," *Acta Phys Sin*, 56(8):4342–4347.

Smith, B. L. and Demetsky, M. J. (1996). "Multiple-interval freeway traffic flow forecasting." *Transp. Res. Rec*, 1554, 136–141.

Wang, J. S. Gao. B. C., and Shi. L.P. (2004). "Application of support vector machines in traffic volume forecast," *Information Technology* 28(4), 8–10.

Wang, W., Guo, X.. C., (2000). "Traffic engineering," *Nanjing: Southeast University Press.*

Whitman, G. B., (1974). "Linear and non-linear waves," *Wiley & Sons, NY.*

William, B. M. (1999). "Modeling and forecasting vehicular traffic flow as a seasonal stochastic time series process." *Doctoral dissertation. Department of Civil Engineering, University of Virginia, Charlottesville* .

Yu, G., Hu, J., and Zhang, C. etc. (2003). "Short-term traffic flow forecasting based on markov chain model." *Proceedings of The 2003 IEEE Intelligent Vehicles Symposium*, 208 – 212.

Zen J. Q., Wang J. J, etc. (2006). "CTM-RH based line control for urban traffic lights," *Journal of China University of Science and Technology*. 36(11):1232-1236.

Zheng, W. Z., Lee, D. H., and Shi, Q. X., (2006). "Short-term freeway traffic flow prediction: bayesian combined neural network approach." *J. Transp. Eng.* 132(2), 114–121.

Spatial and Temporal Aggregation Analysis for Space-Mean Speed Estimates Using Autoscope Video Detection Data

Zhongyi Wu[1,2], William H.K. Lam[2] and Mei Lam Tam[2]

[1] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan, China, 430079; PH: (852) 2766-4113; Fax: (852) 2334-6389; Email: wuzhyi@gmail.com
[2] Department of Civil and Structural Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

Abstract

Space-mean speed is an important indicator of road network performance in traffic operations and management. To collect travel time and speed data, video detectors or loop detectors are usually deployed. However, most detectors, such as Autoscope Video Detection System, measure spot speed or time-mean speed directly but not the space-mean speed over a section of roadway. An approximation method, which is commonly used to estimate space-mean speeds, is based on the average speed of vehicles traveling on a given segment of roadway during a specified period of time. However, such segment of roadway and specified period of time (refer to spatial and temporal aggregations) have significant effects on the estimation of space-mean speed. This paper investigates the effects of spatial and temporal aggregations of Autoscope data on the estimation of space-mean speeds with different length of measurement sections (5m, 10m, 15m and 20m) and measurement time intervals (20s, 30s, 60s and 120s). The results show that the space-mean speed can be better estimated by Autoscope data using 10m measurement section with a time interval of 120s.

Key Words: space-mean speed; time-mean speed; Autoscope video detection system; space aggregation interval; time aggregation interval

Introduction

Travel time, which is easily understood by transport professionals as well as the general public, is an important indicator of road network performance in traffic operations and management. Travel time estimates should be derived based on space-mean speed, which represents the average speed along a road segment. However, most detectors cannot measure space-mean speed directly but instead of time-mean speed which is the average speed of the traffic stream passing a specific stationary point. In order to estimate space-mean speeds, an approximation method is commonly adopted. The method is based on the average speed of vehicles traveling on a given segment of roadway during a specified period of time (Daganzo, 1997). However, such segment of roadway and specified period of time (refer to spatial and temporal aggregation) have significant effects on the estimation of space-mean speed. It is therefore essential to know how the spatial and the temporal aggregations affect the estimation of space-mean speed.

In the literature, most attentions have been given to develop the relationship between space-mean speed and time-mean speed. Wardrop (1952) derived their relationship and applied the results to estimate time-mean speed from space-mean speed. However, in most cases, time-mean speed but not space-mean speed is often available. It is thus desirable to estimate space-mean speed from time-mean speed. In view of this, Rakha and Zhang (2005) studied the differences between these two mean speeds and proposed a new formulation by making use of their variances for space-mean speed estimation.

In recent years, estimation of space-mean speed from loop detectors has been extensively studied (Dailey, 1999; Hazelton, 2004; Li, 2010). For example, Hall and Persaud (1988) developed a procedure for estimating space-mean speed using volume and occupancy measurements from single loop detector. Hellinga (2002) used the lengths of vehicles measured by an exponentially smoothed adjacent dual loop detector to enhance the speed estimation of single loop detectors.

However, little attention has been paid on the impact of spatial and temporal aggregations to the space-mean speed estimates. Knoop et al. (2009) used different aggregation times (10 and 20 seconds and 1, 2, 5 and 15 minutes) to examine the differences between space-mean speed and time-mean speed. The results show that longer aggregation times, the larger differences (space-mean speed / time-mean speed) and so the speed variation grows. However, their studies have not considered the effect of the length of measurement section (i.e. spatial aggregation) on the space-mean speed estimation.

This paper investigates the effects of spatial and temporal aggregations on the estimation of space-mean speeds using Autoscope video detection data. Autoscope Video Detection Systems (Michalopoulos, 1991) were set up at various locations in Hong Kong for collection of relevant traffic data. Based on the collected Autoscope data and video records, different length of measurement sections (5 m, 10 m, 15 m and 20 m) and measurement time intervals (20s, 30s, 60s and 120s) are analyzed in this study. The rest of the paper is organized as follows. The next section presents the data collected for this study and how to estimate space-mean speed and time-mean speed from these dataset. The methodology of spatial and temporal aggregation analysis for space-mean speed estimates and the results are presented in the third and fourth sections, respectively. Finally, conclusions are given together with recommendation for further study.

Data Collection and Extraction

Nine locations with different characteristics such as the number of lanes were chosen for data collection. The Autoscope Video Detection Systems were set up at these 9 locations along the major urban roads of Hong Kong Island North. The locations of the Autoscope video detectors are shown in Figure 1. The data collection surveys were conducted on several weekdays in May 2009 under various traffic conditions. During the surveys, both the traffic speed data and video records were captured by Autoscope video detectors.

There are two independent ways to obtain the individual vehicular speeds. One is recorded by Autoscope Video Detection System, whereas another is extracted from the video record manually assisting by a computer program. Based on these two extraction methods, the time-mean speed and the space-mean speed data utilized in the analysis were calculated using some approximation methods. The details of these approximation methods will be discussed in the next sub-sections.

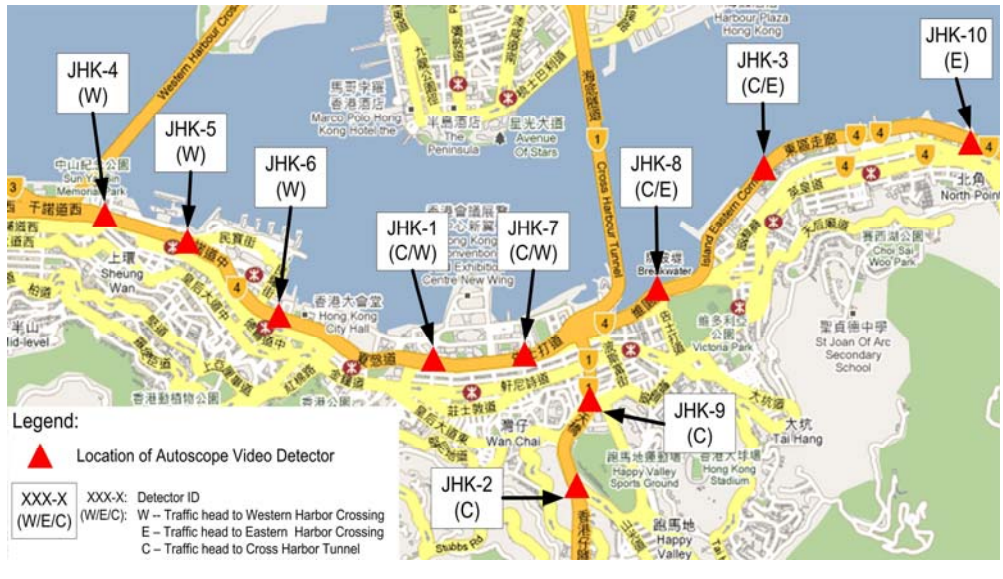


Figure 1. Locations of Autoscope video detectors.

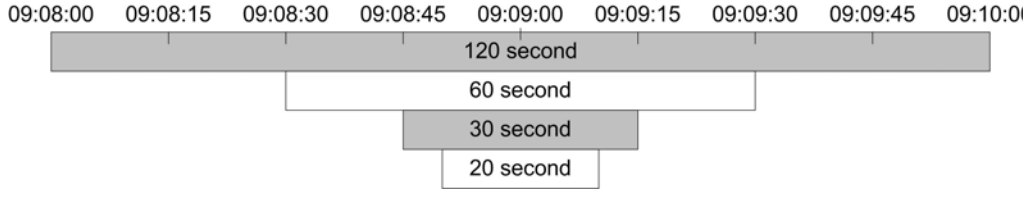


Figure 2. Example for temporal aggregation of space-mean speed measurements.

In order to minimize bias, each measurement time interval is centered as closely as possible on the same collection time period. For example, Figure 2 illustrates the space-mean speed measurements for different temporal aggregations for the time interval at 09:09:00 (hh:mm:ss).

Space-Mean Speed. Average traffic speed can be computed in two different ways: time-mean speed and space-mean speed. The difference in speed computation is attributed to the fact that the time-mean speed reflects the average speed of the traffic stream passing a specific stationary point, while the space-mean speed reflects the average speed over a spatial section of roadway. Daganzo (1997) demonstrated that the space-mean speed is a density weighted average speed, while the time-mean speed is a flow weighted average speed.

Space-mean speed is the average speed of all vehicles occupying a given section of highway or lane over some specified time periods. The average speed of a traffic stream computed as the length of roadway segment divided by the total time required to travel the segment. Therefore, in order to compute accurate travel time, space-mean speed is required. Moreover, space-mean speed is a traffic density speed estimate which reflects the spatial dimension of speed and thus is utilized in the standard speed-flow-density relationship. The space-mean speed can be calculated as follows:

$$u_s = \frac{n * d}{t_1 + t_2 + \cdots + t_n} = \frac{n * d}{\frac{d}{u_1} + \frac{d}{u_2} + \cdots \frac{d}{u_n}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{u_i}} \tag{1}$$

where u_i is the speed of vehicle i, d is the length of roadway segment, n is the vehicle count during a given time period and u_s is the space-mean speed.

Time-Mean Speed. On the other hand, time-mean speed is the average speed of all vehicles passing a particular point on a highway or lane over same specified time period. Time-mean speeds do not provide reasonable travel time estimates unless the speed of the point sampled is representative of the speed of all other points along a roadway segment, or unless there are a large number of closely-spaced detectors

along the segment. Time-mean speed is the arithmetic mean of the speeds of vehicles passing a point on a highway during an interval of time. It can be calculated by Eq. (2):

$$u_t = \frac{u_1 + u_2 + \cdots + u_n}{n} = \frac{1}{n}\sum_{i=1}^{n} u_i \tag{2}$$

where n is the number of vehicles observed during a given time period, u_i is the speed of vehicle i and u_i is the space-mean speed.

Methodology of Spatial and Temporal Aggregation Analysis

Sample Checking. In order to obtain more representative results, adequacy of sample size required for each spatial and temporal aggregation has been checked. Equation (3) is used to calculate the minimum number of samples required for each spatial and temporal aggregation:

$$N' = \left[\frac{z_{\alpha/2}\delta}{E}\right]^2 \tag{3}$$

where N' stands for the required minimum sample size, $z_{\alpha/2}$ is known as the critical value which equals to 1.64 at 10% level of significance ($\alpha = 0.1$), δ stands for the standard deviation of the speed data and E is the permitted error which 10% of the mean space-mean speed is adopted. If the observed sample size N is larger than the required sample size N', the sample size is enough.

Statistical Hypothesis Tests. The hypothesis test described in this section is intended to determine if there is a significant difference between space-mean speeds with different measurement intervals (for example, 20-second vs. 120-second temporal aggregation, 5-meter vs. 20-meter spatial aggregation). The hypothesis test is described below:

$$H_0: \left|u_s^i - u_s^j\right| = 0 \quad \text{km/h}$$

$$H_1: \left|u_s^i - u_s^j\right| > 0 \quad \text{km/h} \qquad (i \neq j)$$

where u_s^i and u_s^j is the space-mean speed with measurement interval i and j, respectively. The t value for the test is calculated by Equations (4) and (5):

$$t = \frac{\Delta\overline{\mu}_{ij}}{\delta_{\Delta\mu}/\sqrt{n}} \tag{4}$$

$$\Delta\overline{\mu}_{ij} = \frac{1}{n}\sum_{k=1}^{n}(\mu_{sk}^i - \mu_{sk}^j) \tag{5}$$

where $\Delta\overline{\mu}_{ij}$ is the average deviation of the space-mean speed with different measurement intervals, $\delta_{\Delta\mu}$ is the standard deviation of the space-mean speed deviation with different measurement intervals, n is the sample size.

At a 5% level of significance ($\alpha = 0.05$), when the test statistic is greater than $t_{0.05}(v)$, the null hypothesis H_0 is rejected, suggesting a significant difference in the space-mean speed between the time intervals or space intervals, where v is the degree of freedom ($v=n-1$). If $t < t_{0.05}(v)$, the null hypothesis cannot be rejected, and one can conclude that there is no significant difference in space-mean speed between the measurement intervals.

Results

As stated earlier, a large set of data was collected from Autoscope Video Detection System for analyzing the spatial and temporal aggregation effects. The results presented in this section illustrate the significant role that measurement interval plays in the estimation of space-mean speed.

Sample Size Checking. It can be seen in Table 1 that the collected data are adequate for all the different time and space aggregation intervals. It is noted that the mean and the standard deviation of the space-mean speed increase with decrease of the time aggregation interval. In another words, the deviation of the space-mean speed decreases with an increase of the time aggregation interval. Conversely, the mean and the standard deviation of the space-mean speed decrease with space aggregation interval.

Table 1. Adequacy of sample size.

Aggregation	Temporal aggregation				Spatial aggregation			
	120 s	60 s	30 s	20 s	20 m	15 m	10 m	5 m
Mean (km/h)	45.1	45.39	45.86	46.28	47.94	47.73	47.13	45.78
Std. Dev (km/h)	26.64	26.87	27.23	27.66	21.35	21.29	21.1	20.36
N' (required)	94	94	95	96	53	54	54	53
N (observed)	241	241	241	241	90	90	90	90
Status	Enough data	Enough data	Enough data	Enough data	Enough data	Enough data	Enough data	Enough data

* $\alpha/2 = 0.05$, $z_{\alpha/2} = 1.64$.

Table 2. Results of hypothesis tests on different spatial aggregations.

Spatial aggregation	Space-mean speed difference					
	15–20 m	10–20 m	5–20 m	10–15 m	5–15 m	5–10 m
Mean (km/h)	–0.21	–0.81	–2.16	–0.6	–1.95	–1.35
Std. Dev (km/h)	0.76	1.52	2.42	1.1	2.15	1.69
t-value	–2.61	–5	–8.4	–5.11	–8.52	–7.5
v=89, $t_{0.05}(89)$=1.99	Reject	Reject	Reject	Reject	Reject	Reject

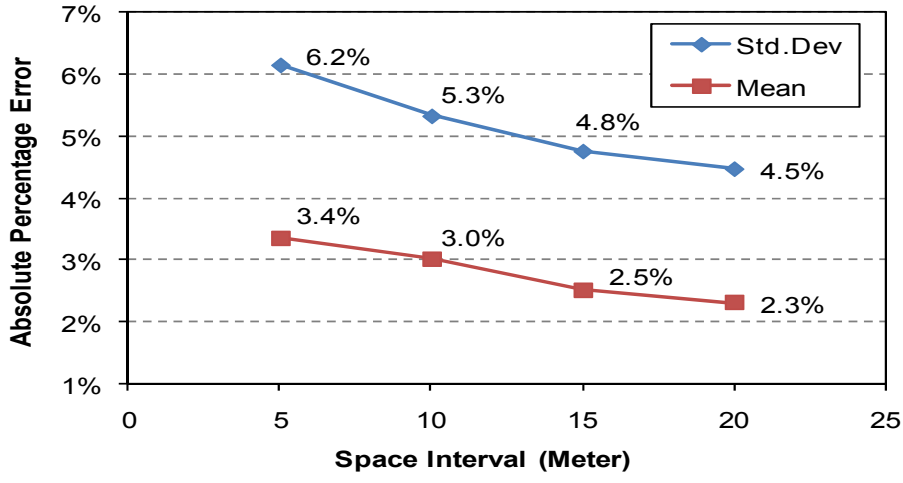


Figure 3. Absolute percentage errors between time-mean speeds and space-mean speeds with difference space intervals.

Spatial Aggregation Analysis. Table 2 presents the results of the hypothesis tests, which designed to determine if there is a significant difference between various spatial aggregations. As shown in Table 2, it can be found in all cases that the null hypothesis was rejected which indicates that the space-mean speeds between different spatial aggregations are significantly difference from each other. The absolute percentage errors between time-mean speeds and space-mean speeds with different spatial aggregations are shown in Figure 3. Both mean and standard deviation of the absolute percentage errors decrease with an increase in space aggregation interval.

Temporal Aggregation Analysis. The results of the hypothesis tests for space-mean speeds with different temporal aggregations are summarized in Table 3. From Table 3, it can be noted that the space-mean speed at time intervals of 60 s and 120 s is failed to reject the hypothesis, which means they are no any significant difference. The greatest deviation of space-mean speed is between measurement time interval of 20 s and 120 s. Figure 4 illustrates the differences in space-mean speed estimated from 120 s vs. 60 s and 120 s vs.20 s.

Table 3. Results of hypothesis tests on different temporal aggregations.

Temporal Aggregation	Space-mean speed difference					
	60–120 s	30–120 s	20–120 s	30–60 s	20–60 s	20–30 s
Mean (km/h)	0.29	0.77	1.18	0.47	0.89	0.42
Std. Dev (km/h)	2.67	4.6	5.54	3.19	4.33	2.6
t-value	1.75	2.64	3.4	2.36	3.28	2.56
v=240,$t_{0.05}$(240)=1.97	Fail to reject	Reject	Reject	Reject	Reject	Reject

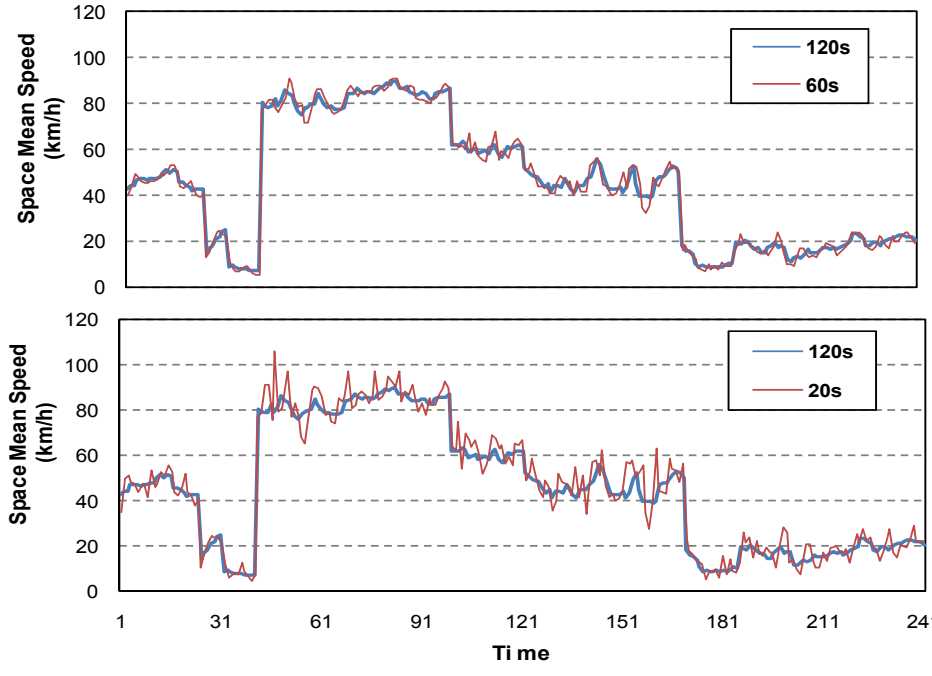


Figure 4. Space-mean speed with different time aggregation intervals.

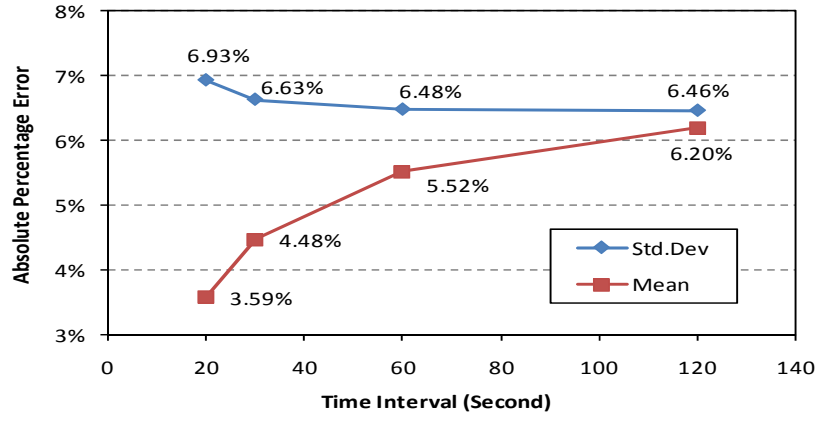


Figure 5. Absolute percentage errors between time-mean speeds and space-mean speeds with difference time intervals.

Figure 5 shows the absolute percentage errors between time-mean speeds and space-mean speeds with different measurement time intervals. The mean of absolute percentage errors increases with the increase of measurement time aggregation interval. However, the standard deviation of the absolute percentage errors decreases with the increase of time aggregation interval.

Space-Mean Speed Estimated by Autoscope. Autoscope Video Detection System measures individual vehicular speed (i.e. spot speed) directly. However, these spot speeds contain certain errors which may cause by light change or vehicle occlusion.

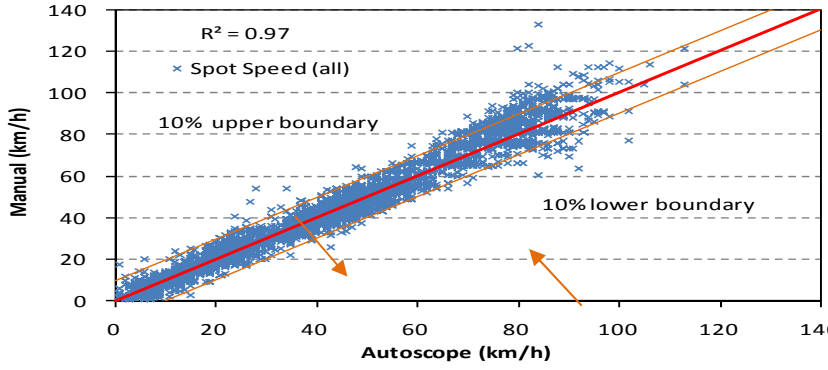


Figure 6. Comparison of spot speeds by manual measurement and Autoscope video detection system.

Table 4. Relative errors of space-mean speed estimated by Autoscope.

Aggregation		Temporal aggregation			
		120 s	60 s	30 s	20 s
Spatial aggregation	5 m	3.36%	5.02%	6.05%	7.43%
	10 m	3.18%	3.47%	3.82%	4.09%
	15 m	5.68%	6.17%	6.58%	7.69%
	20 m	6.49%	7.36%	10.03%	15.27%

To investigate the accuracy of Autoscope speed data, manual measurement from video records were conducted. Figure 6 shows the comparison results of the speed data obtained between manual measurement and Autoscope Video Detection System. It was found that there are 5.1% of spot speeds measured by Autoscope with the absolute error of more than 10 km/h.

Table 4 shows the relative errors of space-mean speed estimated by Autoscope when compared with manual measurement. It was found that the relative errors of space-mean speed decrease with an increase of measurement time interval. However, in the view of spatial aggregation, the relative error is the lowest when space interval equals to 10 m. The results show that the space-mean speed can be better estimated by Autoscope using 10 m measurement section with a time interval of 120 s.

Conclusions

In this paper, different length of measurement (5m, 10m, 15m and 20m) and (20s, 30s, 60s and 120s) time intervals were examined to assess the impacts of spatial and temporal aggregation of Autoscope video detection data on the estimation of space-mean speeds. Speed data were collected at nine selected locations in the urban road network of Hong Kong. The analysis results show that the space-mean speed can be better estimated by Autoscope data using 10m measurement section with a time interval of 120 s.

Further research will be carried out to investigate how to improve the estimation of space-mean speeds using different traffic data such as spot speed, time-mean speed and occupancy data provided by the Autoscope Video Detection System.

Acknowledgements

The work described in this paper was supported by a niche areas funding from the Research Committee of The Hong Kong Polytechnic University and the Autotoll Limited (Project No. J-BB7Q). The first author would like to thank The Hong Kong Polytechnic University for support to carry out his PhD research under the supervision of the second author in Hong Kong.

References

Daganzo, C. (1997). *Fundamentals of Transportation and Traffic Operations*, Elsevier Science.

Dailey, D.J. (1999). "A Statistical Algorithm for Estimating Speed from Single Loop Volume and Occupancy Measurements", *Transportation Research Part B*, 33(5), 313-322.

Hall, F.L and Persaud, G.B. (1988). "Estimating Speeds from Freeway Traffic Management System", *ITE 1988 Compendium of Technical Papers*, 166-171.

Hazelton, M. (2004). "Estimating Vehicle Speed from Traffic Count and Occupancy Data", *Journal of Data Science*, 2(3), 231-244.

Hellinga, B. (2002). "Improving Freeway Speed Estimates from Single-Loop Detectors", *Journal of Transportation Engineering*, 128(1), 58-67.

Knoop, V., Hoogendoorn, S.P., and Zuylen, H.V. (2009). "Empirical Differences between Time Mean Speed and Space Mean Speed", *Traffic and Granular Flow 07*, 351-356.

Li, B. (2010). "Bayesian Inference for Vehicle Speed and Vehicle Length Using Dual-loop Detector Data", *Transportation Research Part B*, 44(1), 108-119.

Michalopoulos, P.G. (1991). "Vehicle Detection Video through Image Processing: The Autoscope System", *IEEE Transactions on Vehicular Technology*, 21-29.

Rakha, H., and Zhang, W. (2005). "Estimating Traffic Stream Space-mean Speed and Reliability from Dual and Single Loop Detectors", *Proceedings of the 84th Transportation Research Board*, Washington D.C., USA [CD-ROM].

Wardrop, J.G. (1952). "Some Theoretical Aspects of Road Traffic Research", *Proceedings of the Institute of Civil Engineers*, 1-2, 325-378.

The Application of Support Vector Machines (SVM) for Traffic Condition Prediction Using ITS Data

Dazhi Sun[1], Qixing Wang[2] and Xinguo Jiang[3]

[1, 3] Southwest Jiaotong University, Chengdu, Email: sbz2000@swjtu.edu.cn
[2]Texas A&M University –Kingsville, Texas, U.S.A., Email: kfds000@tamuk.edu

Abstract

This paper introduced a pilot implementation of Support Vectors Machine (SVM), a new machine learning algorithm, for traffic condition prediction utilizing the data collected in traffic management center. A SVM^{python} based Level of Service Prediction System (LOS-PS) was developed in this study. ITS data collected from Transguide program in San Antonio, Texas were utilized to implement the proposed LOS-PS system in the case study of this paper.
Key Words: support vector machine; level of service prediction; data mining

Introduction

With the increasing installation and implementation of intelligent transportation systems (ITS) detection devices, the huge volume of available ITS data makes it difficult to process and utilize them efficiently for various operational or planning purposes with traditional analytical techniques. Various data analysis methods including Bayesian analysis and fuzzy cerebellar model etc. have been applied to deal with ITS data. For example, Li investigate the recursive estimation of vehicular speed which is a Bayesian analysis using the information provided by a single loop detector (Baibing Li, 2009). Yingyang Fuzzy Cerebellar Model was applied by Nguyen et al. for analyzing experimental ITS data on Pan Island Expressway (PIE) in Singapore to study its effectiveness in universal approximation and prediction(M.N Nguyen, Shi. D,C. and Quek, G.S.Ng). Min et al. investigated spatio-temporal correlation of ITS data for traffic condition prediction (Wanli Min, 2009).

Data mining integrates database management, machine learning and visualization technologies, for extracting hidden patterns from large quantities of data (Smith, B.L.et.al, 2002). Some previous traffic prediction efforts can be classified as

historical, database algorithms; time-series and spatio-temporal models; simulation models and neural network models. However with large quantities of data, these models may have difficulties processing the data in a timely manner. Support vector machines (SVMs), a new generation of learning system, used for classification and regression. Some advantages of using SVM include that it can be trained relatively easy and scaled reasonably to the high dimensional feature space etc. This paper is going introduce a new SVM implementation for predicting traffic condition, in terms of Level of Service (LOS) using the collected ITS data.

Support Vector Machine (Svm) Methodology

Statistical Learning Theory is established by Vapnik as a particular theory for limited sample size machine learning (Rakesh Agrawal.et.al., 2002; Radu Sion.et.al.,2002) Support Vector Machine was developed based on this theory, but for a new general-purpose learning. Support vector machines (SVM) are considered as one of the most robust and accurate methods among all well-known machine learning algorithms (Radu Sionet.al., 2002)

Generally speaking, the first step of learning process is choosing a non-linear mapp $\varphi(x)$, through mapping n-dimensional vector sample $(x_1, y_1),...,(x_i, y_i) \in R^n \times R$ to the feature space, built the optimal linear decision-making function $y = f(x) = w^T \varphi(x) + b$ in this high - dimensional space.. The SVM nonlinear function would transfer to following optimize problem with Vapnik methodology (Rakesh Agrawal.et.al., 2002; Radu Sion.et.al.,2002):

$$\min_{\omega,b,\xi} J(\omega,\xi) = \frac{1}{2}\omega^T\omega + c(\sum_{i=1}^{t}\xi_i + \sum_{i=1}^{t}\xi_i^*)$$

$$s.t: y_i - \omega^T\varphi(x_i) - b \le \varepsilon + \xi_i, i = 1,...,l$$

$$s.t: \omega^T\varphi(x_i) + b - y_i \le \varepsilon + \xi_i^*, i = 1,...l$$

$$\xi_i\xi_i^* \ge 0, i = 1,...l \qquad (1)$$

$\xi_i \in R$ the error variance. Especially $J(\omega,\xi)$ was composited by $\frac{1}{2}\omega^T$ and $\frac{1}{2}\sum_{k=1}^{N} e_k^2$, and γ is a constant which determines the importance of these two values.

Generally, since ω would be infinite-dimensional, directly calculating the function (1) would be extremely difficult, therefore we transfer this regulation problem into dual space.

$$L(\omega,b,e,a) = J(\omega,e) - \sum_{k=1}^{N}\alpha_k[\omega^T\varphi(x_k) + b + e_k - y \qquad (2)$$

$\alpha_k \in R$ is Lagrange multiplier. Yuan and Chen suggested the optimal solution requirement is as follows (Zhangchao Yuan and Chen Li, 2008):

$$\left\{\begin{array}{l}\dfrac{\partial L}{\partial \omega}=0\rightarrow \omega=\sum_{k=1}^{N}\alpha_k\varphi(x_k)\\ \dfrac{\partial L}{\partial b}=0\rightarrow \sum_{k=1}^{N}\alpha_k=0\\ \dfrac{\partial L}{\partial e_k}=0\rightarrow \alpha_k=\gamma e_k, k=1,...,N\\ \dfrac{\partial L}{\partial \alpha_k}=0\rightarrow y_k=\omega^T\omega(x_k)+b+e_k, k=1,...N\end{array}\right\} \quad (3)$$

Through the function (3), take off the ω and e_k, we can get:

$$\begin{bmatrix}0 & (q_N)^T\\ 1_N & \Omega+\dfrac{1}{\gamma}I\end{bmatrix}\begin{bmatrix}b\\ \alpha\end{bmatrix}=\begin{bmatrix}0\\ Y\end{bmatrix} \quad (4)$$

Especially the $\Omega=(\Omega_{ij})_{N\times N}=\left[\varphi(x_i)^T\varphi(x_j)\right]_{N\times N}$. Function (4) is a linear function, we can obtain α and b value. So the function $f(x)$ can transfer to:

$$f(x)=\sum_{i=1}^{t}\alpha_i K(x_i, x)+b \quad (5)$$

The kernel function K (x, Y) should satisfy the Mercer condition.

Real-Time LOS Prediction System (LOS-PS)

Prediction is commonly defined as using past and current observation data to estimate the future data. Actually, prediction assumes that there exist some functional relationships between past and future data. To some extent, the purpose of forecasting is trying to find this functional relationship based on characteristics of historical data and extend this relationship to future conditions. In other word, predication problem is essentially equivalent to approximation of function problem.

Level of service (LOS) is a performance measure to determine the traffic condition and quality of service. The Highway Capacity Manual (HCM) and AASHTO Geometric Design of Highways and Streets ("Green Book") (Level of service Explaination, 2009) define the LOS as follows:
As shown in the table 1, density is the critical factor to decide the LOS for basic segment of freeway. In this study, we use the ITS data collected from the Transguide program in San Antonio, Texas. Since their ITS data provides average speed (u) and flow (q) value, we can use $Density(k) = flow(q)/speed(u)$ to get the LOS of each road section every 5 minutes.

Table 1. Levels of Service (LOS).

LOS	Density (vphpm)
A (best)	0--11
B	12--18
C	19--26
D	27-35
E	36-45
F(worst)	46---

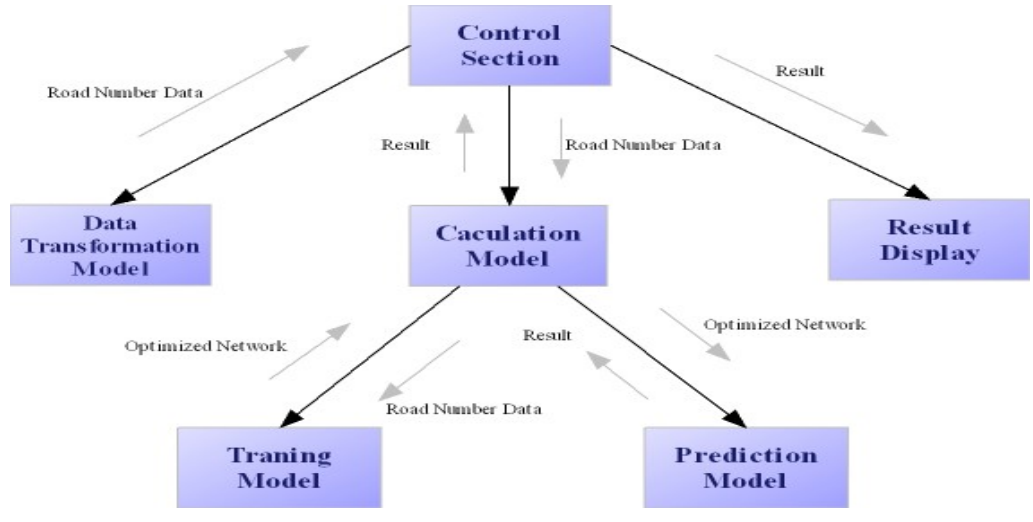


Figure. 1. Structure of LOS-PS.

On a certain road section, the current traffic condition is certainly correlated with the past condition; at same time, road section is a part of the road network, and traffic condition on each road segment is related with upstream and downstream traffic. Therefore the current traffic condition of a particular road segment is intrinsically link to the road network and its past traffic condition.

The LOS-PS was developed on ubuntu8.10 operating system with Python programming language [http://www.python.org/]. Python is an interactive, object-oriented, expandable programming language. There are many convenient open source libraries in Python, which greatly facilitated our implementation. The three modules of LOS-PS are shown in figure 1.

Preprocessing Module. This module serves the following two modules, which is the key component of LOS-PS. Preprocessing mainly check the format of raw data, remove the invalid record. This module was developed because the raw data downloaded from Transguide ITS website contains some invalid records, e.g. the records with "occupancy = 0.08" and "speed = zero" would be considered as invalid. Prior to the prediction process, those invalid data should be identified and reduced.

Data Transformation Module. Data transformation module mainly prepares training and testing data for SVM machine to learn prediction model. This module will perform the following functions: 1) remove unnecessary property (every column of record called property, which also can be called feature in SVM learning model) so we can learn simple, powerful, general prediction model, which will be more robust in prediction. 2) generate training and test data. The raw data is collected from road detectors every 20 second, which is too frequent to be pragmatical. So we will combine raw data record to generate a new record every 5 min. 3) reformat the data record so it can be input into LIBSVM tools.

Prediction Module. In this study, LOS-PS was written in Python programming language, based on the core learning algorithm of the SVM^{light} introduced by Thorsten Joachims (2002,2009). The following figure demonstrates the details of the learning module SVM ^{python} in SVM^{light}(Thorsten Joachims, 2009).

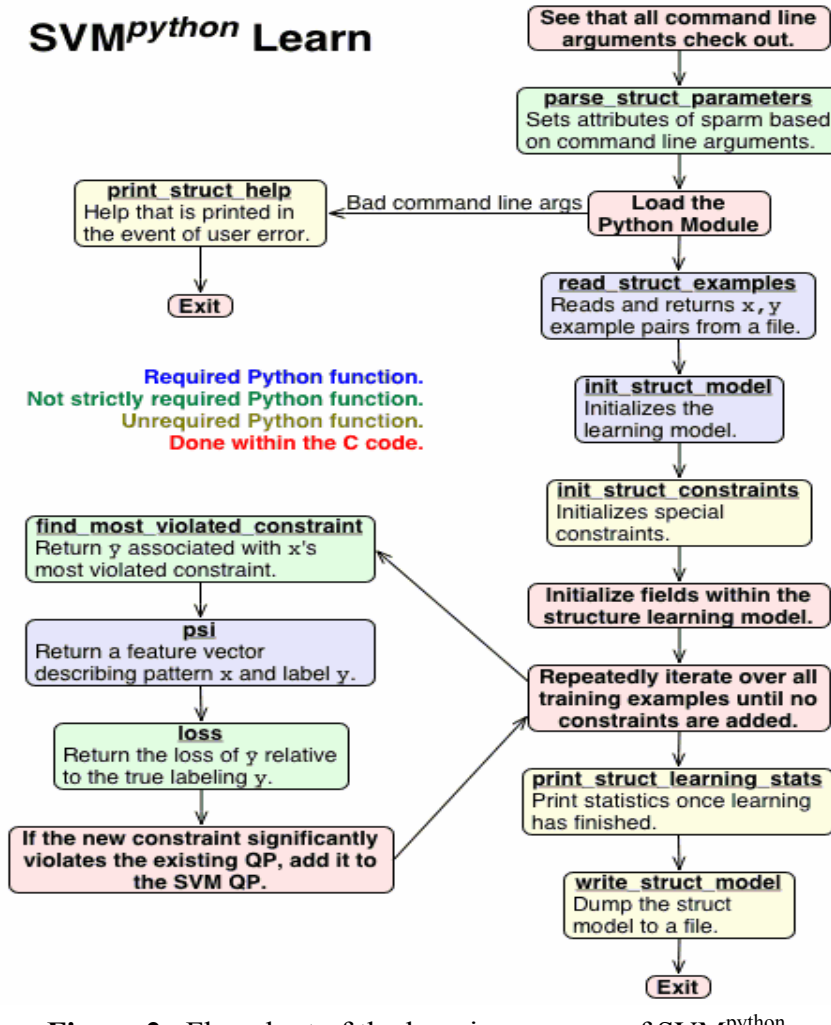


Figure 2. Flowchart of the learning process of SVM^{python} .

Firstly, prediction model will be built from training data set. We adjusted the training parameters and adopted ten-fold verification method to improve our model so at last we can get a stable model.

Secondly, once the stable prediction model was built, we can predict future LOS with some known property (precise time for example) of future (called input). The only thing we need to do is transform input with the same procedure with the ones done in training data generation.

Case Study

TransGuide of San Antonio, Texas, is an ITS program developed by TxDOT to provide traffic conditions information, such as travel time, accidents, congestion, and construction (http://www.transguide.dot.state.tx.us/docs/ atms_info.html,2009). The downloadable ITS data include traffic volume, speed, and occupancy information gathered from the instrumented urban freeways by loop detectors. The distribution of detectors is shown in the figure 3.

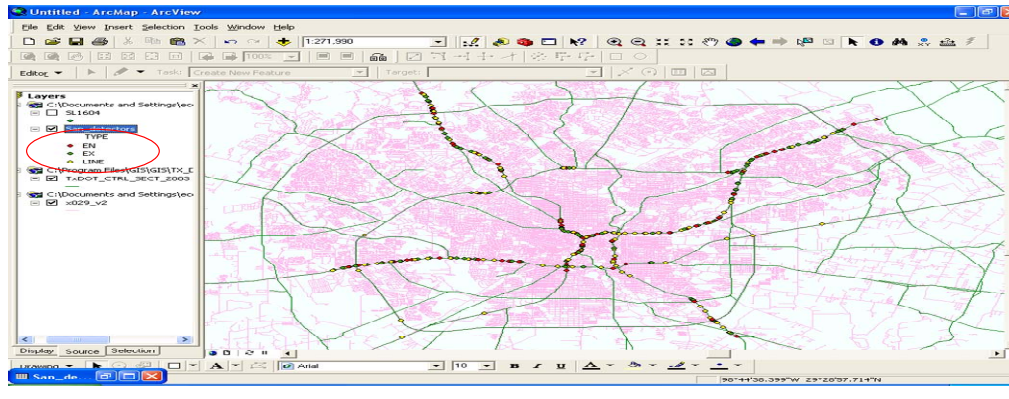


Figure 3. Loop Detector Distribution.

The following is the sample of the downloaded ITS data from Transguide website (http://www.transguide.dot.state.tx.us, 2009). The data was collected every 20sec, which 'Vol' represent the traffic volume of 20sec, 'Occ' represent the traffic occupancy of 20sec.

```
03/09/2009      0:00:31 EN1-0035N-152.804       Speed=60        Vol=006 Occ=004
03/09/2009      0:00:31 EN1-0035N-166.236       Speed=66        Vol=000 Occ=000
03/09/2009      0:00:31 EN1-0035N-166.900       Speed=57        Vol=002 Occ=001
03/09/2009      0:00:31 EX1-0035N-166.836       Speed=58        Vol=000 Occ=000
03/09/2009      0:00:31 EX2-0035N-166.836       Speed=00        Vol=000 Occ=000
03/09/2009      0:00:31 EX1-0035N-153.021       Speed=66        Vol=001 Occ=000
03/09/2009      0:00:31 L1-0035N-152.005        Speed=61        Vol=005 Occ=002
03/09/2009      0:00:31 L1-0035S-152.005        Speed=69        Vol=004 Occ=002
03/09/2009      0:00:31 L2-0035N-152.005        Speed=72        Vol=000 Occ=000
03/09/2009      0:00:31 L2-0035S-152.005        Speed=64        Vol=001 Occ=000
03/09/2009      0:00:31 L3-0035N-152.005        Speed=64        Vol=003 Occ=002
03/09/2009      0:00:31 L3-0035S-152.005        Speed=65        Vol=006 Occ=003
03/09/2009      0:00:31 EX1-0035S-151.952       Speed=55        Vol=002 Occ=002
03/09/2009      0:00:31 EN1-0035S-152.481       Speed=48        Vol=000 Occ=000
03/09/2009      0:00:31 L1-0035N-166.833        Speed=48        Vol=000 Occ=000
03/09/2009      0:00:31 L1-0035S-166.833        Speed=58        Vol=001 Occ=001
03/09/2009      0:00:31 L2-0035N-166.833        Speed=75        Vol=000 Occ=000
03/09/2009      0:00:31 L2-0035S-166.833        Speed=65        Vol=002 Occ=001
03/09/2009      0:00:31 L3-0035N-166.833        Speed=59        Vol=004 Occ=004
03/09/2009      0:00:31 L3-0035S-166.833        Speed=64        Vol=001 Occ=001
03/09/2009      0:00:31 EN1-0035S-153.017       Speed=51        Vol=000 Occ=000
03/09/2009      0:00:31 EX1-0035S-152.789       Speed=50        Vol=005 Occ=001
03/09/2009      0:00:31 L1-0035N-189.000        Speed=63        Vol=001 Occ=001
03/09/2009      0:00:31 L1-0035S-189.000        Speed=58        Vol=001 Occ=000
03/09/2009      0:00:31 L2-0035N-189.000        Speed=72        Vol=000 Occ=000
```

Table 2. Transformation data Table.

LOS	Day of the week	Time (5Min Int)	N-2 LOS	N-1 LOS
1	1:03	2:00	3:01	4:01
1	1:03	2:01	3:01	4:01
1	1:03	2:02	3:01	4:01
1	1:03	2:03	3:01	4:01
1	1:03	2:04	3:01	4:01
1	1:03	2:05	3:01	4:01
2	1:03	2:06	3:01	4:02
2	1:03	2:07	3:01	4:02
3	1:03	2:08	3:02	4:02
1	1:03	2:09	3:01	4:01

Before input these data into the Learn & Prediction model, we first use Transformation Module to train and test data for SVM machine to learn prediction model. Table 2 shows the data after data transformation.

In the case study, the ITS data from January, February, March of 2009 was used as training data to help the Real-time LOS prediction system 'study' the knowledge. Then we chose April 16th, 2009 as the prediction day. Fig 4. illustrates the prediction results comparing with observed value.

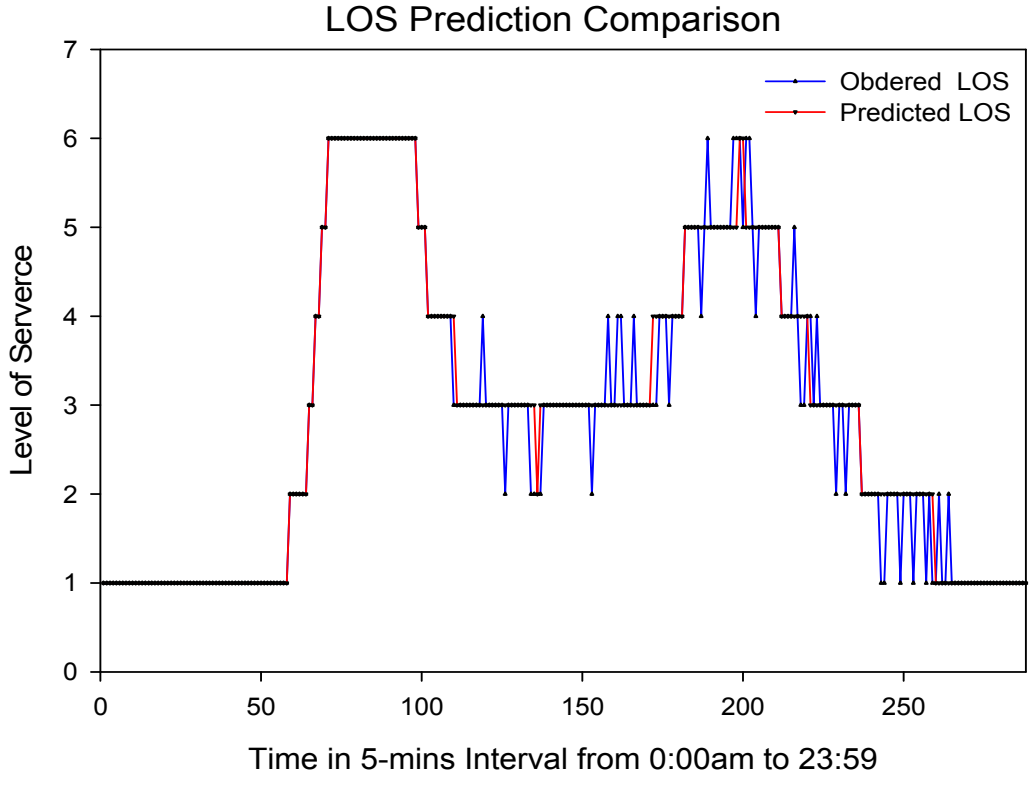


Figure 4. Traffic Condition Prediction Results.

The Level of Service numbers 1 to 7 represent the service level of A to F. The prediction interval is 5 minutes, so there are 288 data points. The comparison results shows 87.6% accuracy of prediction.

Conclusions and Future Work

The SVM implementation for traffic condition LOS prediction is introduced in this paper. The case study of using Transguide ITS data to predict LOS, showed that SVM method is applicable for mining ITS data for operational purpose. It is also found that this new algorithm can reduce the computational complexity, improve the learning speed and improve the prediction precision. For future research, weather condition, incident and highway construction information etc. would be included into the traffic condition prediction system.

References

Baibing Li, (2009).“On the recursive estimation of vehicular speed using data from a single inductance loop detector: a bayesian approach.” *Transportation Research Part B* 43 (2009) 391-402.

Level of service Explaination. http://en.wikipedia.org/wiki/ Level_of_service (accessed 05/09).

M.N Nguyen, Shi. D,C. Quek, G.S.Ng “Traffic prediction using yingyang fuzzy cerebellar model articulation controller.” ICPR’06.

Smith, B.L., Williams, B.M., and Oswalsd, R.K. (2002). “Comparison of parametric and nonpara-metric models for traffic flow forecasting.” *Transportation Research Part C*, 10(4), 303–321.

Rakesh Agrawal , Jerry Kiernan． Watermarking Relational Dat~ [C1． In：Proceedings of the 28th VLDB Conference，Hong Kong，China，2002.

Radu Sion. Mikhail Atallah， Sunil Prabhakar． Watermarking Relational Databases[R]．CERIAS TR 2002-28，20023．

Radu Sion． Mikhail Atallah． Sunil Prabhakar． *On Watermarking Nu—meric Sets*[R]． CERIAS TR 2001-60， Proceedings of the Workshop on Digital Watermarking 1W DW , 2002.

Smith, B.L., Williams, B.M., and Oswalsd, R.K. (2002). “Comparison of parametric and nonpara-metric models for traffic flow forecasting.” *Transportation Research Part C*, 10(4), 303–321.

TransGuide Official Website, http://www.transguide.dot.state.tx.us/docs/ atms_info.html (accessed 05/09)

Traffic Statics Data. http://www.transguide.dot.state.tx.us/docs/ statistics.html (accessed 03/09)

T. Joachims, *Optimizing Search Engines Using Clickthrough Data*, Proceedings of the ACM Conference on Knowledge Discovery and Data Mining (KDD), ACM, 2002.

Thorsten Joachims. "SVM^{light} support vector machine." http://svmlight.joachims.org/ (accessed 03/09)

Wanli Min, L. Wynter, Y. Amemiya "Road traffic prediction with Spatio-Temporal Correlations." http://transp-or2.epfl.ch/tristan/F ullPapers/ 010wynter.pdf (accessed 05/09)

Zhangchao Yuan, and Chen Li,(2008). "Improved algorithm of LS-SVM and its application to traffic flow prediction." *Journal of Kunming University of Science and Technology*, 33 (6).

Applying Locally Linear Embedding on Feature Extraction of Traffic Flow Data

Yenan Chen[1], Jianming Hu[2], Yi Zhang and Di Li

[1] Department of Automation, [2] Associate Professor, Tsinghua University, Haidian District, Beijing 100084, China, Email: chenyn06@mails.tsinghua.edu.cn, hujm@mail.tsinghua.edu.cn

Abstract

We encounter high dimensional data when using time series vector to describe the traffic flow of a certain link during some time period, or using flow data from different links to describe the traffic status of a region at a certain time point. This paper applies a dimensionality reduction method, named Locally Linear Embedding (LLE) to extract temporal and spatial features out of these high dimensional traffic flow data. LLE can visualize our data in a low dimension space, thus giving a vivid perspective on the emerging features. According to these features, we can put links into different clusters and better interpret the evolution of traffic patterns. Furthermore, comparison between linear dimensionality reduction method, PCA and LLE is carried out. The result shows that LLE has better performance.

Key Words: feature extraction; locally linear embedding; nonlinear dimensionality reduction; clustering; traffic status

Introduction

The massive traffic flow data are always intrinsically high-dimensional. When describing the traffic status of a certain link during a period of time, we use a time series of traffic flow, which is a high dimensional vector consisting of flow records. In another case, we describe the traffic status of a certain region by collecting the flow records of each links of that region, which forms a high dimensional vector, too. By inspecting these high dimensional data, we can learn little about the pattern of the link or the traffic status of a region. Only through feature extraction, we can get

useful information, which will facilitate our later research work on traffic flow prediction and traffic pattern evolution.

Acknowledgement: The work described in the paper is partially supported by National Natural Science Foundation of China (NSFC)50708054,(NSFC)60834001, Hi-Tech Research and Development Program of China (863Project) 2007AA11Z222, National Basic Research Program of China (973Project) 2006CB705506 and National Key Technology Re-search and Development Program 2006BAJ18B02.

Lin Hangfei extracted traffic characteristics from historical data by analyzing the framework of the information from four perspectives, i.e. space, time, satistics and parameter (Lin et al., 2007). Samant and Adeli used discrete wavelet transform (DWT) and linear discriminant analysis (LDA) to extract incident-related features from traffic patterns. DWT is applied to raw data and LDA is employed to reduce dimensionality. The result is then used as input for a neural network model to detect traffic incidents. Their feature extraction method significantly reduced the network size (Samant and Adeli, 2000). Boris used two models FOTO (forecasting of traffic objects) and ASDA (Automatische Staudynamikanalyse: Automatic Tracking of Moving Traffic Jams) for automatic recognition and tracking of congested spatial-temporal traffic flow patterns on freeways (Kerner, et al. 2004). In Lawrence Lan's paper, their group explored the features of traffic dynamics by mapping the one-dimensional traffic series into multidimensional spaces by Taken's algorithm (Lan et al., 2008). Guo's group extracted the eigenvectors that describe the satus of intersections using K-means and EM methods (Guo et al., 2005). Wang implemented PCA, rPCA, Kpca on urban traffic data. A transform matrix was designed by selecting eigenvalues according to decreasing magnitude of eigenvectors. These methods had extracted typical features including periodicity (Wang et al., 2009).

In this paper, we use Locally Linear Embedding (LLE), a nonlinear dimensionality reduction method to extract features out of traffic flow data.

The chief topic here is dimensionality reduction. Traditional techniques include PCA and MDS. There are also advanced approaches that nonlinearly reduced dimensionality by globally optimal solutions (Joshua et al., 2000). LLE is an unsupervised learning algorithm that computes low-dimensional, neighborhood-preserving embedding of high-dimensional input (Sam et al., 2000). LLE has been employed in a variety of feature extraction applications such as face recognition (Liu et al., 2008), ear recognition (Xu et al., 2008), speech (Xu et al., 2008) and music (Jain et al., 2004) visualization, and aircraft pose recognition (Yuan et al., 2009). In addition, LLE was implemented in hand gesture recognition and tracking (Ge et al., 2006).

This paper is organized as follows: Section II presents the introduction of LLE's principle and algorithm; Section III gives brief view of the collection of data; Section IV presents the result of feature extraction; Section V analyzes the features we got; Section VI summarizes the whole paper.

LLE Introduction

Suppose the data is consisted of N vectors X_i, each of dimensionality D. We expect each data point and its neighbors to lie on or close to a locally linear patch of the manifold. We choose K nearest neighbors per data point and reconstruct each data point by linear coefficients of it neighbors. The reconstruction errors are measured by cost function:

$$\varepsilon(W) = \sum_i \left| \vec{X}_i - \sum_j W_{ij} \vec{X}_j \right|^2 \tag{1}$$

$\vec{X}_j$ are the K nearest neighbors of $\vec{X}_i$. The weights W_{ij} summarize the contribution of the jth data point to the ith reconstruction. To compute W_{ij}, we have to minimize the cost function. There are two constraints here: (1) If $\vec{X}_j$ is not among the K nearest neighbors of $\vec{X}_i$, $W_{ij} = 0$. (2) $\sum_j W_{ij} = 1$. The optimal weights are found by solving a least squares problem. Under the 2 constraints above, we can write the cost function as:

$$\varepsilon = \left| \vec{x} - \sum_j w_j \vec{\eta}_j \right|^2 = \left| \sum_j w_j (\vec{x} - \vec{\eta}_j) \right|^2 = \sum_{jk} w_j w_k C_{jk}$$

$$C_{jk} = (\vec{x} - \vec{\eta}_j) \cdot (\vec{x} - \vec{\eta}_k)$$

We use a Lagrange multiplier to enforce the constraint that $\sum_j W_{ij} = 1$ and make w_j the variable. By solving the N equations, we have:

$$w_j = \frac{\sum_k C_{jk}^{-1}}{\sum_{lm} C_{lm}^{-1}}$$

The constrained weights that minimize these reconstruction errors obey an important symmetry: for any particular data point, they are invariant to rotations, rescaling, and translations of that data point and its neighbors. Because the construction weights reflect the intrinsic properties of data that are invariant to such transformations, we therefore expect the weights W_{ij}, that reconstruct the ith data point in D dimensions should also reconstruct its embedded manifold coordinates in d ($d<D$) dimensions. This is done by choosing d-dimensional coordinates $\vec{Y}_i$ to minimize the embedding cost function:

$$\Phi(Y) = \sum_i \left| \vec{Y}_i - \sum_j W_{ij} \vec{Y}_j \right|^2$$

We fix the weights W_{ij} while optimizing the coordinates $\vec{Y}_i$. It can be minimized by solving sparse $N \times N$ eigenvector problem. Cost function defines a quadratic form,

$$\Phi(Y) = \sum_{ij} M_{ij} (\vec{Y}_i \cdot \vec{Y}_j)$$

$$M_{ij} = \delta_{ij} - W_{ij} - W_{ji} + \sum_k W_{ki} W_{kj}$$

where δ_{ij} is 1 if $i = j$ and 0 otherwise.

Coordinates $\vec{Y}_i$ can be translated by a constant displacement without affecting the cost. Thus, we remove one degree of freedom by requiring the coordinates to be centered on the origin, and we have:

$$\sum_i \vec{Y}_i = \vec{0}$$

Also, to avoid degenerate solutions, we constrain embedding vectors to have unit covariance, with outer products that satisfy

$$\frac{1}{N} \sum_i \vec{Y}_i \vec{Y}_i^T = I$$

According to Rayleitz-Ritz theorem (Horn, 1990), the optimal embedding is found by computing the bottom d+1 eigenvectors of the matrix, M. We discard the bottom eigenvector which represents a free translation mode of eigenvalue zero. This enforces the constraint that the embedding have zero mean. So the remaining d eigenvectors form the d embedding coordinates found by LLE. Moreover, the matrix M can be stored and manipulated as the sparse symmetric matrix.

$$M = (I - W)^T (I - W)$$

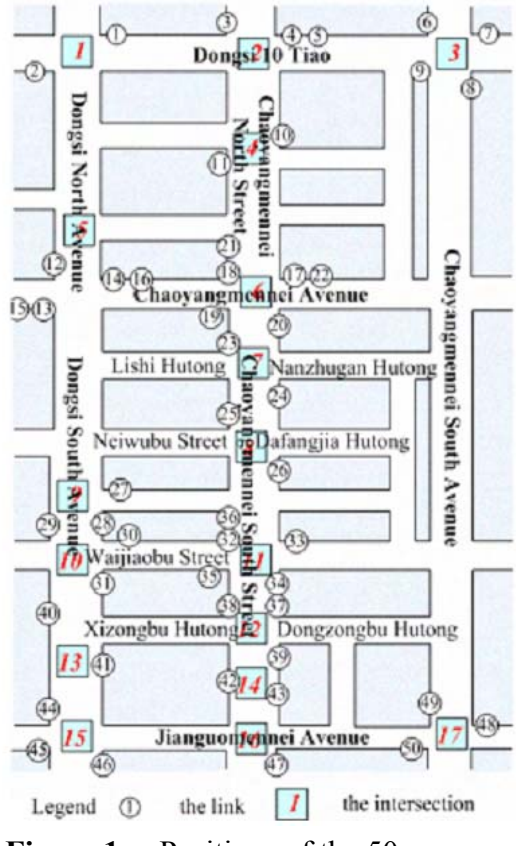
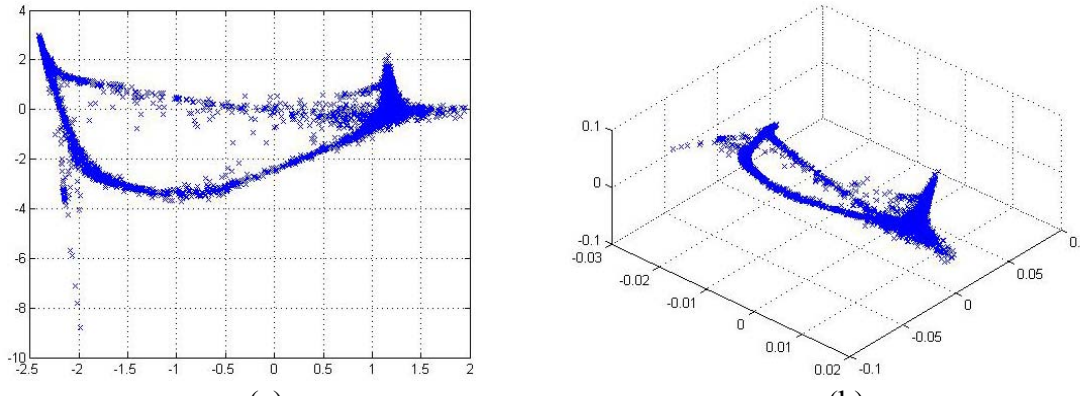


Figure 1. Positions of the 50 sensors.

Table 1. The acquisition time of data

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Day	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mon	Tue	Wed	Thu	Fri	Sat	Sun



(a) (b)

Figure 2. (a) is the result of reducing vectors to 2-dimensional space, and (b) is to 3-dimensional space.

Data Collection

The traffic flow data used in this paper is collected from 50 sensors in the region of Dongcheng District, covering a period of 16 days from Oct. 22, 2006 to Nov. 6, 2006. The data is provided by Beijing Traffic Management Bureau. The sampling interval is 5 min. That means 288 samples a day and 4608 samples for 16 days each sensor. The flow data is store in a matrix of 50×4608. Figure 1 indicates the area of data collected and the positions of all 50 sensors. Table 1 shows the acquisition time of data from each day.

Experimental Results

LLE must be given the K nearest neighbors of each data point and the performance of LLE depends on the value of K. There is still no theoretical rule on how to decide K. Most of the time, K is decided by experiment (Lawrence et al., 2000). However, there is a proposed method that utilizes adaptive neighbors of every data point (Huang et al. 2009). In this section, we tried K=12, 13, ⋯, 20 and we always report the result using the best value of K.

Temporal Feature. We extract time features by applying LLE to 4608 samples of 100-dimension vectors. Through this method, the traffic status of every time point can be visualized on 2-dimensional or 3-dimensional space. In this paper, we use the word status points to call these visualized traffic statuses on low dimensional space. Figure 2 shows the result of dimensional reduction (the 100-dimension vectors

describing the traffic status of a link are reduced to a low-dimension space). Figure 3 plots the status points within one day on a separate sub-figure.

Spatial Feature. Apply LLE to 50 4608-dimension vectors, each representing the traffic flow of a certain link over 16 day period. Result is shown in Figure 4.

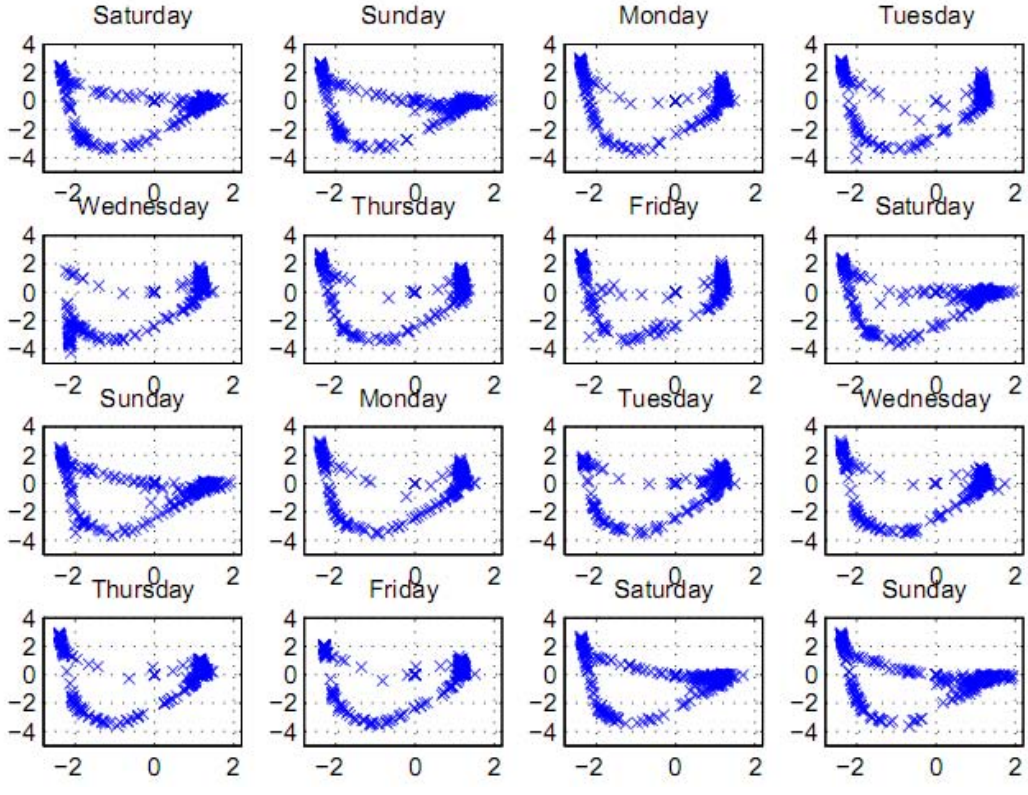


Figure 3. Each sub-figure shows the 288 status points within one day.

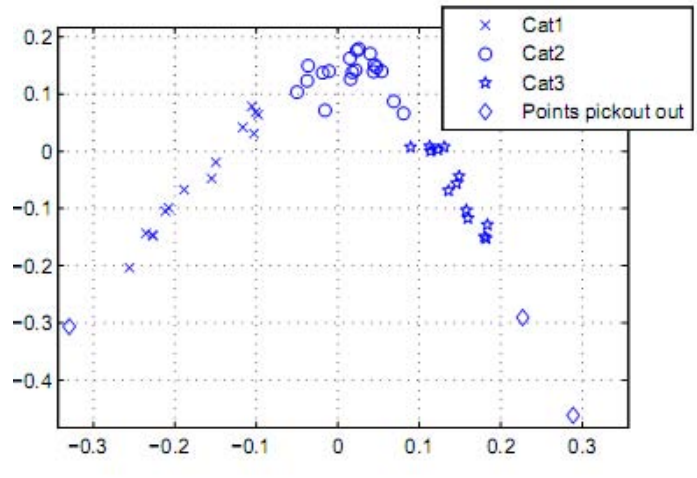


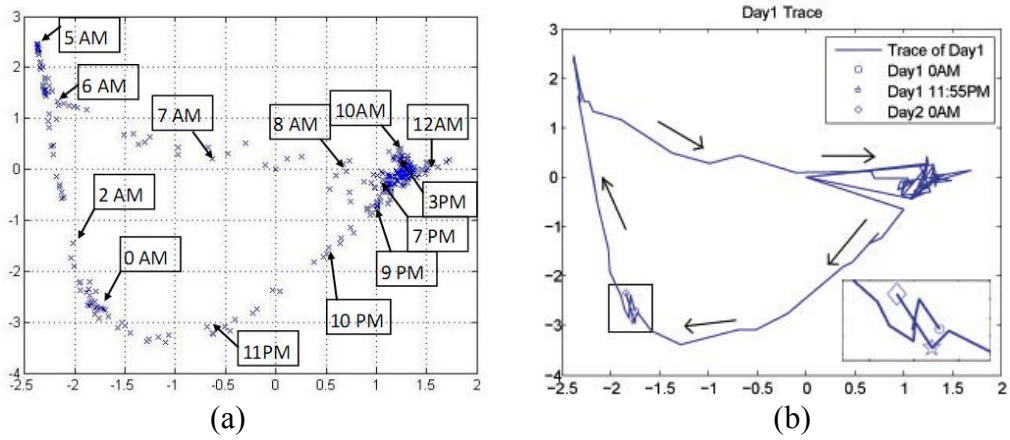
Figure 4. Each marker represents a sensor.

Analysis of Features

Analysis of Spatial Features. Because LLE preserves the topological feature of the original manifold, on the visualized space, points close to each other are still spatially close in the original space. According to this feature of LLE, we can cluster the links into different categories. First, we pick out the 3 points that are comparatively far away from the others. Then put the rest into 3 categories in a way shown in Figure 4. Because the links and the sensors are one to one, this result can also be the clustering result of links.

Analysis of Temporal Features. After dimensionality reduction, we can view each point in low-dimension space as representing the traffic status of the region covered by the 50 sensors at the time point. To have a general idea about how the status point moves with time, we can refer to Figure 5(a), which shows the time of some status points on day 1. Through Figure 5(a), we see that most status points during daytime, from about 10 am to 7 pm, gather densely. In Figure 5, we indicate the time of some status points on small labels with arrows pointing to the status points. The line indicates the trace of the status point on day 1. Small arrows suggest the direction of the trace. The figure embedded on right bottom shows details of the framed area after enlargement. Three different markers represent three specific time point. So the trace of status points evolving during daytime is not as clear as that during night time. If the sampling time longer and draw a line showing the trace, the evolving process of status points is clearer. On Figure 5(b), the sampling time is 3 times longer.

Through studying the low-dimension result, we summarized several features. (1) Periodicity. From Figure 3, the periodicity of 24 hour is very obvious. This can be explained by people's usual behavior. (2) Differences between Weekend and Workdays. First, status points of daytime on weekdays gather at a region different from that of weekends. Based on Figure 3, the 2-dimension figure, the different distribution regions are as listed below in Table 2.



(a) (b)

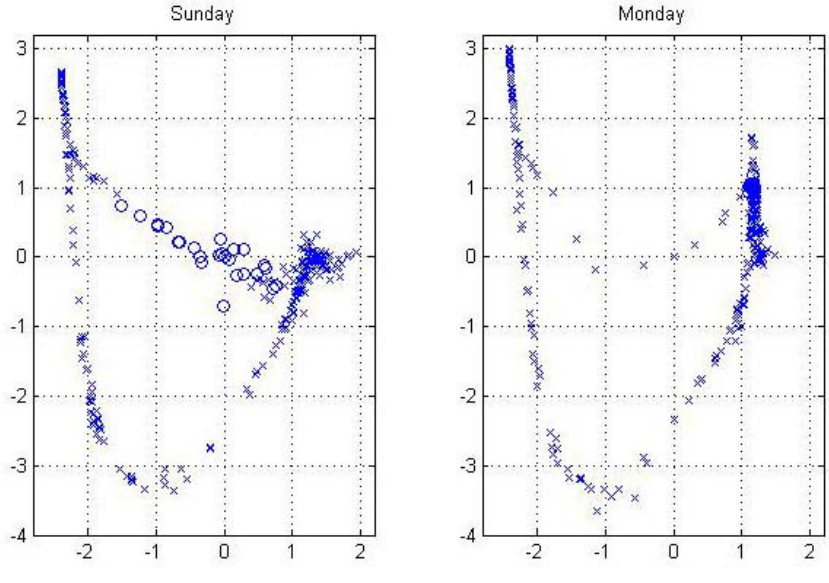
Figure 5. Traffic status of the region covered by the 50 sensors at the time point.

Table 2. Distribution regions of weekday status points and weekend status points.

	Daytime status points of weekdays	Daytime status points of weekends
x-coordinate	[0, 2]	[–1, –1]
y-coordinate	[1, 1.5]	[1, 2]

From Figure 2 (b), the 3-dimension figure, the difference is also apparent. It is easy to understand that people have distinct behavior on workdays and weekend. Weekend daytime status points scatter in a bigger scope than that of workdays. That is because traffic pattern is more regular on workdays. But on weekends, people go to anywhere and traffic patterns vary greatly.

Second, the traffic status leading to morning peak goes distinct ways on workdays and weekends. We can have a close look at Figure 6, which is part of Figure 3. Figure 6 shows the figure of Day 2, Sunday, and Day 3, Monday. The circle markers are status points of morning peak, defined as 7:00–9:00 AM. In Figure 6, the circle markers represent the status points between 7 AM and 9 AM. This time period is defined as morning peak hour in this paper. On Figure 6(a), which depicts traffic status on a weekend day, the traffic status evolves from wee hours to daytime through a steadily changing process. But on Figure 6(b), which represents a workday, the status does not have such an obvious transition. The transition process is rather short and status point quickly merges into the mass status points of daytime. Furthermore, the traces are very different, on workdays, the trace before morning has a concave and goes up, while the trace of weekend is rather straight.



(a) (b)

Figure 6. The status points of day 2 and day 3.

Those distinctions reflect people's different behavior on workdays and weekends. On weekends, the traffic flow grows slowly in the morning, but on workdays, people go to work on comparatively the same time range so there are concentrated outgoings that pump up the traffic load in a very short time. That explains the sudden transition of traffic status on workdays.

Detection of Abnormal Traffic Status. Because of the periodicity, when we plot all the 4608 status points, they tend to gather around a fixed trace. Here we can view it as a typical trace of the region during one day.

Take another look at Figure 3 and we will find some traffic status points deviating from the usual trace. It means that some unusual events were happening, which caused these abnormal traffic status. After analyzing the 16-day figures in 2-dimension, we have found 3 abnormal events. They happened respectively during the time range 2:25 AM-3:15 AM of Day 5, 11:00 PM–11:40 PM of Day 7, and 2:30 AM-5:40 AM of Day 14. Here we detail and analyze the detection of the abnormal event on Day 5 as an example.

On Figure 7, blue markers show all the status points of 16 days. The magenta markers show only status points of day 5, which show all the 4608 status points and the magenta-cross markers show status points of Day 5 (Wednesday) only. We discover that the status point trace departs the usual trace and goes to a totally opposite direction at 2:25 AM and continues until 3:15 AM. After 3:15 AM, it goes back to the normal track.

We then analyze the original flow and occupancy data during that time. Sensor No.1 and No.46 have abnormally higher flow and occupancy rate than the average, and sensor 41 has especially lower flow and occupancy rate than the average. Figure 8 gives detail on the unusual data detected by these three sensors.

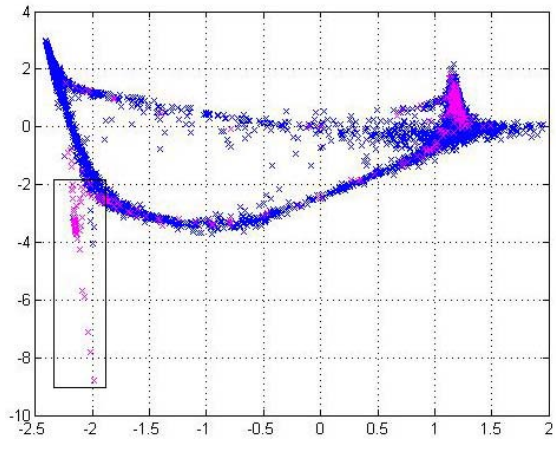


Figure 7. The status points that deviate the typical trace.

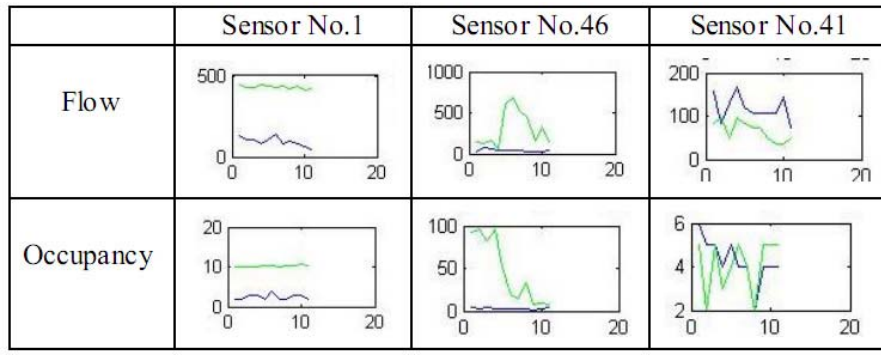


Figure 8. Green line indicates the traffic flow or occupancy rate of the sensor during 2:25 AM and 3:15 AM on day 5. Blue line indicates the average traffic flow o occupancy rate of the sensor during 2:25 AM and 3:15 AM.

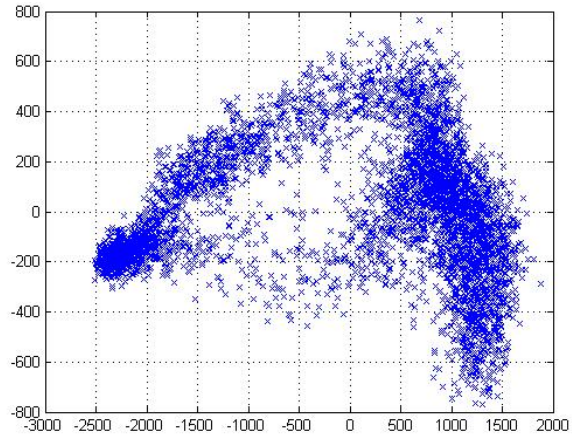


Figure 9. Result of using PCA to reduce dimensionality.

Comparison between PCA and LLE. Figure 9 is yielded by using PCA to reduce dimensionality. We can compare this Figure with Figure 3. It is obvious that the result by PCA did not extract a typical pattern out of massive traffic data. Though we can also find the feature of periodicity in PCA's result, little further information can be provided. So LLE is a better method on feature extraction of traffic data.

The reason may lie in the intrinsic nonlinearity of the traffic flow data itself. LLE, a nonlinear method, performs better than linear method like PCA towards nonlinear dataset.

Conclusion

In this paper, we applied Locally Linear Embedding to extract features out of a group of traffic flow and occupancy data. LLE is a nonlinear dimensionality reduction method. By reducing high-dimensional data to 2-dimension space, we visualized these data and found several features.

(1) Spatial features. On low dimensional space, each link is represented by a point. We cluster these 50 links into 3 categories and marked them by different colors. Links in the same category can be viewed as sharing similar traffic pattern.

(2) Temporal features. Through dimensionality reduction, we have status points on 2-dimensional space. Each status point represents the overall traffic status of a specific area at some time point. Studying the evolving of status point through one day or all 16 days, we can clearly see the periodicity of traffic pattern. Also, the status point's evolving pattern is greatly different between workday and weekends.

The low-dimension result provides us a typical traffic pattern of an area. Compare the typical pattern with real record of a day, we can identify abnormal events when the trace of status point deviates from the typical pattern. At last, traditional feature extraction method, represented by PCA is carried out and compared to LLE. Result shows that LLE performs better than PCA on identifying typical traffic pattern and feature extraction. LLE is only one of the nonlinear dimensionality reduction methods. Its good performance on feature extraction of traffic data inspired us to explore more on such nonlinear methods.

Reference

Samant, A., and Adeli, H. (2000). "Feature extraction for traffic incident detection using wavelet transform and linear discriminant analysis." *Computer-Aided Civil and Infrastructure Engineering*, 15(4), 241-250.

Kerner, B. S., Rehborn, H., Aleksic, M., and Haug, A. (2004). "Recognition and tracking of spatial-temporal congested traffic patterns on freeways." *Transportation Research Part C*, 12(5), 369-400.

Ge, S. S.,Yang, Y., and Lee, T. H. (2006)." Hand gesture recognition and tracking based on distributed locally linear embedding." 2006 *IEEE Conference on Robotics, Automation and Mechatronics*, 2006, 1–6.

Guo, W., Yao, D. Y., Fu Y., and Hu J. M, and Liu N. "Study on the method for regional traffic flow feature extraction and traffic status evaluation." *Journal of Highway and Transportation Research and Development*, 22(7), 101-104.

Han, X., and Mu, Z. C. (2008). "Multi-pose ear recognition based on improved locally linear embedding." 2008 *Congress on Image and Signal Processing*, 39– 43.

Huang, L. Z, Zheng, L. X., Chen, C. Y., and Lu, M. (2009). "Locally linear embedding algorithm with adaptive neighbors." 2009 *International Workshop on Intelligent Systems and Applications*, 23-24, 1–4.

Liu, J. R., Wang, R. S., and Yang, J. X. "A scenario analysis of Beijing's private traffic patterns." *Journal of Cleaner Production*, 15(6), 550-556.

Tenenbaum J. B., de Silva V., and Langford, J. C. (2000). "A global geometric framework for nonlinear dimensionality reduction." *Science*, 290, 2319-2323.

Jain, V., and Saul, L. K. (2004). "Exploratory analysis and visualization of speech and music by locally linear embedding." *IEEE International Conference on Acoustics, Speech, and Signal Processing* (*ICASSP'04*), 3, 17-21, 984-987.

Lin, H. F, Sun, Y., and Shi, X. F. (2007). "Method of extracting traffic characteristics based on historical data information." *Urban Transport of China*, 5, 38-41.

Liu, C., Zhou, J. L., He, K., and Duan, Y. M. (2008). "Facial pose and expression analysis based on locally linear embedding." *2008 International Symposium on Computational Intelligence and Design, ISCID'08*, 2, 152-156.

Saul, L. K., and Roweis, S. T. (2000). An Introduction to Locally Linear Embedding.

Lan, L. W. Sheu, J. B., and Huang Y. S. (2008). "Investigation of temporal freeway traffic patterns in reconstructed state spaces." *Transportation Research Part C*, 16(1), 116-136.

Horn, R. A., and Johnson, C. R. (1990). *Matrix Analysis*, Cambridge University Press, Cambridge.

Roweis, S. T., and Saul, L. K. (2000). "Nonlinear dimensionality reduction by locally linear embedding." *Science* , 290, 2323-2326.

Wang X., Xue L. F., Yang D., and Han Z. Y. (2008). "Speech visualization based on locally linear embedding (lle) for the hearing impaired." 2008 *International Conference on Biomedical Engineering and Informatics*, 2, 502–505.

Wang Y., Hu J. M., Zhang Z. (2009). "Comparative study on feature extraction of mass traffic data using multiple methods." *Intelligent Vehicles Symposium*, 1179-1184.

Yuan W. T., Jia P., Wang L. P., Shao L. (2009). "Aircraft pose recognition using locally linear embedding." 2009 *International Conference on Measuring Technology and Mechatronics Automation*, 3, 454–457.

Study on the Evaluation System and Prediction Approach on Traffic Congestion Index of RTMS Data

Zhenjun Wang[1] Lijuan Que[2] Jing Liu[1] and Jizhen Guan[1]

[1] Engineering, ITS Research Center, Beijing Stone Intelligent Transportation System Co. Ltd., 2F, Xinsanyuan Office, NO.14 Zaojunmiao, Haidian District , Beijing, P. R. China, 100081, Email: greatpromise05@163.com
[2] Lecturer, Department of Financial and Economic Management, Beijing Information Technology College, No.5 Fangyuan Xilu, Chaoyang District, Beijing, P.R.China, 100016, Email: quelijuan@163.com

Abstract

From the perspective of time and space, this paper proposed traffic congestion index system based on data collected from RTMS of Beijing expressway. Traffic congestion condition in various time granularities and various space granularities can be evaluated by traffic congestion index system. Performance characteristics and trend of traffic congestion of Beijing expressway can be obtained and evacuation capacity of road is also evaluated by these indexes. Based on study of traffic evaluation index, this paper further proposed the prediction approach on traffic congestion index.
Key Words： RTMS; expressway; congestion index; evaluation; prediction

Introduction

Remote Traffic Microwave Sensor (RTMS) unit is a traffic sensor which uses microwave signals to detect vehicles. Unlike sensors which use the Doppler Effect, this sensor continuously modulates the transmitted frequency and can therefore detect stationary objects. So it is one kind of sensor for the detection and measurement of traffic on road. And t can be easy and safe to install by roadside and

it provides robust traffic information while requiring no maintenance (Image Sensing Systems Europe Limited, 2009). It is usually detected continuous traffic flow (e.g., traffic flow on urban expressway and traffic flow on highway). It provides per-lane presence indications (e.g., volume, occupancy, vehicle speed, and classification information).

Until 2008 Dynamic Information System of Real-time traffic data on Beijing Road Network had covered 2nd Ring Rd, 3rd Ring Rd, 4th Ring Rd, 5th Ring Rd and some connecting roads in Beijing (Yu, 2002). In urban the travelers prefer to drive on the expressway due to without signal control. So it takes on the most traffic load. Traffic condition of expressway will directly affect the whole road network traffic condition. Hence, congestion evaluation of expressway is one of the most important components of traffic condition evaluation in a city.

This paper is organized into six sections. The first section provides a description of the background for the introduction and the application of RTMS in Beijing. In the second section, a brief literature review is given on traffic congestion evaluation index. The third section provides an introduction of traffic congestion index system and index algorithm based on RTMS data. In the fourth section, this congestion evaluation index system is applied to evaluate traffic congestion condition of 2nd Ring Rd and 3rd Ring Rd in two days of 2008. In the fifth section, traffic congestion prediction based on traffic congestion index is proposed. In the last section, conclusions are presented and study needs for the future are addressed.

Literature Review

Congestion is rising with the rapid growth of economy and people. Road congestion is a universal problem affecting every major city. Most urban areas had implemented a wide variety of projects and programs to deal with traffic congestion. It is important to recognize how current congestion level is and whether these measures can relieve traffic congestion.

There are many researches on traffic congestion evaluation in overseas, which mainly concentrate in American, Europe, and Japan etc. In the congestion management of some cities some mobility measurement indexes (e.g., Volume to Capacity ratio (V/C), Level of Service (LOS), intersection delay and average travel speed during peak hours etc.) are proposed (Tulsa Transportation Management Area, 2009). Researches on traffic congestion evaluation are systematically carried out and set up the comparative perfect congestion evaluation index system in Urban Mobility Report (Schrank et al., 2009) and Congestion Management System (Texas Transportation Institute 3135 TAMU, 2004). Scholars in other countries had also

researches on congestion evaluation index. Evaluation of road traffic congestion is studied by fuzzy techniques (Pongpaibool et al., 2007). This system takes a road-traffic video signal as an input and computes vehicle volume and velocity. However, this method is more applied to the evaluation of one road or several roads. Data from GPS can be also used for the classification of road traffic congestion levels. The ratings and velocity were fed into decision tree learning model. The study successfully extracted vehicle movement patterns to feed into the learning model using a sliding windows technique (Thianniwet et al., 2009).

In abroad some research institutes had carried out the traffic congestion evaluation based on floating-car data (FCD) (Quan et al., 2007). Road congestion is a subjective quantity because it comes from the feeling of drivers. The continuous scale of 0-10 represents overall measure of the traffic conditions on a roadway. Stratified traffic congestion evaluation index system was put forward and applied in traffic information delivery. However, this index is only the figure and does not contain any physical meaning. For the travelers, this index system is little useful. And data from FCD rely on sample size on the road, owing to FCD collecting mechanism, so that the stability of FCD is not better than that of RTMS. The quality measure of traffic status according to a characteristic figure is studied (Liu et al., 2009). It is proposed based on the analysis of RTMS real-time data of Beijing expressway and traffic flow theory and statistics analysis method. In this study, the measurement of different roads speed is uniform. Obviously it is not reasonable.

This paper proposes the time-space congestion index system based on RTMS data. The congestion index system uses vehicle speed, vehicle volume and vehicle occupancy as inputs. Congestion evaluation index is congestion time counted on a road during an ordered period. So the index system can be easily understood by the popular. In the congestion evaluation index system, the minimum time is one hour and the minimum space is one road.

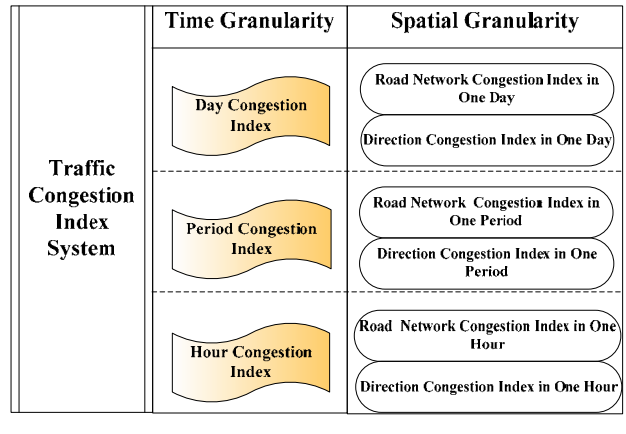


Figure 1. The framework of traffic congestion index system.

Introduction to Congestion Index System

Congestion Index system. From the perspective of time and space this paper proposes, traffic congestion index system. According to time levels, day congestion index, period congestion index and hour congestion index constitute time granularity indexes. This traffic congestion index system is applicable to evaluate the traffic condition for various road grades in various areas and in various periods.

Day Congestion Index (DCI). It is a macro-level index for the comparison of congestion condition on the same road in different days or the evaluation of the road network in one day(6：00~24：00). The range of DCI is from 0 to 18. DCI is composed of Road Network Congestion Index in One Day and Direction Congestion Index in One Day from the perspective of space.

Road Network Congestion Index in One Day (RNCID). RNCID can be applied to evaluate the road network, a specified area, one Ring Rd or one connecting road in one day(6:00~24:00).

Direction Congestion Index in One Day (DCID). One Ring Rd can be divided into outer Ring Rd and inner Ring Rd. And it can be also divided into four directions: East Ring Rd, West Ring Rd, South Ring Rd and North Ring Rd. Connecting road can be divided into inbound and outbound from directions. This index is used to evaluate road congestion condition of any direction in one day (6:00–24:00).

Period Congestion Index (PCI). There is strong time characteristics in traffic flow time series. There are obvious peak hours and off-peak hours in one day. PCI is a medium-level index and is composed of Road Network Congestion Index in One Period and Direction Congestion Index in One Period from the perspective of space. The road condition in any specified period can also be evaluated. If the selected period is a day, PCI becomes DCI; if the selected period is an hour, PCI becomes HCI.

Road Network Congestion Index in One Period (RNCIP). We can calculate road network traffic congestion index in morning peak hours, off-peak hours and evening peak hours. It can be evaluated the condition of road network, a specified area, one Ring Rd or one connecting road in one period.

Direction Congestion Index in One Period (DCIP). One Ring Rd can be divided into outer Ring Rd and inner Ring Rd. And it can be also divided into four directions: East Ring Rd, West Ring Rd, South Ring Rd and North Ring Rd. Connecting road can be divided into inbound and outbound from directions. We can calculate the

congestion index of each direction of the road in the specified period, morning peak hours, off-peak hours and evening peak hours.

Hour Congestion Index (HCI). This index can evaluate road condition continuously. From one day to one hour, time granularity gets small gradually and the tendency of congestion can be easily tracked. Finally the index will be suitable for congestion prediction. The range of Hour Congestion Index is from 0 to 1.

Road Network Congestion Index in One Hour (RNCIH). Each road in the whole road network by calculating RNCIH is evaluated. The whole road network, a specified area, one Ring Rd or one connecting road in an hour can be evaluated.

Direction Congestion Index in One Hour (DCIH). One Ring Rd can be divided into outer Ring Rd and inner Ring Rd. And it can be also divided into four directions: East Ring Rd, West Ring Rd, South Ring Rd and North Ring Rd. Connecting road can be divided into inbound and outbound from directions. Each road from each direction of the road in one hour is evaluated.

Congestion Index Algorithm. This paper proposed the time-space congestion index that is obtained by summing congestion time counted on a road in an ordered period.

$$C = \frac{\sum_{i=1}^{N} T_i}{\sum_{i=1}^{N} L_i}, \quad i=1\ldots N \tag{1}$$

where C is the congestion index, T_i is the congestion time counted on a road, L_i is the length of studied road.

$$T_i = t_j \times \frac{\sum_{j=0}^{m} occ_j}{t_j} = \sum_{j=0}^{m} occ_j, \quad j=0\ldots m \tag{2}$$

where t_j is the number of time interval below the speed threshold, occ_j is congestion time occupancy of the road.

According to the congestion condition in Beijing, this paper gives the congestion speed threshold based on road grade. Congestion speed thresholds of Ring Rd and connecting road are shown in Table 1.

Table1. Congestion speed threshold of Ring Rd and connecting road.

Road grade	Congestion speed threshold
2nd Ring Rd	25 km/h
3rd Ring Rd	30 km/h
4th Ring Rd	40 km/h
Connecting road	25 km/h

Congestion Index Process Flow

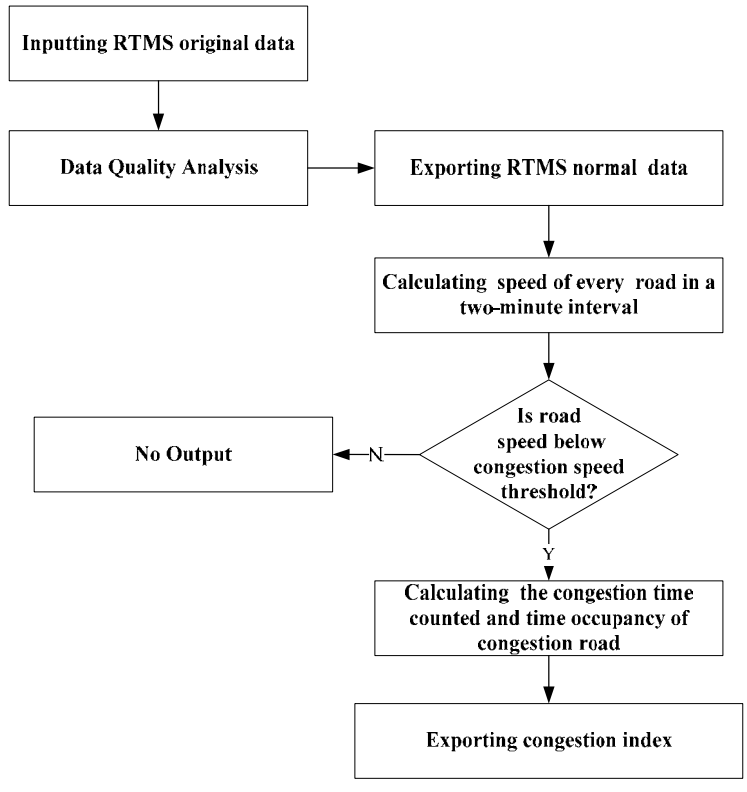


Figure 2. Congestion index process flow.

Experiments Study

2nd Ring Rd and 3rd Ring Rd in two days (6 June 2008 and 13 June 2008) in Beijing are taken as a case study. The two days are Friday. According to Weather Recording, it was raining from 17:40 to 18:30 on 13 June 2008 and was sunny in 6 June 2008. We can find whether the rain can influence traffic condition.

DCI Analysis

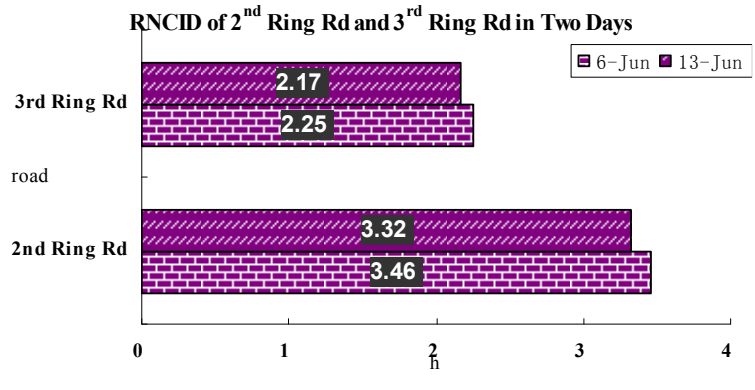


Figure 3. RNCID Comparison of two days.

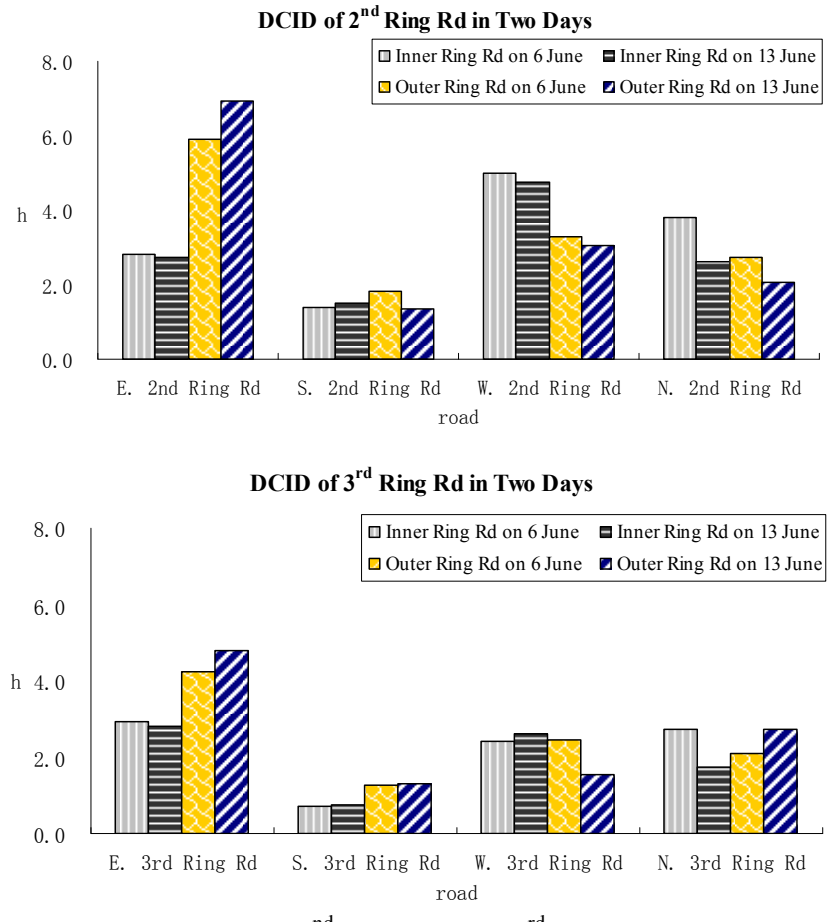


Figure 4. DCID of 2nd Ring Rd and 3rd Ring Rd in two days.

RNCID Comparison of Two Days. RNCID Comparison of two days is shown in Figure3. We can find that all congestion time on 3rd Ring Rd on 6 June is 2.25h. RNCID of 2nd Ring Rd on 6 June is the largest (3.46h) and traffic condition of 3rd Ring Rd on 13 June is the smoothest (2.17h). Generally RNCID of 2nd Ring Rd is 1h more than that of 3rd Ring Rd. Roads condition of 2nd Ring Rd and 3rd Ring Rd on 6 June are both worse than that on 13 June. It shows that rain on 13 June does not affect road condition.

DCID Comparison of Two Days. In two days DCID of E. Outer Ring Rd is obviously larger than that of E. Inner Ring Rd. Only the condition of E.2nd Outer Ring Rd and S. 2nd Inner Ring Rd among four directions of 2nd Ring Road on 13 June is worse than that on 6 June. There are more congestion directions of 3rd Ring Rd on 13 June. As a whole, the Ring Rd condition on 13 June is better than that on 6 June. The results of RNCID are the same as the result of DCID.

PCI Calculation. Because the rain occurs at evening, the specified period in this paper is from 17:00 to 21:00.

RNCIP Comparison of Two Days. In this period the Ring Rd on 6 June runs more smoothly than that on 13 June. And especially the congestion time of 2nd Ring Rd is longer than that of 3rd Ring Rd in two days. It shows that rain should affect travelling in this period.

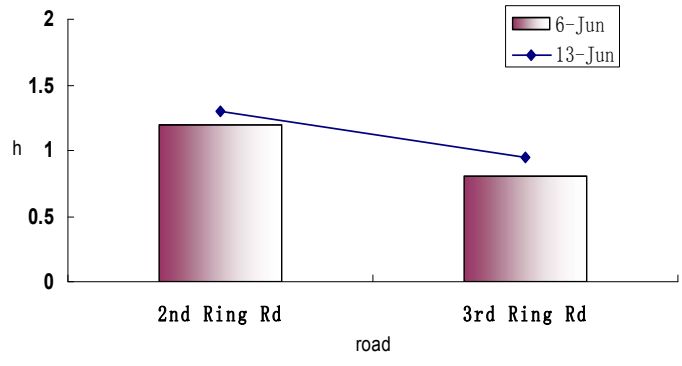
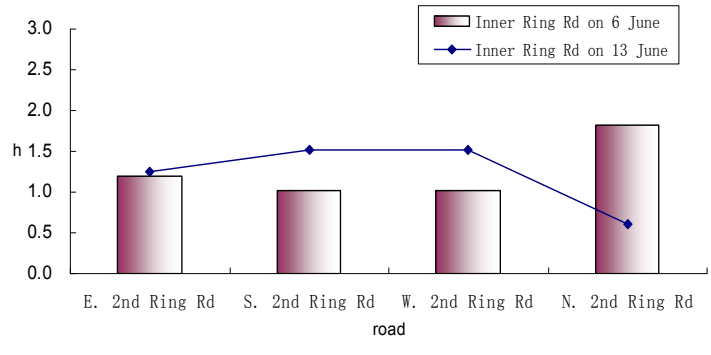
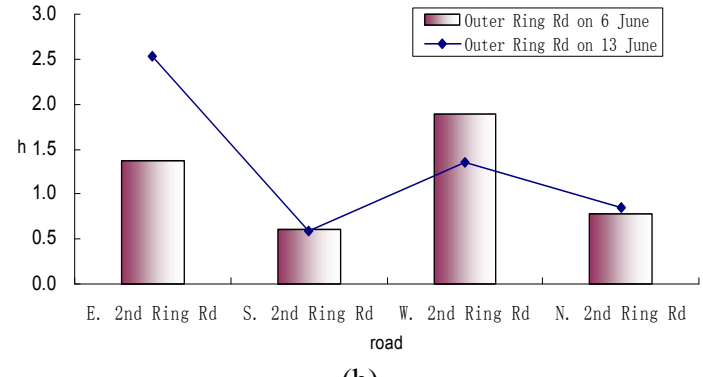


Figure 5. RNCIP Comparison of two days.

DCIP Comparison of Two Days. Except for N. 2nd Inner Ring Rd, W. 2nd Outer Ring Rd, N. 3rd Inner Ring Rd and W. 3rd Outer Ring Rd, the whole road condition on 13 June is worse than that on 6 June. This result is the same as RNCIP.



(a)



(b)

Figure 6. DCIP of 2nd Inner Ring Rd and 2nd Outer Ring Rd in Two Days.

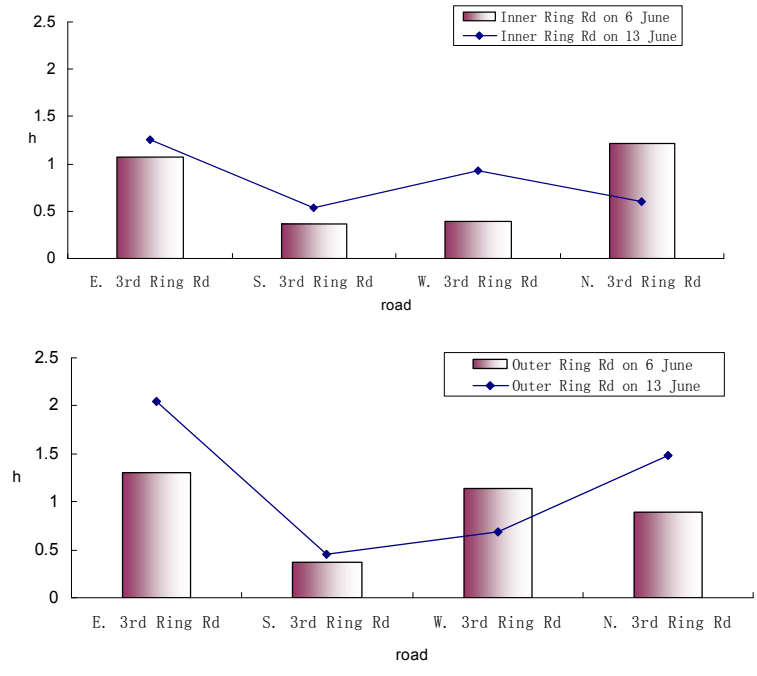
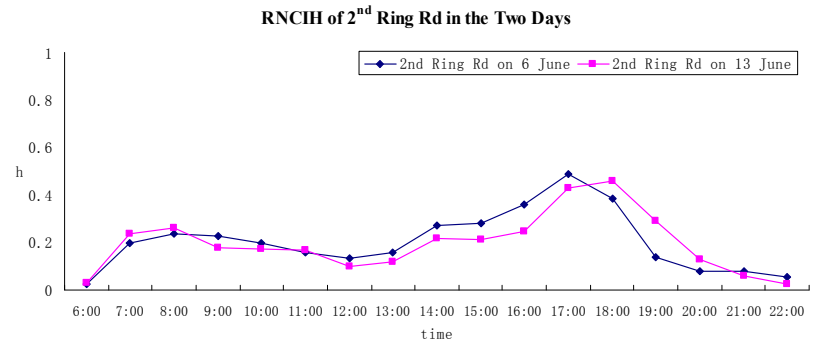
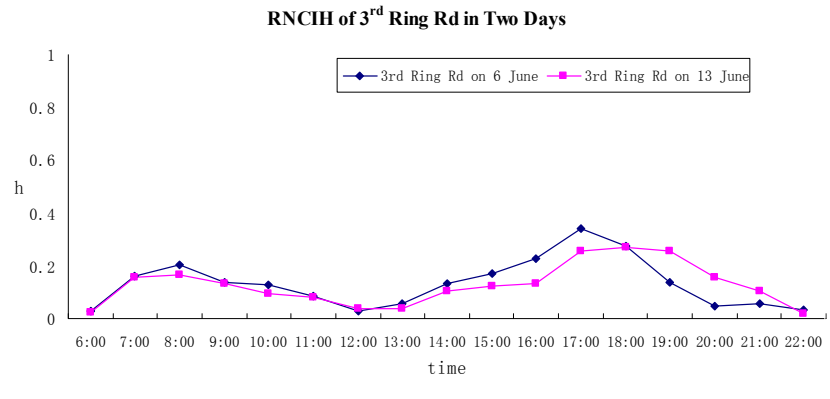


Figure7. DCIP of 3rd Inner Ring Rd and 3rd Outer Ring Rd in Two Days.

HCI Calculation



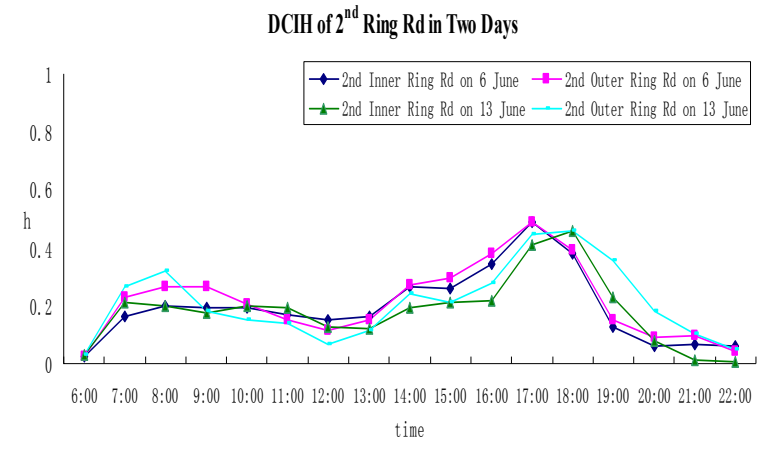
(a)



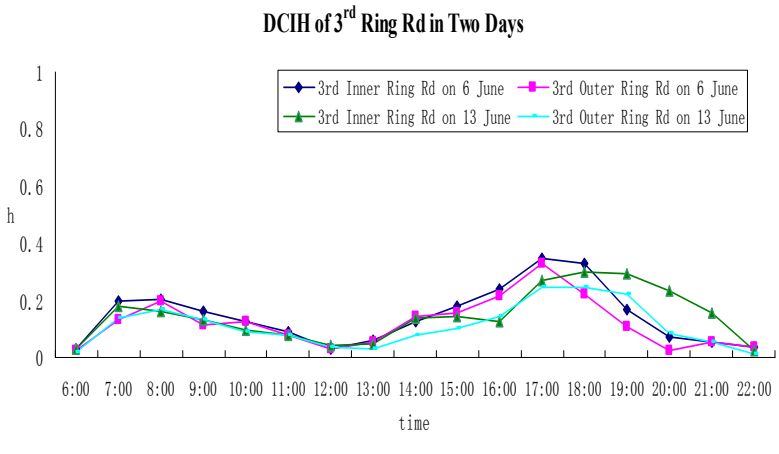
(b)

Figure 8. RNCIP Comparison of two days.

RNCIH Comparison of Two Days. For 2nd Ring Rd in evening peak hours, congestion period of traffic condition on 13 June more serious than that on 6 June is from 17:30 to 21:30. And for 3rd Ring Rd it occurs from 18:30 to 22:00. It indicates that congestion range diffuses from inside to external side and congestion duration becomes gradually short. Except evening period, congestion condition on 6 June is worse than that on 13 June. Congestion condition of 2nd Ring Rd is more serious than that of 3rd Ring Rd.



(a)



(b)

Figure 9. DCIH Comparison of the two days.

DCIH Comparison of Two Days. As shown in Figure 9, rain effect on 2nd Ring Rd and 3rd Ring Rd is obvious. During evening peak hours (17:00-19:00) the congestion condition is more serious than that during morning peak time. As it is raining at night of 13 June, congestion time is lengthened about 2 hours. It shows that most people prefer to drive on 2nd Outer Ring Rd and 3rd Inner Ring Rd while it is raining. Congestion evacuation capacity of 2nd Ring Rd and 3rd Ring Rd are very strong.

It appears that result of DCI is in contradiction to the result of PCI and HCI in the above experiment due to from various aspects. But DCI is also right. Except for the evening period, traffic condition on 6 June is more series than that on13 June. Because HCI can more reflect the detail of traffic condition, the effect of heavy rain on road condition can be only found by it.

Congestion index prediction. Current congestion index mostly can evaluate congestion condition based on historical traffic flow. Drivers pay more attention to real-time data, because traffic condition varies with time. If you want to improve the accuracy of route guidance, real-time data should be considered. The most important reason is that current congestion index (Guo et al., 2007) does not contain physical meaning. Congestion index system mentioned in this paper is congestion time per mileage. In congestion index prediction algorithm PNCIH and DCIH can be used. HCI is a minimal-level index and a continual evaluation index. It is applied to congestion index prediction. For travelers, this prediction interval can be easily accepted. So we can predict congestion time index C_n, according to road congestion time index C_{n-1} at last moment and historical congestion time index L_{n-1} at last moment, historical congestion time index L_n at this moment and adjust factor ς at this moment (e.g., weather, traffic administration, and event etc). In this following formula, α, β, η is the weighted factor of congestion index, $\alpha + \beta + \eta \le 1$ and $0 \le \alpha, \beta, \eta \le 1$.

$$C_n = \alpha \square L_{n-1} + \beta \square C_{n-1} + \eta \square L_n + \varsigma \tag{3}$$

Conclusions

Congestion index system is composed of three levels from one day to one hour, from the road network to one section. DCI is the macroeconomics index. It is more suitable to evaluate road condition in one year or one month, so the changing trend of congestion is found. PCI can be applied to evaluate road condition in the specified period. And HCI is the microcosmic index and can be used to evaluate road condition in one day. Thus Effect of heavy rain on road condition can be obtained by HCI. It can also be applied in the prediction of congestion index. In order to meet to need of traffic management, this congestion index system can also be normalized from 0 to 10. Because the index system is congestion time on the congestion road, the system can be applied to the prediction of congestion time in one-hour interval by HCI. In the future study, congestion time prediction and calibrating factor are studied so as to improve accuracy of congestion time prediction.

Acknowledgement

This work is supported by a grant from the National High-tech Research and Development Program of China (863 Program) (NO.2007AA11Z220).

References

Guo, J. F., Liu, M. H., Yu, L., Guan, J. Z., Guo, S. X., and Zhang, X. L.(2007). "Development and application of macroscopic measurement of traffic congestion in Beijing." *3rd China Annual Conference on ITS*, 26-33.

Image Sensing Systems Europe Limited(2009). "RTMS G4 introduction." *http://www.rtms-by-eis.com/en/RTMS/RTMSG4/tabid/65/Default.aspx.*

Liu, S. Y., Guan, J. Z., Jiang, X. X., and Liu, L.(2009)."Quantify measurement of Beijing freeway traffic status based on characteristic figures." *Journal of Transportation Systems Engineering and Information Technology,* 39-44.

Pongpaibool, P., Tangamchit, P., and Noodwong, K.(2007). "Evaluation of road traffic congestion using fuzzy techniques." *TENCON 2007 - 2007 IEEE Region 10 Conference*, 1-4.

Quan, Y. S., Guo, J. F., Guan, J. Z., Yu, L., and Wen, H. M (2007). "Research on traffic congestion evaluation and empirical analysis of Beijing." *3rd China Annual Conference on ITS*, 1–6.

Schrank, D. and T. Lomax (2009). *The 2009 Urban Mobility Report.* Texas: Texas Transportation Institute.

Texas Transportation Institute 3135 TAMU, Cambridge Systematics(2004). "Current conditions and trends from archived operations data." *Monitoring Urban Freeways in 2003*, 12–18.

Thianniwet, T., Phosaard, S., and Wasan, P. (2009). "Classification of road traffic congestion levels from GPS data using a decision tree algorithm and sliding windows." *The 2009 International Conference of Computer Science and Engineering*, 105-109.

Tulsa Transportation Management Area. (2009).*Congestion Management Process Report.*

Yu, C. Q. (2002). "Dynamic information system of real-time traffic data on Beijing road network." *Communication and Transportati0n Systems Engineering and Information,* 22-28.

Hardware-in-the-Loop Simulation: Challenges and Solutions

Isaac K. Isukapati[1], Alixandra Demers[2], George F. List [3]

[1] Graduate Research Assistant, [2] Research Assistant Professor, [3] Professor and Department Head, Department of Civil, Construction, and Environmental Engineering, North Carolina State University, 208 Mann Hall, Campus Box 7908, Raleigh, NC, USA 27695; Email: isaac_assiduous028@yahoo.com; ademers@ncsu.edu; gflist@ncsu.edu

Abstract

This paper discusses some complex Hardware-In-the-Loop configurations used for evaluating three distinct control strategies aimed at minimizing dilemma zone problems at high-speed, rural intersections. Five intersections were studied; three configurations were tested. The authors show that effective Hardware-In-the-Loop Simulations can be created for complex signal control configurations. They demonstrate that it is important to become familiar with the electrical details of the components and then create software and hardware splices that tie the devices together in such a way that each one functions in the manner intended. The authors show that for speed traps, the time step in simulation models dictates longer-than-field-based separations, about 29 m (95') instead of 6.6 m (20') (as in the field). This clearly suggests that time step duration has a significant impact on the results obtained for HILS, and hence suggests it worthwhile to push that frontier further, striving for even smaller time steps so that HILS can be effective.

Key Words: hardware-in-the-loop simulation; cabinet-in-the-loop; controller interface device; high-speed rural intersections

Introduction

It is common practice to use simulation to evaluate new traffic control strategies. Transyt 7F, and Passer have been used extensively in this regard and they are useful for quickly evaluating benefits and designing fixed-time signal plans (Balk *et al*. 2000, Bonneson *et al*. 2002). In addition, software packages like Synchro can optimize most actuated and coordinated controller systems (Bullock & Catarella 1998). A weakness in most of these traffic simulation models is their ability to

represent complex control strategies, especially those dependent on sophisticated detector placements. Since various systems have different control logic algorithms implemented on unique computing platforms, the best way to evaluate various traffic control systems and their control logic is to use the actual hardware. Hardware-In-the-Loop Simulation (HILS) is an effective way to accomplish this objective. It is important to note, however, that sometimes creating the interface between the traffic controller hardware and the computer running the microscopic traffic simulation is very challenging. Addressing those challenges is the focus of this paper.

This paper discusses three HILS configurations used for evaluating control strategies aimed at minimizing the likelihood that vehicles will be caught in dilemma zones at high-speed, rural, isolated intersections. Five intersections were studied, chosen based on their differences in mainline topography, intersection geometry, signal phasing, and cycle lengths. For each intersection, three configurations were tested, as defined below.

(a) Base case: traffic simulation software (VISSIM)+controller interface device (CID)+2070 Eagle controller running OASIS

(b) Long vehicle system (LVS): traffic simulation software (VISSIM)+CID+2070 Eagle controller running OASIS+upstream speed traps connected to NQ4 LVS devices

(c) Detector-control system (D-CS): traffic simulation software (VISSIM)+CID+2070 Naztec D-CS controller+upstream speed traps

For the second and third configurations, customized test set-ups were created that tie together the components. This paper details the challenges encountered and solutions created to enable HILS of these complex configurations.

Literature Review

Motivation. It is now common to use vehicle detection in signal control. A common feature in all traffic control systems is that they monitor detector calls and control the load switches. Different controllers implement the control logic slightly differently depending on the computing platform and the algorithms employed are typically not available in a stand-alone form for scientific evaluation by traffic engineering community. Several researchers have explored ways to integrate actual controllers into a simulation environment (Bullock *et al*. 1999; Burns & Wellings, 1990; Engelbrecht *et al*. 1999; Haenel and Messer 1974). This Hardware-in-the-Loop (HIL) architecture provides a hardware-based framework for evaluating the performance of a particular controller/control strategy in conjunction with traffic simulation. Figure 1 describes the abstract notion of HILS; four components are involved.
(1) A microscopic simulation model (engine) that moves vehicles through a defined network and tabulates MOE's. The control logic is implemented by an external traffic controller that is hooked up to the computer running a simulation model. All

the detector calls from the simulation model are transferred to the external controller through the CID.
(2) A Controller Interface Device (CID): The purpose of this device is to provide the interface between the computer running a traffic simulation model and the traffic controller. The interface is based upon discrete controller inputs and outputs. These discrete states are exchanged via the voltage levels used to monitor the load switches.
(3) A software interface module: It provides the interface between the CID and microscopic simulation model; it is typically implemented in a dynamic link library (DLL) module.
(4) The controller, and possibly other types of hardware.

Application of Real-time Microscopic Simulation Technology. In a simulation model, the movement of vehicles is performed by recalculating the positions of each vehicle every time step, typically once per second, or in some instances, as frequently as once every 0.1 seconds. In the case of HILS, at each time step, vehicle positions, speeds, and accelerations are updated in response to the traffic conditions and the signal indications obtained from the controller through the CID. Correspondingly, detector states are updated and these actuations are passed via the CID to the controller.

Researchers are finding that HILS is a powerful method to test control strategies and build practitioner confidence. (Roe 2009) used HILS to evaluate advanced Intelligent Transportation Systems. Abbas (2008) used CORSIM-based Cabinet-in-the-Loop Simulation to evaluate the number of vehicles caught in dilemma zones.

Certain aspects of the simulation process are especially important when HILS is used. Four of the most important are:
(1) The simulation must run in real-time in order to ensure synchronization between the microscopic simulation model and the traffic controller,
(2) The simulation must run in evenly-spaced simulation time-steps, and
(3) The simulation of each time-step must run faster than real-time in order to ensure the interface software has sufficient time to run and wait for the real-time clock to reach the start of the next simulation period.
(4) The time steps must be short enough that the controller can respond to the vehicle movements in a way that replicates the manner in which it would function in the field. It is the fourth aspect that plays a prominent role in this paper.

There are several efforts to evaluate the performance of HILS (Husch 2000, Kwon *et al*. 2001, Li *et al*.(undated)). Burns *et al*. conclude that any real-time computer system must produce correct outputs in response to real-world inputs within well-defined deadlines (also see List *et al*. 2007a). An inherent limitation of any real-time simulation is the size of its time step. In an HILS-based simulation, this affects the granularity of the real-time status of the system being passed from the simulation model to the controller and vice-versa. Ideally, the data would be transferred at such frequent intervals, using time steps small enough that the controller cannot tell whether the data are coming from a real world system or not. Then, to the controller,

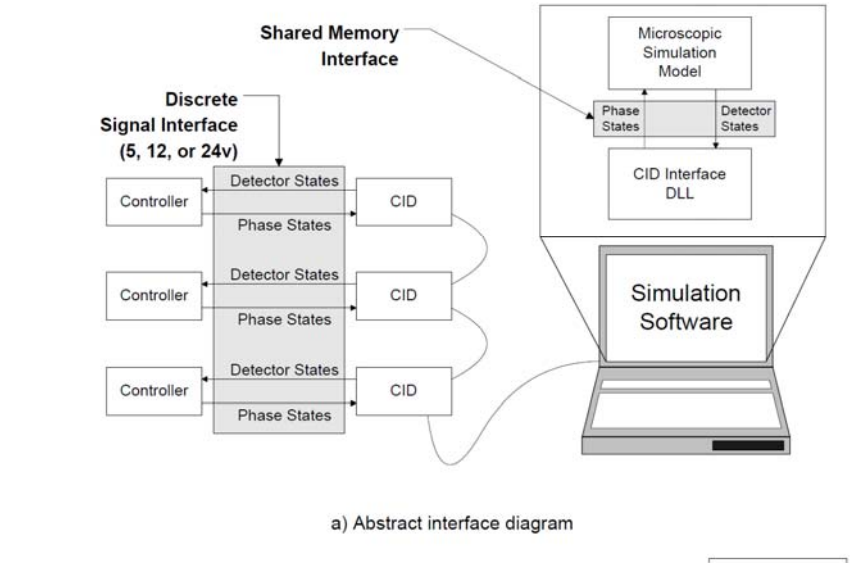


Figure 1. HILS Architecture (Source: Bullock, 2004)

time would appear continuous rather than discrete. Unfortunately, this is not yet feasible; time steps on the order of a millisecond would be needed. In the case of HILS, this means it is very important to understand the impacts of the time step size.

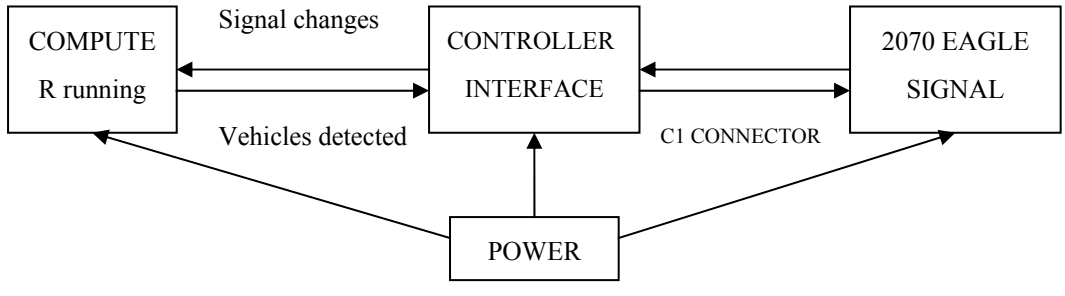
A second issue beyond time step size is latency. Latency occurs when the transmission of information is delayed. Li *et al*. studied the effects of latency on HILS using CORSIM. They concluded that latency in HILS system can be software and hardware dependent. From the traffic simulation model to the controller, latency results from the microprocessor clock speed and data handling routines, the CID interface software, and the USB software. Coming back in the other direction, hardware latency results from USB communications, CID signal conversions, the traffic controller itself, and signal transmission. The experimental results demonstrate that (a) 6 milliseconds of latency exist between consecutive USB data transmissions and (b) the CID hardware requires 1 millisecond to convert data from digital to analog or vice-versa. Therefore, there is a need to assist researchers in dealing with issues related to time step size and latency so that HILS can become a standard research tool.

Challenges and Solutions

Based on our recent HILS experiments, challenges were evident. Solutions were developed through discussions with the North Carolina Department of Transportation (NCDOT) Signal Shop, software developers, hardware manufacturers, electrical engineers, and months of tinkering as issues were realized. Each of the three HILS configurations is now described.

Base Case. The base case for each intersection was the configuration that currently exists in the field. This includes the actuated signal controller employed, its settings, and the detector configuration. An HIL configuration was created for each that included: a laptop computer running VISSIM (traffic simulation program), a CID, and a 2070 Eagle controller running OASIS software. The CID and controller are connected by a wiring harness with A, B, and C connectors on the CID side and a 104-pin C1 connector on the controller side (see Figure 2). The CID enabled detections that occur in the traffic simulation to be conveyed as detector inputs to the signal controller; similarly, it allowed signal indication decisions by the controller to be conveyed back to the traffic simulation.

Running HILS Using VISSIM and a 2070 Controller. Setting up proper communication between the traffic simulation model and the external traffic controller requires attention to detail. In the case of a 170 wiring harness, each detector in the field communicates with the traffic controller through a specific pin in the wiring harness. To represent this in simulation, a mapping has to be created between the VISSIM detector numbers and the C9 pin numbers on the CID end of the 170 wiring harness. Further, each of the C9 pin numbers also has to be mapped to the corresponding C1 pin number on the controller side of the 170 cable as shown in Table 1.



(a) VISSIM – CID – 2070 Eagle Controller Set-up.



(b) CID on Top of 2070 Eagle Controller Connected to Computer Running VISSIM

Figure 2. Base Case HILS Configuration.

It is important to understand how VISSIM and the CID exchange data. VISSIM uses Dynamic Data Exchange (DDE) to pass data to the CID. The DDE module in VISSIM is called the Signal State Generator (SSG) module. VISSIM updates detector states every 0.1 seconds when the simulation model is running and sends the detector states to a CID through a USB interface. The CID converts these digital signals to pulse signals and sends them to the traffic controller. The controller then updates the signal indications and sends those new indications back to VISSIM via the CID. Figure 3 illustrates this process.

Creating a CID Configuration File. A CID configuration file specifies to VISSIM which intersection will be controlled by an external traffic controller, and also which node will be connected to which CID. This file includes (a) the mapping between the VISSIM intersection number (SCJ number) and the CID ID number, and (b) phase pattern information, and traffic controller types for CID-controlled intersections.

Long Vehicle System Scenario. In this configuration, NQ4 Long Vehicle System (LVS) devices in conjunction with lane-specific speed traps are placed about 302.4 m (1,000') upstream of the stop line. The NQ4 triggers main street green holds when it sees long vehicles traveling faster than a threshold speed. This LVS is a supplement to the volume-density control. It represents the manner in which the NCDOT typically addresses the need for advance detection and high-speed protection. The speed traps are formed from loops placed 6.6 m (20') apart. This enables the NQ4 device to sense the speed and length of each oncoming vehicle. When the NQ4 device identifies "long" vehicles (over 6.7 m (22')) traveling at or above a given speed threshold (88 kph (55mph) for our locations), it places a hold (12 seconds for our case) on main street green (using a detector input) to ensure that the oncoming vehicle can pass through the intersection without encountering an amber, even if the maximum green is violated.

To create this configuration, the HILS expands to include four components: a laptop computer running VISSIM, a CID, an NQ4 LVS device, and a 2070 Eagle controller. The CID connects to the controller via a 170 wiring harness as was the case previously. In addition, the NQ4 device is spliced into the wiring harness as shown in Figure 4. The CID passes loop actuations for the upstream speed traps to the NQ4 device; and then it analyzes these inputs and decides if a long vehicle has been detected and if it is traveling fast enough that a hold on main street green is warranted. If the vehicle is long and traveling at a high speed, then the NQ4 device sends a signal directly to the controller to hold the main street green for a fixed number of seconds so as to reduce dilemma zone occurrences.

Incorporating the NQ4 into HILS is not as simple as it might seem. First, transferring the long vehicle detector calls from VISSIM to the NQ4 device through the CID requires attention to detail; the logical pin assignments do not match the physical ones. This required buzzing out all 104 pins in the 170 harness; a sample of matches is shown in 错误！书签自引用无效。.

Table 1. Logical and physical pin assignment matchings

Function	C1 pin #	Logical Pin # C7	Logical Pin # C9	Physical Pin # A	Physical Pin # B	Physical Pin # C
	1			37		
4 DW	2	39				33
4 Walk	3	40				12
4 Red	4	10				3
4 Yellow	5	11				23
4 Green	6	12				2
3 Red	7	7				52
3 Yellow	8	8				51
3 Green	9	9				24
2 E&C	39		3		18	
6 E&C	40		17		11	
4 E&C	41		10		33	
8 E&C	42		24		26	
2 Call	47		5		17	
6 Call	48		19		10	
4 Call	49		12		32	
8 Call	50		26		25	
Ped Inh	51		33		3	

C7 = Signal Head output from CID
C9 = Detector input to CID

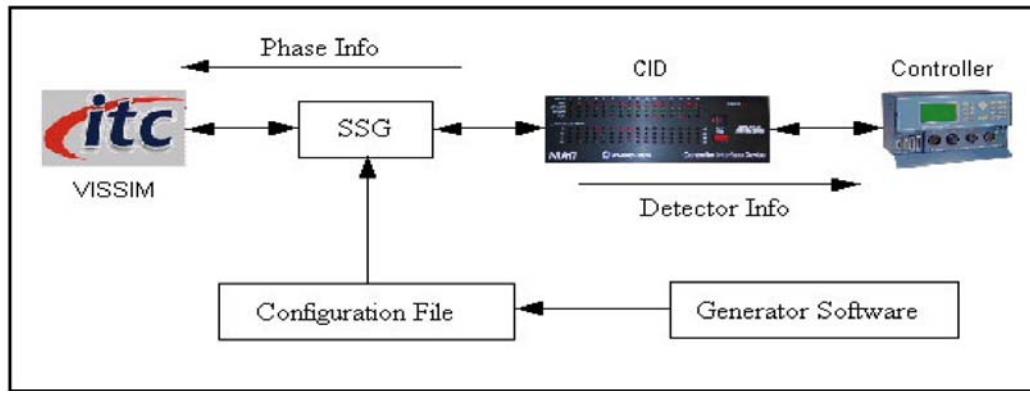


Figure 3. Schematic of the CID system (Source: NIATT)

The next challenge is to make sure the NQ4 can calculate the lengths and speeds of vehicles with an acceptable level of accuracy. This cannot be done if the speed trap detectors are actually placed 6.6 m (20') apart as they would be in the field. The reason is that the 0.1 second time step size is coarse to et an accurate sense of the vehicle speeds. For example, a vehicle traveling 60 mph (26.8 m/s) takes 0.246 seconds to travel the 6.6 m between the detectors. At 65 mph (29.1 m/s) it takes 0.227 seconds. To distinguish between these, the NQ4 device needs to tell that the time at 65 mph is 19 milliseconds shorter than the time at 60 mph. If the simulation

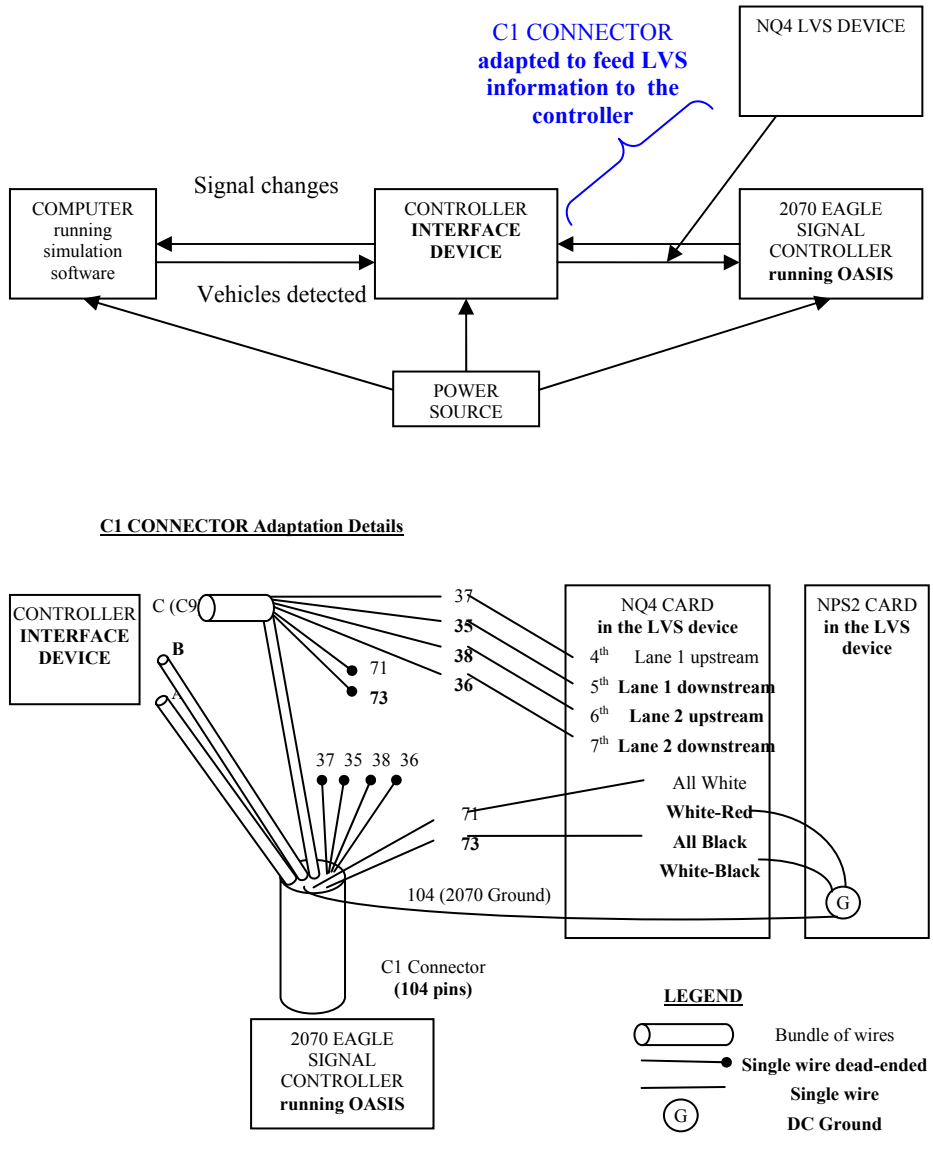
time step were 0.02 seconds instead of 0.1 seconds, this might be adequate. Since the time step could not be adjusted that small, this could not be tested. It was determined, however, that if the detectors for the speed traps were placed 29.0 m (95') apart, the NQ4 could transmit holds when appropriate given the 0.1 second time step.

Naztec D-CS Control Strategy. Bonneson *et al*. (2002) suggested a detection-control system (D-CS) to enhance dilemma zone protection. This is the third configuration created. Functionally, the D-CS decides whether it is better to continue the current green phase or terminate it based on information provided by the "vehicle detection system" and the "traffic control system". The vehicle detection system provides information about vehicle lengths, speeds, and lane locations from upstream major-road detectors; while the traffic control system determines from those inputs and other data when it would be best to transition to yellow on the main line approaches. Bonneson et al. (2002) suggest that system effectiveness is best measured by a tradeoff between two indicators: 1) the number of vehicles caught in the dilemma zone at the onset of yellow versus 2) overall motorist delay. They measure efficiency in terms of the number of detection loops needed for each lane of the major-road approach and the ease with which the system can be installed and operated. They assess the value of the system based on comparison with the "multiple advance detector system".

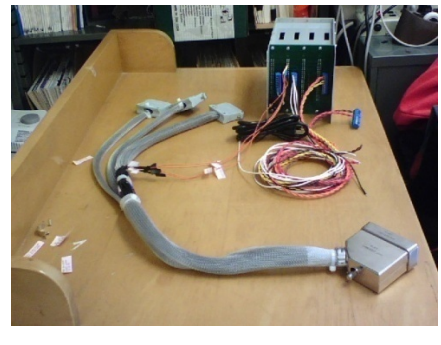
In essence, as said earlier, D-CS aims to minimize the number of vehicles, especially trucks, caught in dilemma zones at the onset of yellow. It tracks the vehicles as they approach the intersection; it projects, based on their speeds, how many vehicles will be in dilemma zones (2-6 seconds upstream of the signal) every 0.1 seconds into the future, and highlights the next time when that total will be zero. If one or more conflicting calls are received, it can use that point in time to end the main street green. The trade off is the amount of delay that will have accrued to the vehicles on the conflicting phases.

D-CS has been implemented in a 2070 controller by Naztec.

To implementing the D-CS control strategy in HILS, a novel way needed to be found to tie the controller to the traffic simulation model. The Naztec controller does not have a way to use a 170 wiring harness for input. The cabinet had to be in the loop (see Figure 5). The NCDOT Signal Shop prepared a C1 harness that could be connected to a de-soldered C1 back panel piece (the 2070 2A Field I/O Module shown in Figure 5(b)) that had all 104 pins exposed. Tagged wires were soldered to all the pins by the Signal Shop and these wires were then soldered or connected to the Naztec controller cabinet's detector and signal panels by the NCSU research team. This allowed VISSIM to interface with the Naztec controller via its normal TS-2 serial port which is connected to the bus interface units in the controller cabinet.

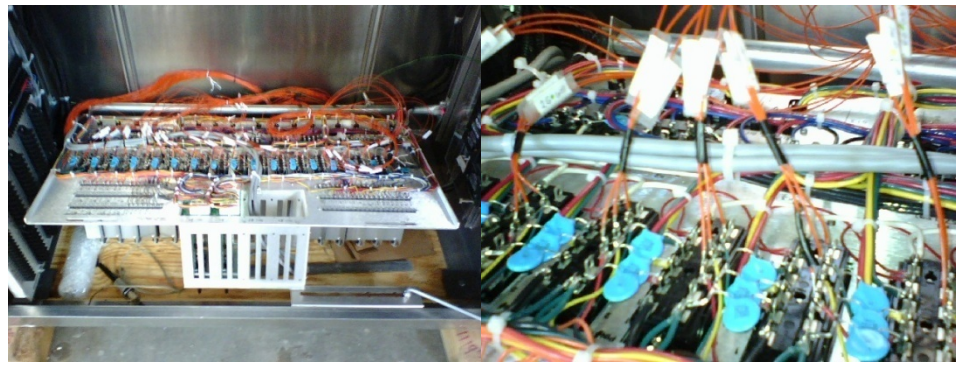


(a) VISSIM – CID – 2070 Eagle Controller & NQ4 LVS Device Set-up

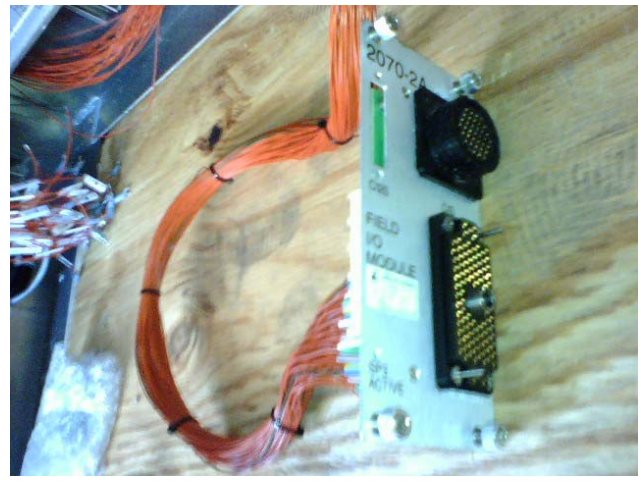


(b) C1 Connector with NQ4 LVS Wires Spliced In

Figure 4. Long vehicle system HILS configuration.



(a) Inside Naztec D-CS Cabinet: Rear Views of Loadswitch Socket Panel Showing the Newly Soldered Connections to the 2070-2A Field I/O Module



(b) Spare 2070-2A Field I/O Module Used to Adapt the Naztec Controller Cabinet to Perform a HILS; picture shows the 104 C1 wires soldered to the Module

Figure 5. Naztec D-CS control strategy

Conclusions and Recommendations

This paper shows how to create effective Hardware-In-the-Loop Simulations for complex signal control configurations. It demonstrates that it is important tobecome familiar with the inner-workings of the components and then create software and hardware splices that tie together the devices in such a way that each one functions in the manner intended.

The authors discovered that the speed trap for dilemma zone minimization needed to be 29 m (95') apart in the simulation instead of 6.6 m (20') to incorporate the NQ4 LVS device into the HILS system. This clearly suggests that latency in the system has a significant impact on the performance of HILS, and hence suggests it worthwhile to study the hardware and software latency issues in an HILS system running VISSIM.

References

Abbas, M. (2009) *Interaction of Simulation and Field Data Collection Equipment*, TRB Signal Systems Committee Summer Meeting, 2009.

Balke, K., Dudek, C., and Urbanik, T. (2000). "Development and evaluation of an intelligent bus priority concept." *Transportation Research Board Annual Meeting*, National Research Council, Washington, DC, Preprint # 00-1543.

Bonneson, J.A., Middleton, D., Zimmerman, K., Charara, H., and Abbas, M. (2002). *Intelligent Detection-Control System for Rural Signalized Intersections*, Report

Number FHWA/TX-03/4022-2, Prepared for Texas DOT, August.

Bullock, D., B. Johnson, R. Wells, M. Kyte, and Z. Li, (2004) "Evaluation procedures for traffic signal equipment using hardware in the loop simulation." Transportation Research Part C, 12(1), 73-89

Bullock, D. and Catarella, A. (1998). "A real-time simulation environment for evaluating traffic signal systems." *Transportation Research Record, #1634, TRB*, National Research Council, Washington, DC, 130-135.

Bullock, D., Morales J.M., and Sanderson B., (1999). "Impact of signal preemption on the operation of the virginia route 7 corridor." *Proceedings of the 1999 ITS America Conference*.

Burns, A. and Wellings A., (1990). "Real-time systems and their programming languages." *International Computer Science Series*. Addison-Wesley Publishing Company.

Engelbrecht, R., Poe C., and Balke, K. (1999). "Development of a distributed hardware-in-the-loop simulation system for transportation networks." *Transportation Research Board Annual Meeting*, National Research Council, Washington, DC, Preprint #990599.

Haenel, H.E., and Messer, C.J. (1974). "A report on the user's manual for progression analysis and signal system evaluation routine--passer ii." Texas Transportation Institute, TTI-218-72-165-14 Interim Report,

Husch, D., (2000). *Synchro Users Manual*, Trafficware, Berkeley, CA.

Kwon, E., Kim S., Kwon, T.M. (2001). "Pseudo real-time evaluation of adaptive traffic control strategies using hatdware-in-loop simulation." *The 27th Annual Conference of the IEEE Industrial Electronics Society.*

Li, Z., Kyte, M., Johnson, B., Wells, R., Abdel-Rahim, A., and Bullock, D. "Traffic real-time hardware-in-the-loop simulation performance analysis." a white paper

found at: http://www.webs1.uidaho.edu/mkyte/research_papers/CIDAnalysis Manuscript_1.pdf.

List, G.F., Demers, A., Isukapati, I.K., Hauser, E. (2007a). *Placement of Detection Loops on High Speed Approaches to Traffic Signals*, HWY 2007-13 Task 1 Report, prepared for North Carolina DOT,.

List, G.F., Demers, A., Isukapati, I.K., Hauser, E. (2007b). *Placement of Detection Loops on High Speed Approaches to Traffic Signals*, HWY 2007-13 Task 2 Report, prepared for North Carolina DOT, June 2007.

NAITT (2002). CID Manual.

Roe, M. (2009) *Development of a Hardware-In-the-Loop Analysis Framework for Advanced ITS Applications*, a thesis for Georgia Institute of Technology.

Robertson, D., *TRANSYT: A Traffic Network Study Tool*, Road Research Laboratory, RrlBpt Lr 253, 1969.

Suwal et al. (2005), "Real-time playback hardware-in-the-loop simulation of traffic systems." IEEE, 2005.

Network-wide Road Travel Time Estimation with Inconsistent Sensor Data

Hu Shao[1, 2], William H. K. Lam[3], Agachai Sumalee[4] and Anthony Chen[5]

[1] Associate Professor, Department of Mathematics, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China; Email: hu_shao2005@yahoo.com
[2] Research Associate, [3] Chair Professor, [4]Assistant Professor , [5]Associate Professor, Department of Civil and Structural Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China

Abstract

This paper proposes a new modeling approach for estimating road travel times in the territory-wide network. Owing to the rapid development of Intelligent Transportation Systems, travel times can be observed on links and/or paths in the road network by different sensor systems. The observed data from multiple sensor systems can provide more comprehensive and complete information on network-wide travel times than those from one sensor system. Thus, the use of the multiple sensor systems may have the potential for more accurate estimation of travel times in the territory-wide road network. However, the traditional travel time estimation methods typically consider the observed data from only one sensor system. The reason is that the observed data from different sensor systems may be inconsistent, which may bring difficulties in using these inconsistent input data for travel time estimation. This paper proposes a new two-stage model to overcome such data inconsistency difficulty for travel time estimation in the territory-wide road network with partial observed data from multiple sensor systems. In the first stage, the link travel time covariance matrix estimation model is proposed to estimate the covariance matrix of the link travel times for the whole network with the use of the second-order statistical property of the observed data from multiple sensor systems. Then, mean travel times on the links without sensors are updated according to the resultant estimated covariance matrix. In the second stage, a generalized least squares model is adopted to explicitly consider the data inconsistency issue for estimating the network-wide mean link and path travel times using the first-order statistical property of the observed data from multiple sensor systems. A heuristic solution algorithm is proposed for solving the two-stage model. A local road network in Hong Kong is used to demonstrate the applications of the proposed modeling approach.

Key Words: travel time estimation; data inconsistency; Intelligent Transportation Systems; generalized least squares method

Introduction

Travel time is widely recognized as an important measure for evaluating the operating efficiency of road networks, assessing performance of traffic management strategies, developing vehicle route guidance systems, etc. It is also regarded as the most useful traffic information available to road users (Margulici and Ban, 2008). Therefore, road travel time estimation problem is becoming an increasingly important issue in the field of transportation research and traffic studies.

Most of the existing travel time estimation models are mainly focused on estimation of travel time on a particular road link or path. Specifically, the day-to-day link-specific or path-specific travel time during a particular time period can be estimated (or observed directly) by many traffic detecting devices, such as loop detector, automatic vehicle identification (AVI) reader and Global Positioning System (GPS)-Geographical Information System (GIS). For convenience, these devices for travel time estimation are collectively referred to as "sensors" in this paper. For example, Robinson and Polak (2005) employed the k-nearest neighbor (k-NN) method for estimating the urban link travel time with inductive loop detector data. Guo and Jin (2006) adopted the cross-correlation method to estimate link travel time using traffic flow data collected from single loop detectors. Yeon et al. (2008) developed a discrete time Markov chain model to estimate travel time on a freeway where the states correspond to whether or not the link is congested. Taylor et al. (2000) presented a GPS-GIS system for collecting on-road traffic data, including path travel time from a probe vehicle. Tam and Lam (2008) estimated the path travel time with the use of the AVI tag readers' information. Liu and Ma (2009) proposed a virtual vehicle probe model for estimating time-dependent travel time on a signalized arterial road. It should be noted that all of the above studies are site-specific instead of the network-wide travel time estimation.

In principle, if each road link of the whole network can be covered by the sensor system, the network-wide travel time can be estimated by the conventional methods. In reality, a single sensor system covering the whole network may not be feasible for practical applications due to the budgetary constraints and physical limitations (Lam et al., 2002; Hu et al., 2009). Therefore, it is desirable to estimate network-wide travel times when only partial links or paths are installed with sensors or even different types of sensors. In view of this, this paper proposes a new method to estimate the travel times in territory-wide road network with focus on simultaneous use of different types of available observed travel time data. Several new features of the proposed new method are summarized as below in Table 1.

- Travel time data from all available sensor systems are simultaneously used to provide a wider-coverage of traffic information within the road network.

Table 1. Comparison between the conventional methods and the proposed new method for network-wide travel time estimation.

Features	Conventional methods	Proposed new method
Sensor system	Single	Multiple
Additional sensors	Needed	Not needed
Consideration of data inconsistency	No	Yes
Use of link travel time covariance	No	Yes

- It does not need additional or excess sensors to cover all the road links in the network.

- The data inconsistency issue is explicitly considered. The data from different sensor systems are usually inconsistent due to the measurement errors of various sensors at different locations.

- The spatial covariance structure of link travel times is fully utilized on the basis of available observed travel time data. Although the available data may not cover all links or paths in the road network, the spatial covariance structure of link travel times is used in the new approach so as to provide the additional information for estimation of travel times on the links without sensor data.

This paper is organized as follows. The next section presents the formulation of the proposed two-stage model for network-wide road travel time estimation. It follows with a solution algorithm for solving the proposed model. A numerical example is used to demonstrate the applications of the new model. Finally, conclusions are drawn together with discussion on further studies.

Model Formulation

The Flowchart of ohe Proposed Modeling Approach. As shown in Figure 1, a two-stage model is proposed to estimate the network-wide road travel times. In Stage I, the link travel time covariance (LTTC) matrix estimation model is employed to estimate the covariance matrix of the link travel times for the whole network. Then, mean travel times on the links without sensors are updated using the estimated covariance matrix. In Stage II, a generalized least squares (GLS) model is adopted to explicitly consider the data inconsistency issue for estimating mean travel times by link and path within the whole network.

Observed Link and Path Travel Times. Denote $\widetilde{\mathbf{A}}$ as the subset of link set $\mathbf{A}$. The travel time on link $a \in \widetilde{\mathbf{A}}$ can be observed by the sensor system during the study time period. Owing to uncertainties in road networks, the link travel time is stochastic across the observed time periods (Lam et al. , 2008). $\tilde{T}_{a,o}$ is denoted as the observed

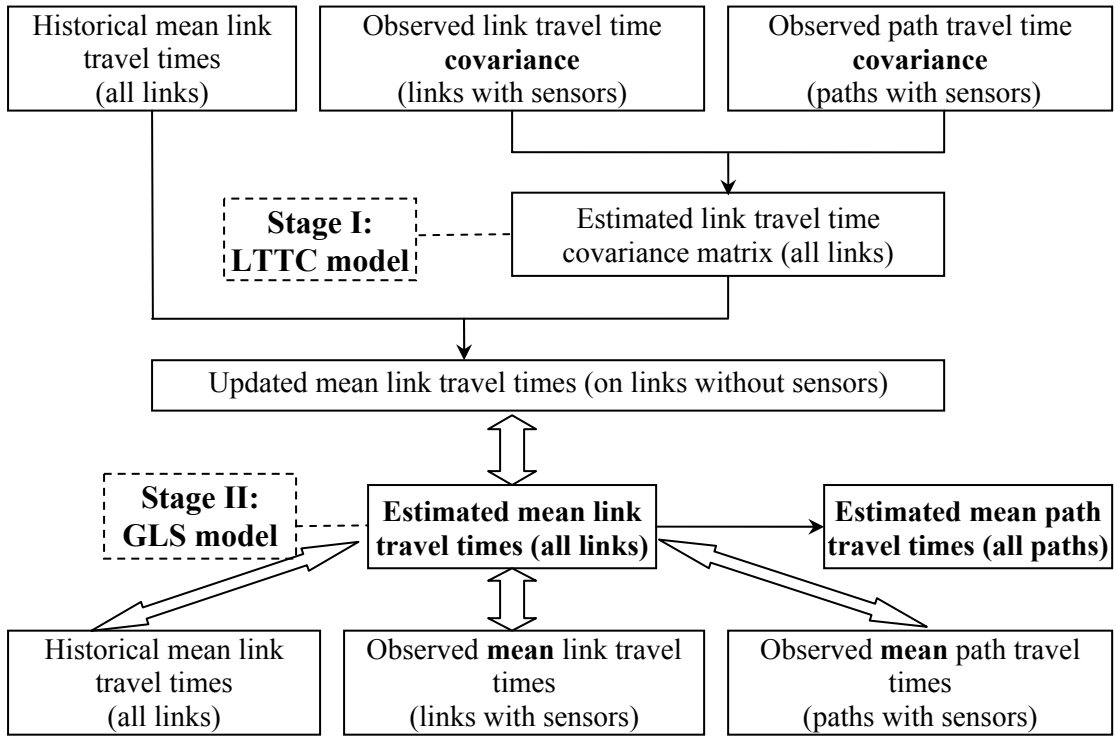


Figure 1. The flowchart of the proposed modeling approach.

travel time on link $a \in \widetilde{\mathbf{A}}$. $\tilde{t}_{a,o}^{(i)}$ is denoted as the observed travel time on link $a \in \widetilde{\mathbf{A}}$ during the study time period on date i $(i = 1,2,\cdots,l)$. Then, $\tilde{t}_{a,o}^{(1)}, \tilde{t}_{a,o}^{(2)}, \cdots, \tilde{t}_{a,o}^{(l)}$ is a sample of $\widetilde{T}_{a,o}$ with sample size l. For convenience, $\tilde{\mathbf{t}}_o^{(i)}$ is denoted as the vector of $\left\{\tilde{t}_{a,o}^{(i)}\right\}$ for all $a \in \widetilde{\mathbf{A}}$. Then, the sample mean of observed link travel times can be calculated as:

$$\tilde{\mathbf{t}}_{link}^{o} = \frac{1}{l}\sum_{i=1}^{l} \tilde{\mathbf{t}}_o^{(i)} \tag{1}$$

The sample covariance matrix of travel times on all links with sensors can be calculated as:

$$\tilde{\mathbf{\Sigma}}_{link}^{o} = \frac{1}{l-1}\sum_{i=1}^{l} \left\{\left(\tilde{\mathbf{t}}_o^{(i)} - \tilde{\mathbf{t}}_{link}^{o}\right)\left(\tilde{\mathbf{t}}_o^{(i)} - \tilde{\mathbf{t}}_{link}^{o}\right)^T\right\} \tag{2}$$

The travel time on paths with sensors can also be observed by another sensor system. Similar to Equations (1) and (2), the sample mean ($\tilde{\mathbf{t}}_{path}^{o}$) and sample covariance matrix ($\tilde{\mathbf{\Sigma}}_{path}^{o}$) of observed path travel times can be obtained, which are not presented in this paper due to the length limitation.

$(\Delta)_{n\times m}$ is denoted as the link-path incident matrix. $\mathbf{t}_{link}$ and $\mathbf{t}_{path}$ are denoted as the estimated mean link and path travel times, respectively. $\mathbf{\Sigma}_{link}$ and $\mathbf{\Sigma}_{path}$ are denoted as the estimated covariance matrix of link and path travel times, respectively. Then, the travel time conservation conditions in terms of mean and covariance are given as follows:

$$\mathbf{t}_{path} = \Delta^T \mathbf{t}_{link} \tag{3}$$

$$\mathbf{\Sigma}_{path} = \Delta^T \mathbf{\Sigma}_{link} \Delta \tag{4}$$

Stage I: LTTC Model. The order of the links can be rearranged so that the links with sensors appear first. Then, the estimated and historical mean link travel times can be partitioned according to whether the links are installed with sensors or not as below:

$$\mathbf{t}_{link} = \{t_{a_1}, t_{a_2}, \cdots, t_{a_o} \ \vdots \ t_{a_o+1}, t_{a_o+2}, \cdots, t_{a_n}\}^T = \{\mathbf{t}^o_{link}, \mathbf{t}^u_{link}\}^T \tag{5}$$

$$\bar{\mathbf{t}}_{link} = \{\bar{t}_{a_1}, \bar{t}_{a_2}, \cdots, \bar{t}_{a_o} \ \vdots \ \bar{t}_{a_o+1}, \bar{t}_{a_o+2}, \cdots, \bar{t}_{a_n}\}^T = \{\bar{\mathbf{t}}^o_{link}, \bar{\mathbf{t}}^u_{link}\}^T \tag{6}$$

Similarly, the estimated link travel time covariance matrix $\mathbf{\Sigma}_{link}$ can also be partitioned as follows:

$$\mathbf{\Sigma}_{link} = \left[\begin{array}{ccc:ccc} \sigma_{a_1,a_1} & \cdots & \sigma_{a_1,a_o} & \sigma_{a_1,a_o+1} & \cdots & \sigma_{a_1,a_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{a_o,a_1} & \cdots & \sigma_{a_o,a_o} & \sigma_{a_o,a_o+1} & \cdots & \sigma_{a_o,a_n} \\ \hdashline \sigma_{a_o+1,a_1} & \cdots & \sigma_{a_o+1,a_o} & \sigma_{a_o+1,a_o+1} & \cdots & \sigma_{a_o+1,a_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{a_n,a_1} & \cdots & \sigma_{a_n,a_o} & \sigma_{a_n,a_o+1} & \cdots & \sigma_{a_n,a_n} \end{array}\right] = \left[\begin{array}{cc} \mathbf{\Sigma}^o_{link} & \mathbf{\Sigma}^{o,u}_{link} \\ \mathbf{\Sigma}^{u,o}_{link} & \mathbf{\Sigma}^u_{link} \end{array}\right] \tag{7}$$

Similar to Eq. (4), the covariance of observed link and path travel times should also satisfy the conservation condition. However, due to the measurement errors, the sample covariance of links and paths with sensors may not be consistent. In view of this, the GLS method is employed to minimize the difference between the estimated covariance and observed covariance of link and path travel times. As multiple solutions may exist, the historical link travel time covariance matrix is adopted as the target matrix in the objective function (8). Therefore, the LTTC model is formulated as the following GLS problem:

$$\begin{gathered} \min\{\gamma_1 \left\| \widetilde{\mathbf{\Sigma}}^o_{path} - \widetilde{\Delta}^T \mathbf{\Sigma}_{link} \widetilde{\Delta} \right\|^2 + \gamma_2 \left\| \widetilde{\mathbf{\Sigma}}^o_{link} - \mathbf{\Sigma}^o_{link} \right\|^2 + \gamma_3 \left\| \overline{\mathbf{\Sigma}}_{link} - \mathbf{\Sigma}_{link} \right\|^2 \} \\ s.t.\ \mathbf{\Sigma}_{link} \succeq 0 \end{gathered} \tag{8}$$

where $\widetilde{\Delta}$ is the link-observed path incidence matrix; $\mathbf{\Sigma}^o_{link}$ is a sub-matrix of $\mathbf{\Sigma}_{link}$ which is defined in Eq. (7); $\overline{\mathbf{\Sigma}}_{link}$ is the historical link travel time covariance matrix; γ_1, γ_2 and γ_3 are non-negative weighting parameters, $\mathbf{\Sigma}_{link} \succeq 0$ represents that $\mathbf{\Sigma}_{link}$ is a positive semi-definite matrix.

According to Lam et al. (2002), the mean travel time on links without sensors can be updated using the estimated link travel time covariance matrix in the GLS problem (8):

$$\hat{\mathbf{t}}_{link}^{u} = \bar{\mathbf{t}}_{link}^{u} + \boldsymbol{\Sigma}_{link}^{u,o}\left(\boldsymbol{\Sigma}_{link}^{o}\right)^{-1}\left(\tilde{\mathbf{t}}_{link}^{o} - \bar{\mathbf{t}}_{link}^{o}\right) \tag{9}$$

where $\hat{\mathbf{t}}_{link}^{u}$ is the updated mean travel time on links without sensors; $\boldsymbol{\Sigma}_{link}^{u,o}$ is a sub-matrix of $\boldsymbol{\Sigma}_{link}$ as shown in Eq. (7).

Stage II: GLS Model for Network-Wide Travel Time Estimation. To overcome the aforementioned data inconsistency problem, the GLS method is adopted in an attempt to improve estimation of the mean travel times on all links within the road network. As mentioned above, there are in total four sets of travel time data available: (1) historical mean travel times on all links; (2) observed mean travel times on links with sensors; (3) observed mean travel times on paths with sensors; (4) updated mean travel times on links without sensors (Equation (9)). Therefore, the objective function is to minimize the difference between these four sets of mean travel times and the estimated mean link travel times. It can be expressed as the GLS model below:

$$\begin{aligned} \min\{&\omega_1\left\|\bar{\mathbf{t}}_{link} - \mathbf{t}_{link}\right\|^2 + \omega_2\left\|\tilde{\mathbf{t}}_{link}^{o} - \mathbf{t}_{link}^{o}\right\|^2 + \omega_3\left\|\tilde{\mathbf{t}}_{path}^{o} - \tilde{\Delta}^T\mathbf{t}_{link}\right\|^2 + \omega_4\left\|\hat{\mathbf{t}}_{link}^{u} - \mathbf{t}_{link}^{u}\right\|^2\} \\ &s.t.\ \mathbf{t}_{link} \geq 0 \end{aligned} \tag{10}$$

where ω_1, ω_2, ω_3, and ω_4 are the non-negative weighting parameters for different sets of mean travel times. Then, the mean path travel times can be calculated by using Eq. (3).

Solution algorithm

In this paper, a heuristic solution algorithm is proposed for solving the two-stage network-wide travel time estimation problem in practice. The procedure of the proposed solution algorithm is described as follows:

Step 1: Estimate the link travel time covariance matrix using the second-order property of observed link and path travel times together with the historical data as below:

Step 1.1: Find a solution of the following minimization problem:

$$\min\{\gamma_1\left\|\tilde{\boldsymbol{\Sigma}}_{path}^{o} - \tilde{\Delta}^T\boldsymbol{\Sigma}_{link}\tilde{\Delta}\right\|^2 + \gamma_2\left\|\tilde{\boldsymbol{\Sigma}}_{link}^{o} - \boldsymbol{\Sigma}_{link}^{o}\right\|^2 + \gamma_3\left\|\overline{\boldsymbol{\Sigma}}_{link} - \boldsymbol{\Sigma}_{link}\right\|^2\}$$

Step 1.2: If $\boldsymbol{\Sigma}_{link} \succeq 0$ (positive semi-definite), to go step 2. Otherwise, go to step 1.3.

Step 1.3: Carry out eigenvalue decomposition for $\boldsymbol{\Sigma}_{link}$, i.e.

$$\boldsymbol{\Sigma}_{link} = \mathbf{Q}^T diag(\lambda_1, \cdots, \lambda_n)\mathbf{Q}$$

where $\lambda_1, \cdots, \lambda_n$ is the eigenvalue of $\boldsymbol{\Sigma}_{link}$, $\mathbf{Q}$ is an orthogonal matrix.

Step 1.4: $\hat{\lambda}_i = \max\{\lambda_i, 0\}$, $i = 1,2,\cdots,n$.

Step 1.5: Calculate the link travel time covariance matrix as follows:

$$\boldsymbol{\Sigma}_{link} = \mathbf{Q}^T diag(\hat{\lambda}_1, \cdots, \hat{\lambda}_n)\mathbf{Q}$$

Step 2: Update the mean link travel times on links without sensors according to Equation (9).
Step 3: Find $\mathbf{t}_{link}$ by solving GLS problem (10).

Note: Steps 1.3-1.5 can guarantee that the constraint of GLS problem (8) is satisfied, i.e., $\mathbf{\Sigma}_{link}$ is a positive semi-definite matrix.

Numerical Example

To demonstrate the applications of the proposed road travel time estimation model, the Tuen Mun Road Corridor Network in Hong Kong is used as a study network. This network consists of 3 zones (C1, C2, and C3), 4 nodes, 10 links and 20 paths as shown in Figure 2. The simulation method is used to generate the observed link travel times and path travel times. A standard logit-based SUE traffic assignment model is used for the simulation experiment with dispersion parameter $\theta = 100$. In each simulation, the link capacity is generated from a continuous uniform distribution and the O-D demand is generated from a normal distribution. The historical mean link travel times, shown in Table 2, are generated from the SUE model on the basis that the mean values of the OD demand and of the link capacities are used as the inputs.

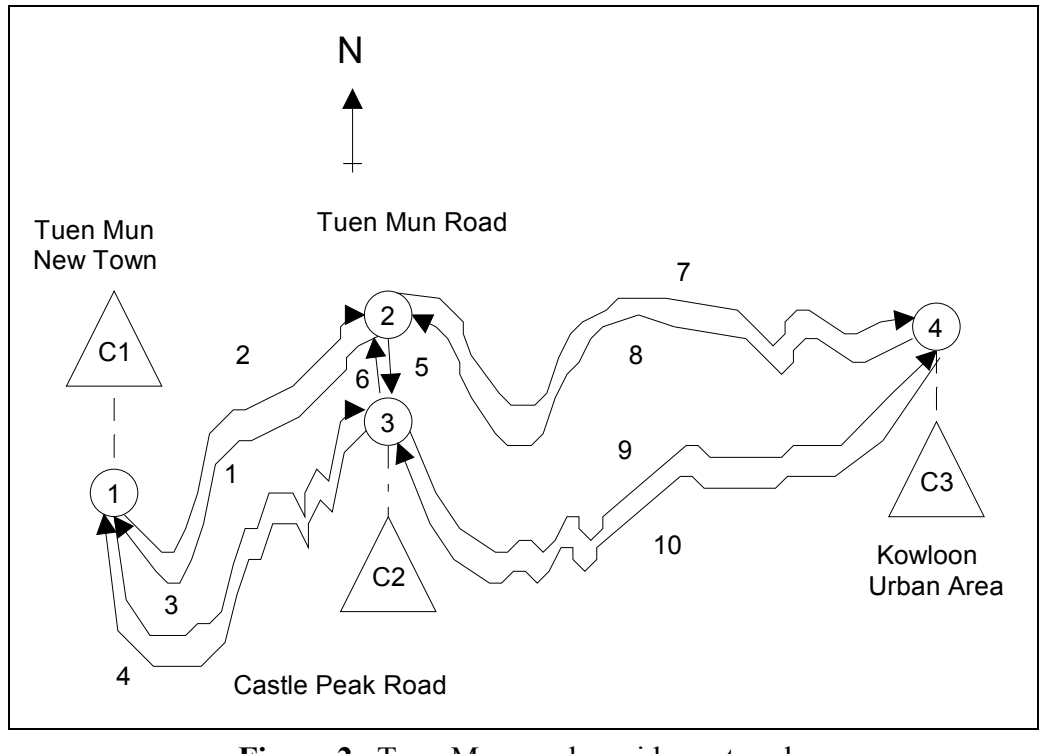


Figure 2. Tuen Mun road corridor network.

Table 2. Historical and observed mean link travel times.

Link no.	1	2	3	4	5	6	7	8	9	10
$\bar{\mathbf{t}}_{link}$ (hour)	0.1399	0.2742	0.2897	0.1506	0.0118	0.0086	0.1146	0.0590	0.1285	0.0804
$\tilde{\mathbf{t}}^{o}_{link}$ (hour)	0.1465	0.3104	0.3252	0.1574	0.0117	0.0096	0.1141	0.0618	0.1292	0.0843

Table 3. Number of reliable estimates under different scenarios.

Scenario No.	Number of links with sensors	Locations of sensors on links	Number of reliable estimates
A	**2**	2, 3	**5**
B	***5***	2, 3, 5, 7, 10	***6***
C	***5***	1, 4, 6, 7, 8, 9	***8***
D	**10**	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	**8**

In this numerical example, it is assumed that the link travel times are observed by loop detectors and the path travel times are observed by AVI readers. The standard SUE model is solved for 100 times. The observed mean and covariance matrix of link travel times are generated from the first 50 simulations. These 50 simulations represent the travel times observed over 50 homogeneous, independent time periods of typical weekdays. The observed mean link travel times are shown in Table 2. Similarly, it is supposed in this example that only 2 path travel times can be observed, which are path 1 (link 3) and path 2 (links 2 and 5). The observed mean and covariance matrix of path travel times are generated from the last 50 simulations. Because the link-based and path-based observed data are generated from different simulation results, data inconsistency exists between these two types of data. Due to the length limitation of this paper, some input data and results of simulation are not provided.

The conventional measures, such as the root mean square percentage error, for quantifying the quality of the estimates are not applicable in this example due to the data inconsistency from different sensor systems. In view of this, the concept of confidence interval is used as an alternative to examine the reliability or accuracy of an estimate. For a given confidence level $1-\eta$, the confidence interval for the mean travel time can be calculated as follows:

$$\left(\tilde{t}_{a_i} - \tilde{\sigma}_{a_i,a_i} t_{\eta/2}(l-1),\ \tilde{t}_{a_i} + \tilde{\sigma}_{a_i,a_i} t_{\eta/2}(l-1)\right)\ i = 1,2,\cdots,n \qquad (11)$$

where l is the sample size; $t_{\eta/2}(l-1)$ is the inverse of Student's T cumulative distribution function with $l-1$ degrees of freedom at the value $\eta/2$. If the estimated mean link travel time falls within the confidence interval, the estimated mean travel time is considered as a "reliable estimate". In this example, the confidence level is set to be 90% , i.e., $\eta = 0.05$. All the weighting parameters are assumed to be equal to 1, i.e., $\gamma_1 = \gamma_2 = \gamma_3 = \omega_1 = \omega_2 = \omega_3 = \omega_4 = 1$.

In total, there are four scenarios with different numbers of sensors and locations of sensors for testing the proposed model. The corresponding results are shown in Table

3. It can be seen that the number of reliable estimates is non-decreasing with the number of sensors. For example, the number of reliable estimates is 5 in scenario A with 2 link travel time sensors. It is increased to 8 when the number of sensors reaches to 10 in scenario D. This result is sensible because more sensors can provide more travel time information. Thus, the number of reliable estimates in scenario D is greater than that in scenario A.

An interesting finding is that the sensor locations can also influence the accuracy of travel time estimation. For example, although the numbers of sensors are equal in scenarios B and C (each with 5 link travel time sensors), the numbers of reliable estimates are 6 and 8, respectively. This finding indicates that the sensor locations should be optimized in practical application of the proposed model for network-wide travel time estimation. Particularly, when different sensor systems are used to provide travel time and traffic flow data, how to optimize the sensor locations using different mix of various sensor systems to improve the estimation accuracy and save the total sensor system cost would be a meaningful extension of the proposed model for further studies.

Conclusions and Further Studies

In this paper, the network-wide road travel time estimation problem has been formulated as a two-stage model with particular attention on inconsistent input data from multiple sensor systems. The input data are the observed link and path travel times, which are assumed to be collected by loop detector and AVI sensor systems, respectively. In the first stage, the link travel time covariance matrix is estimated using the second-order properties of the observed link and path travel times. On the basis of the resultant covariance matrix information, the travel times on links without sensor data can also be updated. In the second stage, a generalized least squares model is employed to minimize the difference between the estimated mean link travel times and the observed data (i.e. observed mean link and path travel times) as well as the historical data. A heuristic solution algorithm is proposed to solve the two-stage model. This proposed modeling approach has the potential to provide better travel time estimates in the territory-wide road network as it makes full use of the comprehensive information of the available data and explicitly accounts for the inconsistency issue of these data. It was illustrated in the numerical example that the number and location of sensors have significant impacts on the accuracy of the travel time estimation.

On the basis of the proposed network-wide travel time estimation model, some further extensions can be envisaged as follows: (1) The observed travel time data from two sensor systems have been studied in this paper. It would be useful to extend the proposed model to consider both the observed travel time and traffic flow data collected from different types of sensor systems for improving travel time estimates within the territory-wide road network. (2) As sensor location can influence the accuracy of the travel time estimation, it is required to investigate how to optimize the sensor location so as to minimize the number of sensors as well as enhance the

estimation reliability. (3) It should be noted that multiple solutions may theoretically exist in the proposed model. Further research should be carried out to develop a more promising and robust solution algorithm in order to find a better solution.

Acknowledgements

The work described in this paper was jointly supported by a research grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (No. PolyU 5195/07E), a research grant from Research Committee of The Hong Kong Polytechnic University (Nos. J-BB7Q) and a research grant from National Natural Science Foundation of China (No. 70901073).

References

Guo, H. R. and Jin, J. H. (2006). "Travel time estimation with correlation analysis of single loop detector data", *Transportation Research Record*, 1968, 10-19.

Hu, S. R., Peeta S., and Chu, C. H. (2009). "Identification of vehicle sensor locations for link-based network traffic applications", *Transportation Research Part B*, 43(8-9), 873-894.

Lam, W. H. K., Chan, K. S. and Shi, J. W. Z. (2002). "A traffic flow simulator for short-term forecasting of travel times", *Journal of Advanced Transportation*, 36(3), 265-291.

Lam, W. H. K., Shao, H. and Sumalee, A. (2008). "Modeling impacts of adverse weather conditions on a road network with uncertainties in demand and supply", *Transportation Research Part B*, 42(10), 890-910.

Liu, H. X. and Ma, W. T. (2009). "A virtual vehicle probe model for time-dependent travel time estimation on signalized arterials", *Transportation Research Part C*, 17(1), 11-26.

Margulici, J. D. and Ban, X. (2008). "Benchmarking travel time estimates", *IET Intelligent Transport Systems*, 2(3), 228-237.

Robinson, S. and Polak, J. W. (2005). "Modeling urban link travel time with inductive loop detector data by using the k-NN method", *Transportation Research Record*, 1935, 47-56.

Tam, M. L. and Lam, W. H. K. (2008). "Using automatic vehicle identification data for travel time estimation in Hong Kong", *Transportmetrica*, 4(3), 179-194.

Taylor, M. A. P., Woolley, J. E. and Zito, R. (2000). "Integration of the global positioning system and geographical information systems for traffic congestion studies", *Transportation Research Part C*, 8(1-6), 257-285.

Yeon, J., Elefteriadou, L. and Lawphongpanich, S., (2008). "Travel time estimation on a freeway using Discrete Time Markov Chains", *Transportation Research Part B*, 42(4), 325-338.

A Robust Kalman Filter Based Data Fusion Method for Traffic Speed Estimation

Depin Peng[1], Jianping Wu, Chunlu Wang and Xingquan Zuo

[1] Ph.D. candidate, School of Computer, Beijing University of Posts and Telecommunications, NO. 10 Xitu Cheng Road, Haidian District, Beijing, China 100876, Email: depinpeng@gmail.com

Abstract

Traffic data is very important for Intelligent Transportation Systems. The traffic data is collected by different types of detectors. However, the data collected by these detectors cannot meet the precision and completeness requirement of ITS. To improve traffic data's completeness and comprehensiveness, this paper addresses a problem when road sections have no GPS sampling signals based on floating vehicle systems. A robust Kalman Filter fusion algorithm is proposed to fuse the information of associated road sections, of which traffic information is known, and to obtain the traffic information of the road sections without GPS sampling signals. The method is verified with the real GPS traffic data of Hangzhou city, China, and compared with the standard Kalman Filter algorithm. Finally, experimental results show that the proposed fusion method is effective and can provide more precise and comprehensive traffic information for traffic managers.

Key Words: Intelligent Transportation Systems; Data Fusion; Robust Kalman Filter; Floating Vehicle

Introduction

The basic traffic data collected by floating vehicle information management systems plays an important role in Intelligent Transportation Systems (ITS).The traffic data may be collected by many kinds of detectors, such as GPS, loop detector, ultrasonic wave detector, infrared detector, microwave detector, and video detector. However, the format and content of the data collected by these detectors are very different. And the data needed by the floating vehicle information management systems includes many multi-level aspects. In order to completely evaluate the traffic parameters of a road section, it is necessary to fuse the different data sources to avoid single data

source failure and improve the precision, especially under the circumstance of road sections without GPS sampling signals. The precision of traffic state estimation, evaluation, and decision making greatly depend on the reliability, completeness, and precision of the basic traffic data. So, the study looks specifically at the method of processing traffic data for the floating vehicle information management system and researches how to use the speed information of associated road sections to acquire the speed information of the road sections without GPS sampling signals.

As all known, for linear dynamic systems with white process and measurement noise, the Kalman filter is known to be the minimum variance linear state estimator. In the case that the random quantities are Gaussian, then the Kalman filter is the minimum variance state estimator (Chia et al., 2006). However, in the application of Kalman filters there is often a known model or signal information that is either ignored or dealt with heuristically. When the process and measurement noise are not Gaussian, the Kalman filter is not the optimal filter any more. In order to solve the problem in this paper, a robust Kalman filtering algorithm is proposed.

Literature Review

Information fusion is an increasing study topic recent years. In the ITS field, one aspect of the research of traffic information fusion focuses on the GPS positioning, tracking, and navigation. Atherton and Bather (2005) studied the information fusion from several Kalman filters tracking the same target, examined the accuracy of the Bar-Shalom formula, and compared the method with the optimal least squares estimator. Chin (1994) proposed a method to incorporate a neural network learning mechanism into the standard Kalman filter to reduce the estimation error and improve the multiple-target tracking solution. Fuerstenberg and Dittmer (2003) described a method for ego-motion estimation using the velocities of the objects tracked by Laserscanners, which are able to observe the vehicle's environment. Lobbia and Kent (1994) proposed an approach for fusing the off-board track-level data at a central fusion node, and presented an inequality constraint to get the proper sampling interval for the incoming state estimates to the node. Bin et al. (2003) presented a GPS/INS integrated navigation system based on the theory of multi-sensor data fusion, generated error models for the inertial measurement unit, and developed an improved decentralized Kalman filter for INS to eliminate obvious error of GPS data.

Another aspect of research of traffic information fusion focuses on vehicle assistant systems. Martin and Becker (Jan, 2005) proposed an approach to evaluate the performance of automotive sensor data fusion algorithms, which calculates performance indices on several evaluation levels, and the indices can be arranged in a pyramidal structure since the number of performance indices decreases from low to high evaluation levels. Zhu et al.(2006) developed a robust detection approach integrated dynamic scene modeling, hypothesis testing, and robust information fusion to achieve reliable detection of overtaking vehicles. Its solution is integrated into a monocular vision system onboard for obstacle detection. Xiang and Ozguner

(2005) described a multi-sensor data fusion system for off-road autonomous vehicles to fuse the data from a camera, four laser range finders, and several ultrasonic sensors. The D-S evidence theory was adopted to classify each grid in the fusion map, and a weighted evidence combination rule is proposed to improve the decision results under high confliction circumstance.

Except for the two aspects above, the traffic information fusion is applied to the field of traffic flow fusion, travel time prediction, and occupancy fusion. Bao et al. (2007) studied the application of self-organizing maps algorithm in traffic data fusion from the point of view of route travel time of urban traffic flow guidance and traffic measurement of intersections. These references mainly researched the traffic flow fusion; however the traffic speed fusion is rarely studied, especially in the condition of road sections without GPS sampling signals. In practical floating vehicle information management system, the speed of the road sections without GPS sampling signals is replaced with the historical speed data, which cannot reflect the real traffic speed estimation, so that the accuracy of the traffic state of the road network is influenced. In fact, the problem has not been solved effectively in practices. In this paper, we study how to use the speed information of associated road sections to acquire the speed estimation of the road sections without GPS sampling signals.

Problem Description. Floating vehicle is the car on which a GPS receiver is installed. It is an effective way to collect traffic state information. But using the floating vehicle to collect traffic data always has some disadvantages. One of the main problems is that sometimes there are no GPS sampling signals on some road sections, because the signals are blocked by a viaduct, tunnel, or high buildings, or there is no floating vehicle running on the road sections.

The problem can be described as Figure 1. Assume that the road section x_2 (include uplink and downlink) has no GPS sampling signals. We can't get the traffic information of this road section, while the traffic information of other road sections is known. The problem is how to use the traffic information of correlative road sections to acquire the speed information of the road section x_2 by an information fusion method.

Robust Kalman Filtering. Consider the following discrete-time system:

$$x_{k+1} = [A_k + \Delta A_k]x_k + w_k \tag{1}$$

$$y_k = [C_k + \Delta C_k]x_k + v_k \tag{2}$$

where $x_k \in R^n$ is the state vector, $y_k \in R^p$ is the measured output, $w_k \in R^{q1}$ is the process noise, $v_k \in R^{q2}$ is the measurement noise, w_k and v_k are uncorrelated, zero-mean, white noise signals with unknown variance, A and C are known constant matrices, and ΔA_k and ΔC_k are unknown matrices representing parameter uncertainties. The admissible uncertainties are assumed to be of the form of Eq. (3).

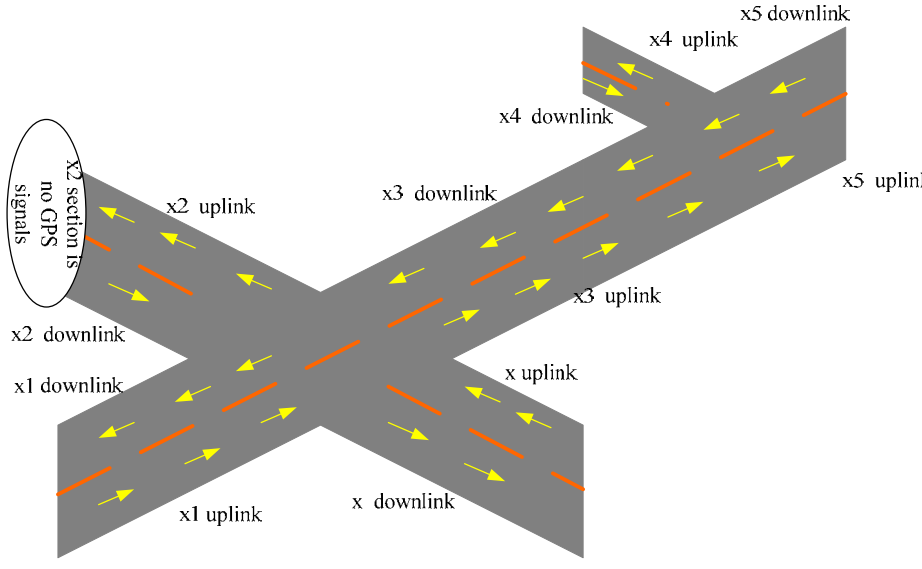


Figure 1. Problem description of traffic speed fusion.

$$\begin{bmatrix} \Delta A_k \\ \Delta C_k \end{bmatrix} = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix} F_k E \tag{3}$$

where $F \in \Re^{i \times j}$ is an unknown matrix satisfying

$$\| F \| \leq I \tag{4}$$

and H_1, H_2, and E are known constant matrices of appropriate dimension that specify how the uncertain parameter in F enters the nominal matrices A and C. Throughout this paper, $\| X \|$, for $X \in \Re^{m \times n}$, denotes its largest singular value.

And our purpose is designing the state estimator as Eq. 5 and applying for traffic speed estimation.

$$\hat{x}_{k+1} = \hat{A}_k \hat{x}_k + \hat{K}_k (y_k - C_k \hat{x}_k) \tag{5}$$

where $\hat{A}_k, \hat{K}_k$ are the matrices to be determined.

Assuming the noise statistics satisfy:

$$E[w_k] = 0, E[v_k] = 0, \; E\{\begin{bmatrix} w_k \\ v_k \end{bmatrix} [w_k^T \quad v_k^T]\} = \begin{bmatrix} Q_k & 0 \\ 0 & R_k \end{bmatrix} = \tilde{\Phi}_k \tag{6}$$

For all admissible uncertainties of F_k, exist a positive definite matrix θ_k that satisfies: $E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T] \leq \theta_k, \forall k$. Therefore, redefine a new state vector:

$$\tilde{x}_k = \begin{bmatrix} x_k \\ \hat{x}_k \end{bmatrix} \tag{7}$$

Then design the augmented filter as following:

$$\widetilde{x}_{k+1} = (\widetilde{A}_k + \widetilde{H}_k F_k \widetilde{E}_k)\widetilde{x}_k + d_k \tag{8}$$

where

$$d_k = \begin{bmatrix} w_k \\ v_k \end{bmatrix}, \widetilde{A}_k = \begin{bmatrix} A_k & 0 \\ \hat{K}_k C_k & \hat{A}_k - \hat{K}_k C_k \end{bmatrix}, \widetilde{H}_k = \begin{bmatrix} H_{1,k} \\ \hat{K}_k H_{2,k} \end{bmatrix}, \widetilde{E}_k = \begin{bmatrix} E_k & 0 \end{bmatrix} \tag{9}$$

Denote the state covariance matrix of Eq. (7) by

$$\widetilde{\Sigma}_k = E[\widetilde{x}_k \widetilde{x}_k^T] = E\{\begin{bmatrix} x_k \\ x_k \end{bmatrix}\begin{bmatrix} x_k \\ x_k \end{bmatrix}^T\} \tag{10}$$

According to Eq. 8 and Eq. (10), the Lyapunov equation that governs the solution of the covariance matrix $\widetilde{\Sigma}_k$ can be represented as:

$$\widetilde{\Sigma}_{k+1} = (\widetilde{A}_k + \widetilde{H}_k F_k \widetilde{E}_k)\widetilde{\Sigma}_k(\widetilde{A}_k + \widetilde{H}_k F_k \widetilde{E}_k)^T + \widetilde{\Phi}_k \tag{11}$$

where $\widetilde{\Phi}_k$ is defined in Eq. 6. The initial value is $\widetilde{\Sigma}_0$, for later use, we set $\begin{bmatrix} I & 0 \end{bmatrix}\widetilde{\Sigma}_0\begin{bmatrix} I & 0 \end{bmatrix}^T = S_0$, $\begin{bmatrix} I & -I \end{bmatrix}\widetilde{\Sigma}_0\begin{bmatrix} I & -I \end{bmatrix}^T = E[(x_0 - \hat{x}_0)(x_0 - \hat{x}_0)^T] := S_1$. It is noted that uncertainty F_k appears in Eq. (1). Therefore, it is impossible to give the exact value of the covariance matrix $\widetilde{\Sigma}_k$. An alternative way is to find a set of upper bounds for $\widetilde{\Sigma}_k$ and then obtain the minimum with respect to the filter parameters $\hat{A}_k$ and $\hat{K}_k$. So, the algorithm procedure is shown in Figure 2.

1. Define state variable $\widetilde{x}_k = \begin{bmatrix} x_k \\ \hat{x}_k \end{bmatrix}$;
2. Design the filter: $\widetilde{x}_{k+1} = (\widetilde{A}_k + \widetilde{H}_k F_k \widetilde{E}_k)\widetilde{x}_k + d_k$; where $d_k = \begin{bmatrix} w_k \\ v_k \end{bmatrix}$, $\widetilde{A}_k = \begin{bmatrix} A_k & 0 \\ \hat{K}_k C_k & \hat{A}_k - \hat{K}_k C_k \end{bmatrix}$, $\widetilde{H}_k = \begin{bmatrix} H_{1,k} \\ \hat{K}_k H_{2,k} \end{bmatrix}$, $\widetilde{E}_k = \begin{bmatrix} E_k & 0 \end{bmatrix}$.
3. Calculate the state covariance matrix : $\widetilde{\Sigma}_k := E[\widetilde{x}_k \widetilde{x}_k^T]$;
4. Find an appropriate upper bound of the state covariance matrix: $\widetilde{\Sigma}_k \le \Sigma_k$;
5. Then, select appropriate $\hat{A}_k$ and $\hat{K}_k$ which make the covariance error minimum;
6. Finally, determine the innovation Kalman filter: $\hat{x}_{k+1} = \hat{A}_k \hat{x}_k + \hat{K}_k(y_k - C_k \hat{x}_k)$; and
7. Let $k \to k+1$, repeat the above steps.

Figure 2. Algorithm procedure.

We now define some useful lemmas.

Lemma 1 (Boyd et al. 1994): Let Y be a symmetric matrix, A, H, and E be real matrices, X be a symmetric positive-definite matrix, and matrix F satisfies $FF^T \le I$. Then

$$(A+HFE)X(A+HFE)^T - Y \le 0 \tag{12}$$

If and only if there exists a constant $\alpha > 0$ such that

$$\alpha^{-1}I - EXE^T > 0, \tag{13}$$

$$A(X^{-1} - \alpha E^T E)^{-1} + \alpha^{-1}HH^T - Y \le 0 \tag{14}$$

Lemma 2 (Xie et al., 1994): Given matrices A, H, E, and F with compatible dimensions such that $FF^T \le I$. Let X be a symmetric positive-definite matrix and $\alpha > 0$ be an arbitrary positive constant such that $\alpha^{-1}I - EXE^T > 0$, then the following inequality holds.

$$(A+HFE)X(A+HFE)^T \le A(X^{-1} - \alpha E^T E)A^T + \alpha^{-1}HH^T \tag{15}$$

Lemma 3 (Zhu, 2002): For $0 \le k \le N$, suppose that $X = X^T > 0$, $f_k(X) = f_k^T(X) \in R^{n\times n}$, and $g_k(X) = g_k^T(X) \in R^{n\times n}$. If there exists $Y = Y^T \ge X$, such that:

$$f_k(Y) \ge f_K(X) \tag{16}$$

and

$$g_k(Y) \ge f_K(Y) \tag{17}$$

then the solutions A_k and B_k to the following difference equations:

$$A_{k+1} = f_k(A_k), B_{k+1} = g_k(B_k), A_0 = B_0 > 0 \tag{18}$$

satisfy $A_k \le B_k$

Remark 2: Lemma 1 is known as the S-procedure technique, which is often utilized to convert the inequality involving norm-bounded uncertainty like Eq. (11) into an equivalent matrix inequality with an extra scalar parameter α. Lemma 2 is a direct result. Lemma 3 will be used to give the upper bound for the covariance matrix $\tilde{\Sigma}_k$.

Now, we compute the parameters $\hat{A}_k$ and $\hat{K}_k$ of filter (5).

Definition 1: The filter (5) is a quadratic filter associated with Σ_k and P_k if there exist symmetric positive-definite matrices Σ_k and P_k ($0 \le k \le N$) such that for all admissible uncertainty F_k satisfying (4), the following equations holds

$$(\tilde{A}_k + \tilde{H}_k F_k \tilde{E}_k)\Sigma_k(\tilde{A}_k + \tilde{H}_k F_k \tilde{E}_k)^T - \Sigma_{k+1} + \tilde{\Phi}_k \le 0, \text{ and } \alpha_k^{-1}I - \tilde{E}_k \Sigma_k \tilde{E}_k^T > 0 \tag{19}$$

Definition 2: The filter (5) is a identity quadratic filter associated with $\sum_k$ and P_k if and only if there exist symmetric positive-definite matrices $\sum_k$ and P_k and $\alpha_k > 0$ ($0 \le k \le N$) such that for all admissible uncertainty F_k satisfying (4), the following equations holds

$$\widetilde{A}_k(\sum_k^{-1} - \alpha_k \widetilde{E}_k^T \widetilde{E}_k)^{-1} \widetilde{A}_k + \alpha_k^{-1} \widetilde{H}_k \widetilde{H}_k^T - \sum_{k+1} + \widetilde{\Phi}_k \le 0, \text{ and } \alpha_k^{-1} I - \widetilde{E}_k \sum_k \widetilde{E}_k^T > 0 \quad (20)$$

$$\text{and } \sum_{k+1} = \widetilde{A}_k(\sum_k^{-1} - \alpha_k \widetilde{E}_k^T \widetilde{E}_k)^{-1} \widetilde{A}_k + \alpha_k^{-1} \widetilde{H}_k \widetilde{H}_k^T + \Phi_k \quad (21)$$

Based on definition 2, we get the conclusion as follows:

$$\text{Theorem 1:} \quad \text{if } \sum_0 = \widetilde{\sum}_0 \text{ and } \widetilde{P}_k \le P_k, \text{ then } \widetilde{\sum}_k \le \sum_k \quad (22)$$

Where $\widetilde{\sum}_k$ and $\sum_k$ satisfy Eqs. (11) and (21), respectively. Then we design the filter with optimized upper bound θ_k .

Theorem 2: Consider (1) and (2). Let $\alpha_k > 0$ be a sequence of positive scalar. If the following two discrete-time Riccati difference equations:

$$\begin{aligned} \theta_{k+1} = &-[\alpha_k^{-1} H_{1,k} H_{2,k}^T + A_k(\theta_k^{-1} - \alpha_k E_k^T E_k)^{-1} C_k^T] \cdot R_{1,k}^T [\alpha_k^{-1} H_{2,k} H_{1,k}^T + C_k(\theta_k^{-1} \\ &- \alpha_k E_k^T E_k)^{-1} A_k^T] + A_k(\theta_k^{-1} - \alpha_k E_k^T E_k)^{-1} A_k^T + \alpha_k^{-1} H_{1,k} H_{1,k}^T + Q_k, \theta_0 = S_1 \end{aligned} \quad (23)$$

$$\text{and } P_{k+1} = \alpha_k^{-1} H_{1,k} H_{1,k}^T + A_k(P_k^{-1} - \alpha_k E_k^T E_k)^{-1} A_k^T + Q_k, P_0 = S_0 . \quad (24)$$

have positive-definite solutions θ_k and P_k such that $\alpha_k^{-1} I - E_k P_k E_k^T > 0$ and $P_k > \theta_k$, then an identity quadratic filter(5) with parameters:

$$\hat{A}_k = A_k + (A_k - \hat{K}_k C_k)\theta_k E_k^T (\alpha_k^{-1} I - E_k \theta_k E_k^T) E_k \quad (25)$$

$$\hat{K}_k = [\alpha_k^{-1} H_{1,k} H_{2,k}^T + A_k(\theta_k^{-1} - \alpha_k E_k^T E_k)^{-1} C_k^T] R_{1,k}^{-1} \quad (26)$$

where $$R_{1,k} = \alpha_k^{-1} H_{2,k} H_{2,k}^T + R_k + C_k(\theta_k^{-1} - \alpha_k E_k^T E_k)^{-1} C_k^T \quad (27)$$

will be such that the state estimation error variance satisfies bounded condition

$$E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T] \le \theta_k, \forall k \quad (28)$$

Moreover, the filter (5) with parameters (Eqs. (25) and (26)) minimizes the bound θ_k .

Remark 3: Similar structure to Σ_k have been used in (Fu et al., 1999) and (Theodor and Shaked, 1996). The choice of this special structure has been motivated by arguments related to the minimization of the upper boundary of the state estimation error variance. Note that the different Riccati Eqs.23 and 24 involve the scalar parameter α_k. Detailed discussions on the feasibility and convergent properties of such kind of difference Riccati equations can be found in (Zhu. X, et al., 2002).

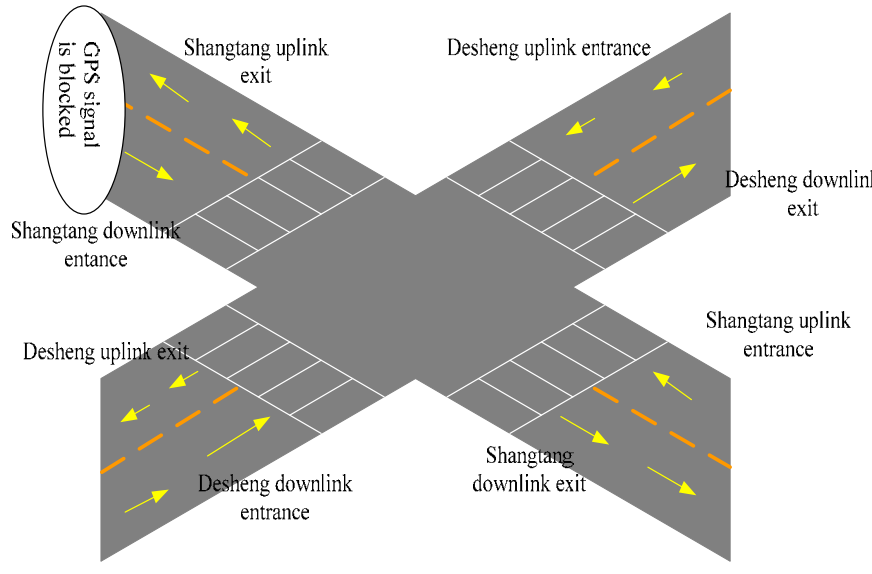


Figure 3. The intersection of Desheng Road and Shangtang Road.

Experimental Results and Analysis. We select the real GPS traffic data of Hangzhou city in 2007. Analyzing the correlation of each road section shown in Figure 1, we find that road section x_2 do not correlate with road section x_4 and road section x_5, and road section x_2 just correlates with road section x_1, x_2, and x_3. So we choose some road sections to verify our fusion method, and select the data from 7:00 am to 9:00 am of May 11, 2007, so there are total 120 data, and the sampling interval of GPS signal is 1 minute.

In the experiment, we research the intersection of Desheng Road and Shangtang Road. The intersection is shown as Figure 3. Suppose that the uplink and downlink of Shangtang Road has no GPS sampling signals, and we consider the speed of this road section as the real speed data. According to the correlation of each road section, we select the average speed of uplink of Desheng Road, uplink of Shangtang Road, and downlink of Desheng Road for the observed values. We consider the average speed of uplink of Shangtang Road as the state variable to be estimated. Choose the Kalman filter parameters by trial-and-error method, obtain the parameters as follows:

$$A = 1,\ H = (1,1,0.91)^T,\ R = \begin{pmatrix} 25 & 0 & 0 \\ 0 & 25 & 0 \\ 0 & 0 & 25 \end{pmatrix}, Q = 25 \tag{29}$$

And consider the following uncertain system:

$$\begin{aligned} x_{k+1} &= [A + \delta_k]x_k + w_k \\ y_k &= Hx_k + v_k \end{aligned} \tag{30}$$

Where the precise boundary on the uncertain parameter δ_k has been assumed to be exactly known, i.e., $|\delta_k| \le 1$. However, in practice, such a boundary can often be obtained from the statistical estimation. It is usually the case that the probability of the uncertain parameter δ_k satisfying $|\delta_k| \le 1$ is close to one, but not equal to one. In this experiment, we assume that $P(|\delta_k| \le 1) = 0.998$ where P means the probability and the statistical law for the uncertain parameter δ_k to satisfy $|\delta_k| > 1$. w_k, v_k are uncorrelated zero Gaussian white noise with the same R and Q as Eq. 29. Note that the system (Eq. 30) is of the form of system (1, 2) with $H_{1,k} = [1 \quad 0 \quad 0]^T$, $H_{2,k} = 0$ and $E_k = [0 \quad 0 \quad 0.03]$. Suppose that the initial conditions are as $x_0 = 0$, $\hat{x}_0 = 0$, $S_0 = 0$ and $S_1 = 0$ .The experiments have been conducted with $\alpha_k = 1.15$. And then we can calculate $\hat{A}_k$ and $\hat{K}_k$ by (Eqs. (25) and (26)) .Use the relative mean error (RME) to evaluate the fusion algorithm:

$$RME = \frac{1}{n}\sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{31}$$

where y_i is the real speed, $\hat{y}_i$ are the fused results. The results are compared with standard Kalman filter in Figure 4, and the fusion errors are shown in Figure 5. The results illustrate the fusion speed based on standard Kalman filter and robust Kalman filter are approximate to real speed. According to Eq. 31, the RME of standard Kalman filter is 19.609%, and that of robust Kalman filter is 12.78%. It is shown that the robust Kalman filter fusion algorithm is better than standard Kalman filter algorithm for acquiring the traffic information of road sections without GPS sampling signals.

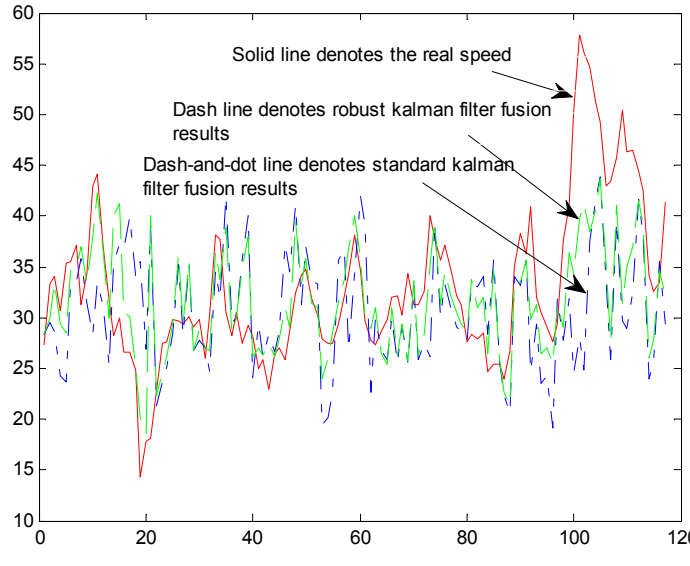


Figure 4. The results comparison with real speed, standard Kalman filter, and robust Kalman filter.



Figure 5. The fusion errors of standard Kalman filter and robust Kalman filter.

Conclusions

The problem of road sections without GPS sampling signals is one of the critical bottlenecks of the floating vehicle information management systems. This paper addressed how to use the traffic information of associated road sections to estimate the speed information. And a robust Kalman filter algorithm was designed to fuse and obtain more precise and comprehensive traffic information. The robust Kalman filter algorithm was verified with the real GPS traffic data of Hangzhou city of China in 2007, and the fusion results were compared with the standard Kalman filter algorithm. The error of the fusion results was relatively small, which proved that the traffic speed information fusion based on robust Kalman filter algorithm is effective; the robust Kalman filter is more robust than Kalman filter for uncertainties of system parameters and noises. We concluded that the information fusion method proposed in this paper can provide more precise and comprehensive traffic information for traffic managers to improve the efficiency of the floating vehicle information management systems.

References

Atherton, D. P. and Bather, J. A. (2005). "Data fusion for several Kalman filters tracking a single target", *IEE Proc-Radar Sonar Navigation*, 152(5), 372-376.

Bao, L., Hu, J., and Qing, G. (2007). "Application of self-organizing maps in traffic data fusion", *14th World Congress on ITS*,paper ID.3015.

Boyd, S., Ghaoui, L. E., Feron, E. and Balakrishnan, V. (1994). "Linear matrix inequalities in system and control theory", Philadelphia, PA: *SIAM, SIAM Studies in Applied Mathematics.*

Chia, T. L., Simon, D. and Chizeck. H. J.(2006). "Kalman filtering with statistical state constraints", *Control and Intelligent Systems,*34(1),73-79.

Chin, L. (1994). "Application of neural networks in target tracking data fusion", *IEEE Transactions on Aerospace and Electronic Systems,*30(1), 281-287.

Fu, M., de Souza, C. E. and Luo, Z. (1999). "Finite horizon robust Kalman filter design" , *in Proc. IEEE Conf. Decision Control,* Phoenix, AZ, 4555-4560.

Fuerstenberg, K. Ch. and Dittmer, M. (2003). "Data fusion and ego-motion estimation using laserscanners" , *Intelligent Transportation Systems proceedings,IEEE,*2, 1224-1228.

Lobbia, R. and Kent, M. (1994). "Data fusion of decentralized local tracker outputs" , *IEEE Transactions on Aerospace and Electronic Systems,* 30(3), 787-799.

Martin, S., and Becker, J. C. (2005). "Performance evaluation of automotive sensor data fusion", *Proceedings of the 8th IEEE Conference on Intelligent Transportation Systems,* 613-618.

Theodor, Y., and Shaked, U. (1996). "Robust discrete-time minimum-variance filtering", *IEEE Trans. Signal Processing*, 44, 181-189.

Wang, B., Wang, J., Wu, J., and Cai, B. (2003). "Study on adaptive GPS INS integrated navigation system", *IEEE Proc on Intelligent Transportation Systems 2003,* 2, 1016-1021.

Xie, L., Soh, Y. C., and de Souza, C. E. (1994). "Robust Kalman filtering for uncertain discrete-time systems", *IEEE Trans. Automat. Contr.,* 39, 1310-1314.

Xiang, Z., and Ozguner, U. (2005). "Environmental perception and multi-sensor data fusion for off-road autonomous vehicles", *Proceedings of the 8th IEEE Conference on Intelligent Transportation System,* 420-425.

Zhu, X., Soh, Y. C., and Xie, L. (2002). "Design and analysis of discrete-time robust Kalman filters", *Automatica,* 38, (6), 1069-1077.

Zhu, Y., Comaniciu, D., Pellkofer, M., and Koehler, T. (2006). "Reliable detection of overtaking vehicles using robust information fusion", *IEEE Transactions on Intelligent Transportation Systems*, 7, (4), 401-414.

A new Concept of Level of Service for Freeway Weaving, Merge and Diverge Segments and Its Applications

Ning Wu[1]

[1] Professor, Institute for Traffic Engineering, Ruhr University Bochum (IA 2/126) D-44780, Bochum, Germany; Email: ning.wu@rub.de

Abstract

In most of the existing highway capacity manuals, the traffic quality of critical areas (weaving, merge and diverge segments) at intersections on freeways is investigated separately. A total assessment of an intersection is therefore not available. This paper presents a new model considering the desired traffic quality of the total intersection. A combined degree of saturation which is defined as a function the degrees of saturation of the critical areas is used for defining the traffic quality. This model is already incorporated in the upcoming German Highway Capacity Manual (HBS2011).
Key words: capacity; ramps junctions; weaving areas; concept of level of service.

Introduction

In the existing highway capacity manuals, e.g. the US Highway Capacity Manual (HCM2000, TRB 2001) and the Germany Highway Capacity Manual (HBS2001, FGSV 2001), Level of Service (LOS) at intersections on freeways (weaving, merge and diverge segments) is traditionally defined by the flow speed, density, or rate in critical area of merging, diverging and weaving manoeuvre. As a Measure of Effectiveness (MOE), the flow speed in the critical areas under consideration is applied. The service speed is then correlated with the predefined LOS. The service flow rate or density can be then obtained according to the existing flow-speed-density relationship.

That traditional concept of LOS for intersections on freeways investigates the traffic quality of different areas (cf. the circles in Figure 1) separately. Normally, the traffic

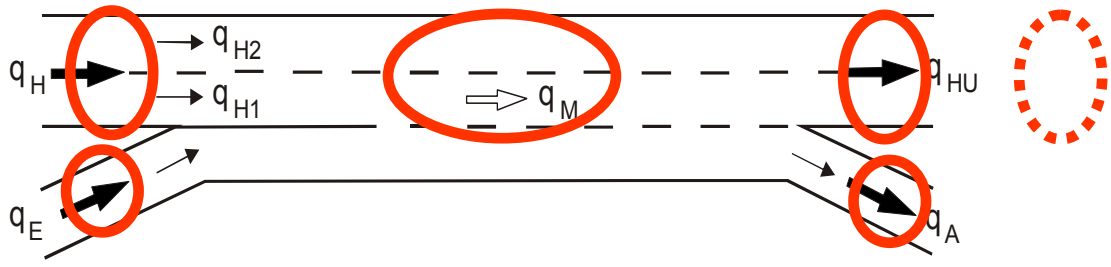


Figure 1. Critical areas at intersection on freeways.

quality of the so-called critical area (merging area for on-ramps, diverging area for off-ramps, and weaving area for weaving sections) is most important. The flow rate of this critical area (cf. the area for q_M in Figure 1) is predominately used for assessing the traffic quality of the intersection. The LOS of this critical area is normally a function of this critical flow (q_M) rate and thus a function of the flow rate on the major road, in the on-ramp (merge segment/entrance), and in the off-ramp (diverge segment/exit). Depending on the geometrical configuration, e.g. the number of lanes on the major road and in the ramps as well the length of the critical area (merging/diverging/weaving area), this functional relationship can vary. Nevertheless, those relationships are clearly defined if the geometrical configurations are known.

As mentioned previously, the traditional concept of LOS only considers the traffic situation in the major flow and in the in- and out-flow separately. The intersection (weaving, merge or diverge segments) is not treated as a whole object. An overall assessment of the intersection is not possible or very difficult.

For example, in the HBS2001, Chapter 4, the traffic quality at intersections on freeways is determined by examining the capacity on the major road, both directly upstream (qH) and downstream (qHU) of the intersection, in the on-ramp (qE), and in the off-ramp (qA) separately (indicated in as red circles in Figure 1). Sometimes it is also needed to check the capacity of the freeway section more distantly downstream of the intersection (doted red circles in Figure 1) because the normal freeway section can have lower capacity as at the intersection due to different speed level and lane utilisation.

Table 1. Degree of saturation for defining LOS in HBS2001.

LOS	degree of saturation x_{LOS} [-]
A	≤ 0,30
B	≤ 0,55
C	≤ 0,75
D	≤ 0,90
E	≤ 1
F	-

The desired traffic quality of the total intersection can only be achieved if all of the mentioned areas are examined to be of necessary quality. The traffic quality of the worst area is then decisive for the total intersection. For example, the capacity of an on-ramp intersection (entrance) must be examined in three areas: a) the capacity CM of the merging area, b) the capacity CE of the on-ramp, and c) the capacity CH of the upstream major street (cf. the dotted polygon Figure 3). In the HBS2001, the degree of saturation x is used for defining the Level of Service (cf. Table 1). If the capacities of the different areas are given, the LOS of those areas can be obtained according to the corresponding degree of saturation.

The New Model

In this paper, a new model is introduced for determining the LOS at intersections on freeways. The new model considers the intersection as a whole object. According to the probability and queuing theory, the degree of saturation of the whole intersection can be considered as a combination of degrees of saturation in the different areas under consideration.

The theoretical consideration is presented as following.

In general, for a road section with n traffic lanes, the combined total degree of saturation of the total road section is equal to the probability that there are vehicles on all available lanes simultaneously. On the other side, the probability that there is a vehicle on a certain lane is equal to the degree of saturation of the lane. Thus, we have

$$x_{total} = \prod_{i=1}^{n} x_{lane,i}$$

In case of a unique flow distribution on all lanes we have $x_{lane,i} = x_{lane} = \text{const.}$ and then

$$x_{total} = x_{lane}^{\,n}$$

At a merge or diverge point the degree of saturation has to be the same from both sides of the merge or diverge point (Wu, 1997). Thus, the follow equations hold (cf. the dotted polygon Figure 3) for the case that the traffic flows are uniquely distributed on all road sections under consideration:

$$x_E^{\,n_E} + x_H^{\,n_H} = x_{ME}^{\,n_{ME}} \quad \text{and} \quad x_M^{\,n_M} = x_A^{\,n_A} + x_{HU}^{\,n_{HU}}$$

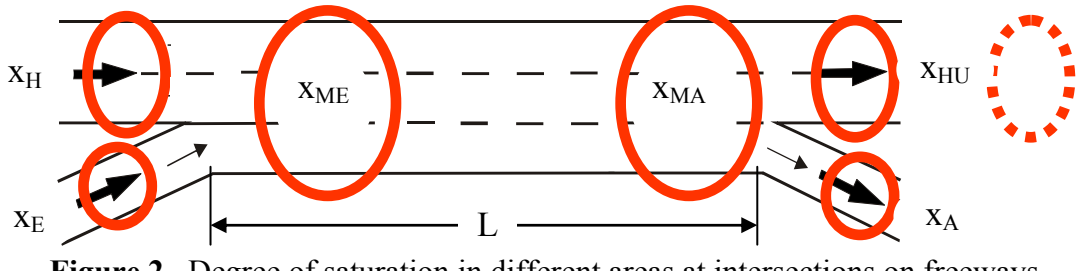


Figure 2. Degree of saturation in different areas at intersections on freeways.

In realty the traffic flow is seldom uniquely distributed on all lane because different traffic behaviours and boundary conditions. The power n to the degree of saturation x is normally not equal to the number of lanes n. In general the following equations always hold:

$$x_E^{a1} + x_H^{b1} = x_{ME}^{c1} \quad \text{and} \quad x_M^{c2} = x_A^{a2} + x_{HU}^{b2}$$

The parameters a1, a1, b1, and a2, b2, c2 can be calibrated against field data.

If there are two intersections in succession (cf. Figure 3, area ME and M), the first one (downstream) can impact the following one (upstream) significantly. The impedance depends mainly on the degree of saturation at the first intersection and the distance between both intersections. According the queuing theory the second intersection is occupied by the queue caused by first intersection if the queue length N behind the first intersection is large than the storage places between both intersections. For example, in case of an M/M/1 queuing system the probability that the second intersection is impact by the first one is (cf. Figure 3)

$$Pr_{imp,ME} = Pr(N > N_L) = x_M^{N_L+1}$$

where x_M is the degree of saturation of the first intersection, N_L is the available storage places between both intersections. For uniquely distributed traffic flow is

$$N_L = L \cdot k_{L,opt} \cdot n_L = \frac{L}{v_{L,opt}} \cdot C_L \cdot n_L$$

where L is the length between both intersections, $k_{L,opt}$, $v_{L,opt}$, n_L, and C_L are the optimum density and speed at capacity, the number of lanes, and the capacity of the road section between both intersections.

As an approximation, the following equation for a general queuing system holds:

$$Pr_{imp,ME} = Pr(N > N_L) = x_M^{f(N_L+1)}$$

The function f(NL+1) is monotonically increasing against the number of available storage places NL and therefore against the length of the road section L between both intersections.

The actual combine total degree of saturation of the second intersection - here, it corresponds to the degree of saturation of a weaving area - is then (cf. Figure 3)

$$x_{MV,total} = x_{MV}^{c3} = x_{ME}^{c1} + \left(x_M^{c2}\right)^{f(L+1)} = \left(x_E^{a1} + x_H^{b1}\right)^{c1} + \left(x_A^{a2} + x_{HU}^{b2}\right)^{c2 \cdot f(N_L+1)}$$

Using the derived equations, the combined degree of saturation for different intersections, e.g. for a grid of road network, can be estimated. However, a huge among of field data would be necessary for the calibration work.

Following the logic above, the relationships between the degrees of saturation in the merge, diverge and weaving segments can be established:

- *In the merging and diverging area*: The degree of saturation in the critical areas (areas ME and M in Figure 2) can be expressed as the superposition of the degrees of saturation in the merging or diverging flows. That is,

$$x_E^{a1} + x_H^{b1} = x_{ME}^{c1} \quad \text{and} \quad x_M^{c2} = x_A^{a2} + x_{HU}^{b2}$$

  The superposition can be calibrated by the parameters a1, a1, b1, and a2, b2, c2 in order to taken into account the geometrical configuration of the intersection. Normally, those parameters correlated strongly to the number of lanes in the road sections under consideration.
- *In the weaving area*: The weaving area can be considered as a combination of the merging and the diverging area. Again, the degree of saturating in the weaving area can be expressed as the superposition of the degrees of saturation in the merging and diverging areas. That is,

$$x_{MV}^{c3} = x_{ME}^{c1} + \left(x_M^{c2}\right)^{f(N_L+1)} = \left(x_E^{a1} + x_H^{b1}\right)^{c1} + \left(x_A^{a2} + x_{HU}^{b2}\right)^{c2 \cdot f(N_L+1)}$$

This superposition can be calibrated by the parameters c3 and a function $f(N_L+1)$ which is monotonically increasing against the length of the weaving area L.

The model parameters a1, a2, b1, b2, c1, c2, c3, and the function f(NL+1) can be calibrated with field data. Those parameters are normally functions of the geometrical configurations, such as the number of lanes in the legs of intersection (major road/on-ramp/off-ramp), and the length of the accelerating (or decelerating, weaving) lanes etc.

Application in the HBS

In case of L = infinite one obtains $\left(x_M^{c2}\right)^{f(N_L+1)} => 0$ and $x_{MV}^{c3} = x_{ME}^{c1}$. The weaving area is transferred into an on-ramp intersection. In the HBS2001, the length of the weaving area is predefined as a constant for a standard design and thus is not more taken into account in the calculation. The difference between a weaving area (weaving section) and a merging area (on-ramp intersection) is then considered by different model parameters.

Normally, the values of the model parameters a, b, c are different and they are dependent on geometrical configurations of the corresponding legs (the major road, the on-ramp, or the off-ramp) of the intersection under investigation. As an approximation and simplification, we can assume a = b = c. Thus, we have

$$x_{ME}^{a} = x_E^{a} + x_H^{a} \quad \text{or} \quad x_E = \sqrt[a]{x_{ME}^{a} - x_H^{a}} = x_{ME} \cdot \sqrt[a]{1 - \left(\frac{x_H}{x_{ME}}\right)^{a}}$$

and

$$x_M^{a} = x_A^{a} + x_{HU}^{a} \quad \text{or} \quad x_A = \sqrt[a]{x_M^{a} - x_{HU}^{a}} = x_M \cdot \sqrt[a]{1 - \left(\frac{x_{HU}}{x_M}\right)^{a}}$$

At capacity (x = 1) we have for the merging and diverging area

$$x_{ME}^{\ a} = x_E^{\ a} + x_H^{\ a} = 1 \text{ or } x_E = \sqrt[a]{1 - x_H^{\ a}}$$

and

$$x_M^{\ a} = x_A^{\ a} + x_{HU}^{\ a} = 1 \text{ or } x_A = \sqrt[a]{1 - x_{HU}^{\ a}}$$

This corresponds to

$$q_{E,max} = C_E \cdot \sqrt[a]{1 - \left(\frac{q_H}{C_H}\right)^a} \quad \text{and} \quad q_{A,max} = C_A \cdot \sqrt[a]{1 - \left(\frac{q_{HU}}{C_{HU}}\right)^a}$$

C_A, C_E, C_H and C_{HU} are the capacities of the off-ramp, the on-ramp, the upstream major road, and the downstream major road.

As an example, Figure 3 depicts the typical shape of the proposed function for an on-ramp intersection. Is can be seen, that the boundary conditions

$$q_E = C_E \text{ at } q_H = 0 \text{ and } q_E = 0 \text{ at } q_H = C_H$$

are satisfied. Similarly, also the following conditions hold:

$$q_A = C_A \text{ at } q_{HU} = 0 \text{ and } q_E = 0 \text{ at } q_{HU} = C_{HU}$$

Varying the model parameter a the function can be fit to the exiting capacity models for the merging/weaving area (cf. CM in Figure 3 for the example of an on-ramp intersection). Normally, the values of the model parameter a depend on the flow rate qA in the following off-ramp (in case of a weaving section), the number of lanes on the major road, and the number of lanes in the on-ramp. For a diverging area, the values of the model parameter a depend predominately on the number of lanes on the major road and in the off-ramp. In addition, the values of the model parameter a depend on the length L of the accelerating/decelerating/weaving section. However, for intersections on freeways with a standard design, the length L can be considered as a constant and can be then neglected. Furthermore, in the HBS2001, only a blanket exiting flow rate qA is considered. Thus, the model parameter a is actually only dependent on the geometrical configuration, that is, on the predefined type of intersections.

If the degrees of saturation on the major road (xH) and in the ramps (xA, xE) are given, the combined degree of saturation of the total intersection (here xME and xM) can be directly calculated as follows:

$$x_{ME} = \sqrt[a]{x_E^{\ a} + x_H^{\ a}} \text{ and } x_M = \sqrt[q]{x_A^{\ a} + x_{HU}^{\ a}}$$

The total LOS of the intersection can be then obtained according to the Table 1.

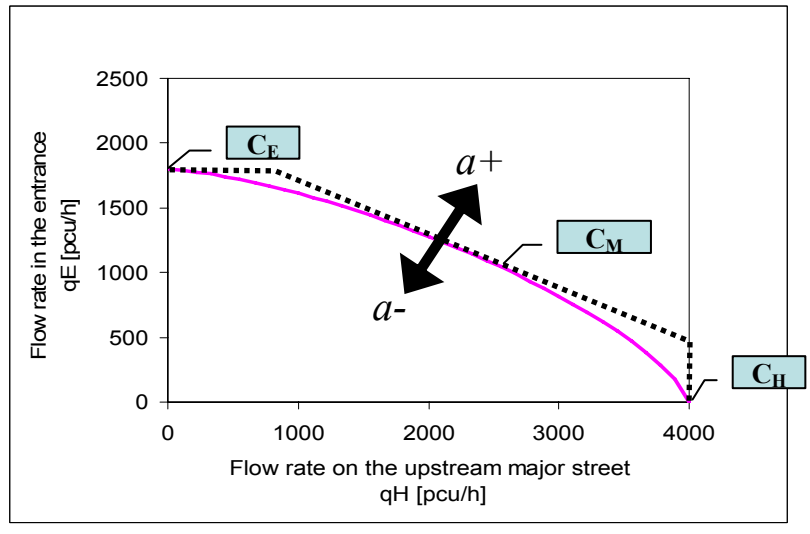


Figure 3. Fitting the new function to the capacity limits of different areas under investigation.

If the saturation degree of LOS (cf. Table 1) for the intersection is given (xME,LOS and xM,LOS), one can calculated the allowed degrees of saturation on the major road (xH) and in the ramps (xE and xA) with the following equations:

$$x_E^{\,a} + x_H^{\,a} = x_{ME,LOS}^{\,a} \quad \text{and} \quad x_A^{\,a} + x_{HU}^{\,a} = x_{M,LOS}^{\,a}$$

Those equations yield

$$x_E = x_{ME,LOS} \cdot \sqrt[a]{1 - \left(\frac{x_H}{x_{ME,LOS}}\right)^a}$$

and

$$x_A = x_{M,LOS} \cdot \sqrt[a]{1 - \left(\frac{x_{HU}}{x_{M,LOS}}\right)^a}$$

This corresponds to

$$q_E = C_E \cdot x_{ME,LOS} \cdot \sqrt[a]{1 - \left(\frac{q_H}{C_H \cdot x_{ME,LOS}}\right)^a} = C_{E,LOS} \cdot \sqrt[a]{1 - \left(\frac{q_H}{C_{H,LOS}}\right)^a}$$

and

$$q_A = C_A \cdot x_{M,LOS} \cdot \sqrt[a]{1 - \left(\frac{q_{HU}}{C_{HU} \cdot x_{M,LOS}}\right)^a} = C_{A,LOS} \cdot \sqrt[a]{1 - \left(\frac{q_{HU}}{C_{HU,LOS}}\right)^a}$$

With those two functions, the traffic quality (LOS) for different values of qH and qE (or qA respectively) can be obtained directly. Figure 4 depicts a typical shape of the new function for an on-ramp intersection.

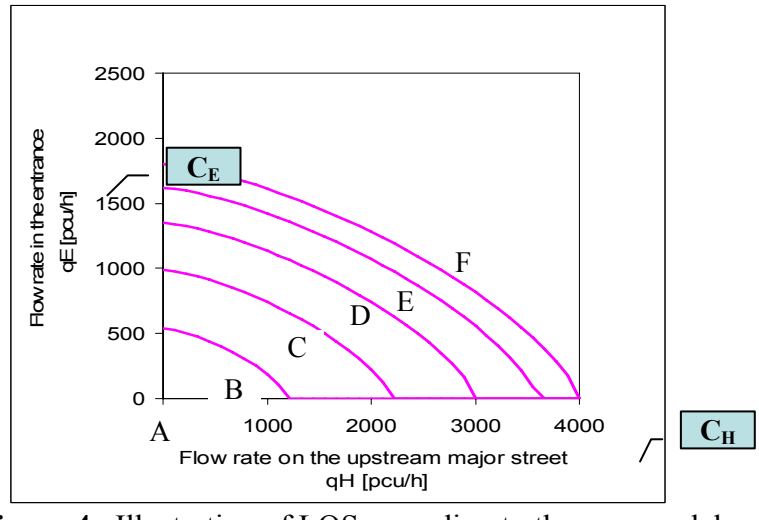


Figure 4. Illustration of LOS according to the new model.

It can be clearly seen, that at qH = 0 (for the y-intercept) and at qE = 0 (for the x-intercept) the traffic quality of the on-ramp and the traffic quality of the major road are decisive for the total intersection.

The new model is incorporated into the new edition of the German Highway Capacity Manual (HBS2011). Compared to the model in the old edition (HBS2001), the new model has following advantages:

- A uniform function for all type of intersections on freeways
- One step traffic quality assessment for three areas:
  - on-ramp / off-ramp
  - major road upstream / downstream
  - critical area (merging/diverging/weaving)
- All boundary conditions satisfied, and
- Easy calibration

In Germany, the capacity and performance of entrances, weaving areas and exits has been studied by Westphal (1995). Although this investigation focused on urban freeways with speed limits of 80 or 100 km/h, Westphal's procedures are also recommended by the HBS2001 for all types of freeways. Based the data of HBS2001, the new model is calibrated.

Using the new model, graphics for all types of on-ramps (3 types), weaving sections (2 types), and off-ramps (7 types) are constructed in the new edition of the German Highway Capacity Manual (HBS2011).

The capacity for diverge segments (exits) is expressed by the following equation:

$$q_{a,LOS} = C_{a,LOS} \cdot \sqrt[a]{1 - \left(\frac{q_h}{C_{h,LOS}} \right)^a} \quad (1)$$

with $q_{a,LOS}$ = maximum flow of the diverge segments (exit) by a given LOS

$C_{h,LOS} = C_h \cdot x_{LOS}$

$C_{a,LOS} = C_a \cdot x_{LOS}$

Ch = capacity of the major freeway section (Table 2)

Ca = capacity of the road section at the exit (Table 2)

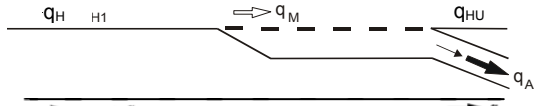
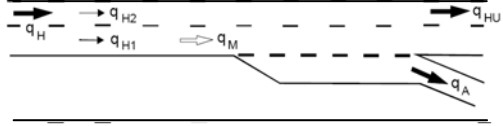
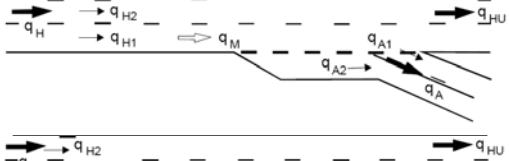
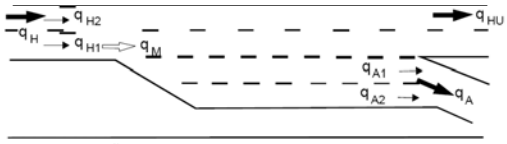
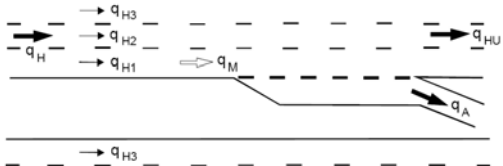
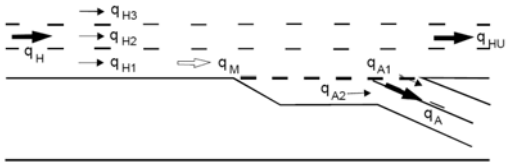
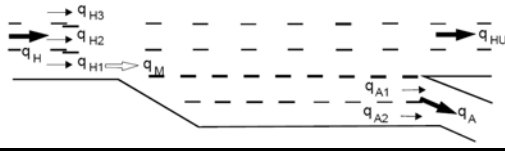
a = model parameter (Table 2)

x_{LOS} = degree of saturation for a given LOS (Table 1)

The capacity for merge segments (entrances) and weaving areas is expressed by the equation

$$q_{e,LOS} = C_{e,LOS} \cdot \sqrt[a]{1 - \left(\frac{q_h}{C_{h,LOS}}\right)^a} \tag{2}$$

Table 2. Parameters for diverge segments (equation (1)).

Type of Exits	C_h	C_a	A
	2000	1800	1,05
	4000	1800	1,9
	4000	3060	1,9
	4000	3600	1,9
	5700	1800	1,9
	5700	3060	1,9
	5700	3600	1,9

with $q_{e,LOS}$ = maximum flow of the merge segment (entrance) by a given LOS

$C_{h,LOS} = C_h \cdot x_{LOS}$

$C_{e,LOS} = C_e \cdot x_{LOS}$

C_h = capacity of the major freeway section (Table 3)

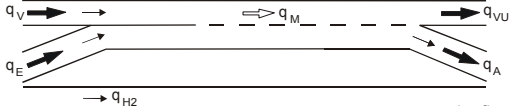
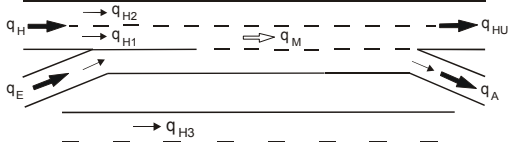
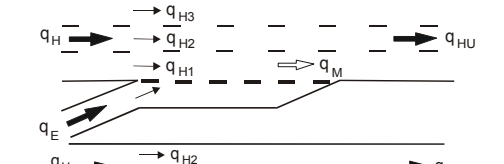
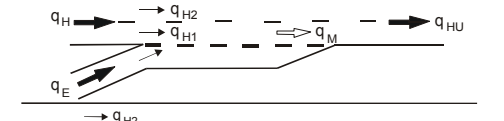
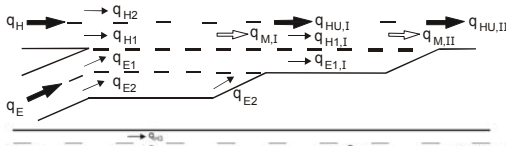
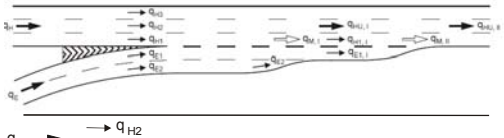
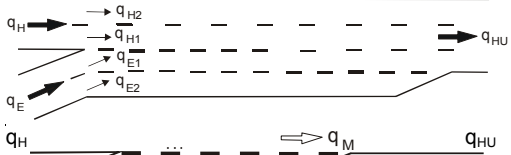
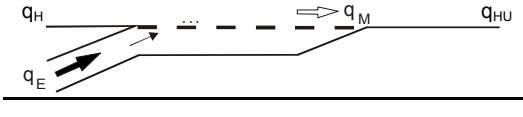
C_e = capacity of the road section at the entrance (Table 3)

a = model parameter (Table 3)

x_{LOS} = degree of saturation for a given LOS (Table 1)

According to the equations (1) and (2), the LOS for the on-, off-ramps, or the weaving areas can be defined depending on the traffic volume on the major freeway section and the entrance/exit. As an example, Figure 5 shows a picture depicting the equation (2) for a 1-lane entrance to a 3-lane freeway. Similar pictures are constructed for all types of entrances, exits, and weaving areas in the proposed HBS2011. The users can use those pictures directly for estimating traffic flow quality

Table 3. Parameters for merge and weaving segments (equation (2)).

Type of Entrance or Weaving Area	C_h	C_e	a
	2000	1800	1,4
	4000	1800	1,3
	5700	1800	1,9
	4000	1800	1,5
	4000	2700	1,5
	5700	3800	1,5
	4000	3800	1,9
	2000	1800	1,2

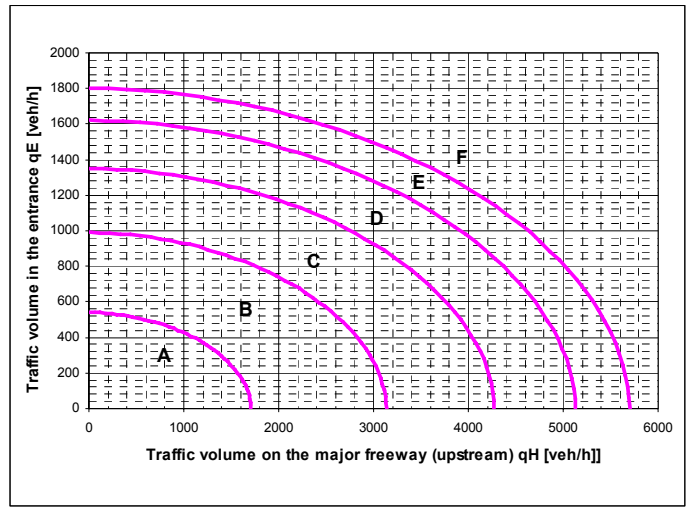


Figure 5. An Example for the Definition of LOS (equation (2)) for a 1-lane entrance to a 3-lane freeway.

Summary and Conclusions

According to the probability and queuing theory, a new model considering the combined degree of saturation for the whole intersection on freeways is derived. The model has a uniform function for all type of intersections on freeways. The traffic quality can be then assessed in one step for the three areas under consideration: a) on-ramp/off-ramp, b) major road upstream/downstream, and c) critical area (merging/diverging/weaving). According to the new model all boundary conditions can be satisfied. Furthermore, because of the simple model parameters, the new model can be easily calibrated. The new model delivers a very useful procedure for determining traffic quality at intersections on freeways. This procedure is always incorporated in the new German Highway Capacity Manual (HBS2011).

References

FGSV (2001). "Handbuch für die bemessung von straßenverkehrsanlagen (HBS 2001)." Forschungsgesellschaft für Straßen- und Verkehrswesen (Hrsg.), Nr. 299, FGSV Verlag GmbH, Köln.

TRB (2000). "Highway capacity manual (HCM 2000)." TRB, National Research Council, Washington, D.C.

Westphal, S. (1995). "Bemessung von Einfahrten an planfreien Knotenpunkten. Institute for Transportation and Traffic Engineering", Ruhr-University Bochum, No. 13.

Wu, N. (1997)."Capacity of shared-short lanes at unsignalized intersections." Proceedings of the Third International Symposium on Intersections without Traffic Signals, Portland, Oregon.

A Simulation Study of Environmental Impact of Automated Truck Lanes on Intercity Expressways

Xun Sun[1], Tomio Miwa[2], Toshiyuki Yamamoto[3] and Takayuki Morikawa[4]

[1] Ph.D candidate, [3] Dr. Eng., [4] Ph.D., Graduate School of Environmental Studies, Nagoya University, Furo-cho, Chikusa-ku, Nagoya Japan464-8603, Email: sunxun@trans.civil.nagoya-u.ac.jp, yamamoto@civil.nagoya-u.ac.jp, morikawa@nagoya-u.jp
[2] Dr. Eng., EcoTopia Science Institute, Nagoya University, Furo-cho, Chikusa-ku, Nagoya Japan 464-8603, e-mail: miwa@civil.nagoya-u.ac.jp

Abstract

Automated truck lanes (ATL) can solve some future problems caused by greater freight transportation demands. This paper proposes an integrated microscopic-mesoscopic simulation framework for precisely estimating the environmental impact of ATL systems on large traffic networks. Microscopic car-following and lane-changing models are extended to reflect the detailed interactions between cars and trucks. Fundamental experiments on one segment of an ATL show that the ATL system is the best of the various available strategies. Finally, Analysis on large traffic networks indicates that the ATL system has the lowest environmental impact among the cases, with the lowest CO_2 emissions not only on the expressway itself but also on associated expressway networks and the total road networks.
Key Words: automated truck lane; CO_2 emissions; integrated simulator; intercity expressway

Introduction

Intercity expressways are suffering from increasing congestion as freight transportation grows rapidly. The percent of trucks on these major transportation networks affect traffic flow characteristics, since trucks differ from passenger cars in terms of acceleration capabilities, maximum speed, length, and other characteristics. As the traffic flow becomes less homogeneous, efficiency and safety are likely to decrease, while CO_2 emissions increase, which will severely affect the environment. Given these circumstances and the likelihood of increased transport volumes, current

freight transport systems require major innovations. Automated truck lane (ATL) is a new concept in freight transportation systems that can prevent some of the expected effects of growing truck usage. But determining the extent to which ATL systems can help and the impact of such a system on the other traffic requires effectively estimating the operational performance of the ATL system.

The Tomei-Meishin expressway is a main transportation artery connecting the Tokyo, Chukyo, and Hanshin areas, Japan's three main metropolitan areas. Most of the freight movements among the three metropolitan areas take place via the Tomei-Meishin expressway. A second expressway, the New Tomei-Meishin Expressway, is expected to go into service in 2020. In order to evaluate the impact of the new ATL logistic system on traffic flow and the environment, the performance of the ATL system is simulated. Previous studies (Nishida 1998 and Ishizawa 2006) have focused on traffic flow changes by designating one lane of the New Tomei-Meishin Expressway as an ATL. Many studies have evaluated traffic flows for alternative strategies involving trucks. These studies (Garber and Gadiraju 1990; Peeta 2005) mostly focus on the basic analyses of the impacts on the expressway itself and assume a fixed demand. However, ATL systems in large traffic networks should be evaluated because the ATL likely will affect not only its own traffic flow but also those of other expressways and surface roads.

This paper introduces an integrated simulator framework for conducting simulation studies of ATL systems in large traffic networks. This paper also analyzes the fundamental environmental impact characteristics of an ATL system and compares them to those of alternative strategies, and then uses the integrated simulator to evaluate the environmental impact of ATL systems in large traffic networks.

Integrated Simulator

A new simulator is needed to simulate traffic flow on the large traffic networks that will be affected by the New Tomei-Meishin Expressway. Generally, the macroscopic or mesoscopic simulator is used, but evaluating the environmental impact of an ATL system on the New Tomei-Meishin Expressway and other expressways requires analyzing changes in the speed of passenger cars and trucks, since current speed and acceleration are the main factors in calculating the CO_2 emissions of a vehicle. The mesoscopic simulator can only provide the average speed of a vehicle in each link and can't fully evaluate the environmental impact of an ATL. Speed changes should be estimated using the microscopic simulation model, which can reflect the detailed relationship of the movements of a passenger car and a truck.

However, simulating large traffic networks with huge numbers of cars requires considerable computer time, so a microscopic simulator is not cost effective for simulating in detail the behaviors of entire traffic networks. Moreover, not all of the detailed road information needed for calibrating the parameters of models required for microscopic simulation of large traffic networks can be obtained. Therefore, ATL systems will be simulated using an integrated framework that combines the

microscopic and mesoscopic models. The proposed framework consists of a common module, meso and micro simulator which are shown in Figure 1.

Mesoscopic Simulator. A mesoscopic simulator will simulate the movement of vehicles on a surface road and provide the microscopic simulator with vehicles that enter the expressway from the surface road. The mesoscopic simulator also will call the common module to calculate the link travel time and provide dynamic route information for the total road network using the DUO (dynamic user optimum) assignment method. A vehicle group contains a maximum of five vehicles. The scanning interval is 15 seconds.

Microscopic Simulator. The microscopic simulator simulates the movement of vehicles on the expressways and provides real-time results of the link performance to the mesoscopic simulator. The expressways considered in this study consist of the road network in 2020, when New Tomei Meishin Expressway begins service. The inner two lanes of the six-lane expressway are designated automated truck lanes. The microscopic simulator updates each vehicle's position every 0.5 second, and simulates two types of vehicles—passenger cars and trucks. The system is capable of simulating expressway characteristics, including car-following, lane changes, on- and off-ramps, and merging and diverging sections.

A complicated model may fail because of the way the parameters are calibrated for large road networks. Therefore, the model should be general and simple. The objective of this study of integrated simulators is to evaluate the impact of ATL systems on large-scale road networks. In order to simulate the various behaviors of drivers before and after executing ATL strategies, the simulator must reflect the relationship between a car and a truck. Hence, the simulation logic of the model in this paper considers not only truck operation characteristics, but also car-truck interactions. To simulate the ATL system, the micro simulator mainly utilize a truck-following and modified lane-changing models to incorporate car-truck interactions that is based on the car-following and lane-changing logics in the FRESIM freeway traffic flow simulator.

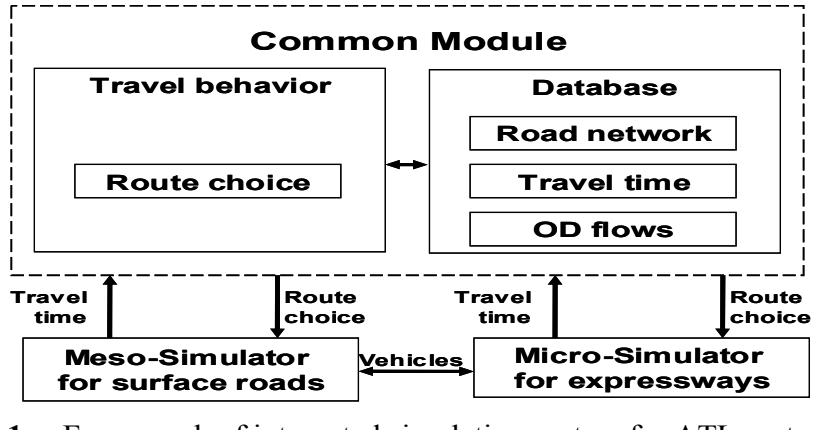


Figure 1. Framework of integrated simulation system for ATL systems.

Truck-following Model. The base model is the FRESIM car-following model. It is selected as a result of insights from a previous study (Aycin and Benekohal 1999), which compares several popular car-following models. That study concludes that the FRESIM car-following model replicates the field data more closely than other models when the driver sensitivity factors are finely calibrated. FRESIM uses a new set of values to represent the driver sensitivity factors, and the desired acceleration of the following vehicle is calculated according to the driver preferred time headway and relative speed. The model is expressed as follows.

$$H = L + qV_f + bq\,(V_l - V_f)^2$$

where H is the space headway (ft) between the leading and following vehicles, L is the lead vehicle length (ft), q is the driver sensitivity factor for the following vehicle (a time headway expression for drivers), V_f is the speed of the following vehicle at time t(fps), V_l is the speed of the leading vehicle at time t(fps), and b is a calibration constant defined as follows.

$$b = \begin{cases} 0.1, & V_l < V_f \\ 0, & otherwise \end{cases}$$

From the above formula, car-following acceleration for simulation scanning interval T is defined as follows.

$$a = \frac{2\left[x_{t+T} - y_t - L - V_f^t(q+T) - bq\,(V_l^{t+T} - V_f^t)\right]}{T^2 + 2qT}$$

Previous studies have indicated that car drivers dislike trucks and so the headway for following a car is lower than the headway for following a truck. This model is simply extended by including a term to represent the additional headway due to following a truck. The additional term is the coefficient for the car-truck headway. Based on car-following model, the truck-following headway can be expressed as follows.

$$H_{truck} = L + \beta\,(qV_f + bq\,(V_l - V_f)^2)$$

The truck-following acceleration can be expressed as follows.

$$a = \frac{2\left[x_{t+T} - y_t - L - V_f^t(\beta q + T) - b\beta q(V_l^{t+T} - V_f^t)\right]}{T^2 + 2\beta qT}$$

The value of coefficient β is estimated from the simulator results of a study (Yoo and Green 1999). In that study, the subjects followed cars about 10% closer than they followed trucks. Hence, β has a value of 1.1 in this paper.

Modified Lane-changing Model. The lane-changing model is another important model that is related to car-truck interactions. Drivers make discretionary lane changes to obtain a more favorable position and/or a higher speed. The desire to make a discretionary lane change is determined as follows.

$$d = \begin{cases} 100, & v \le v_i \\ 100\left[1 - \dfrac{v - v_i}{v_f - v_i}\right], & v_i < v < v_f \\ 0, & v \ge v_f \end{cases}$$

where v_i is an intolerable speed level below which the driver is highly motivated to change lanes. v_f is the desired free-flow speed and v is the current speed of the vehicle that is changing lanes. The motivation of the car driver also is affected by the car-truck interactions. A car-driver behavior survey indicates that these drivers are more willing to change lanes when they follow a truck. Therefore, the FRESIM lane-changing "desire" component is modified and modeled as follows.

$$d = \begin{cases} 100, & v \le v_i \\ \min\left\{100\left[1 - \dfrac{v - v_i}{v_f - v_i} + c\,\dfrac{(1.5 - q)}{1.5 - 0.6}\right], 100\right\}, & v_i < v < v_f \\ 100\,c\,\dfrac{(1.5 - q)}{1.5 - 0.6}, & v \ge v_f \end{cases}$$

where q is the driver sensitivity factor ranging from 1.5 to 0.6 (the smaller the value of q, the greater the motivation to bypass a truck) and c is the coefficient. This paper assume a value of 0.38 so that when q=1, the probability of the driver changing lanes is about 20%, even when the truck ahead travels at free-flow speed.

Data Exchange Procedure. Because microscopic and mesoscopic simulations differ in their basic characteristics, a specific procedure for data exchanges must be followed to ensure proper data exchanges between the two systems. The time step is shorter for microscopic simulator than for mesoscopic simulator. Moreover, the basic unit for microscopic simulations is one vehicle, while for mesoscopic simulation, it is one group of vehicles. In the data exchange procedure, the mesoscopic simulator reads the dynamic OD data to obtain vehicle data for the current time segment and then randomly assigns the departing vehicles to different simulation steps within the current time segment. One time segment is 30 minutes. Hence, one time segment contains 120 mesoscopic simulation steps and one mesoscopic simulation step is equivalent to 30 microscopic steps. The vehicle update process is shown in Figure 2.

CO_2 Emissions Models

The models used to determine CO_2 emissions are associated with the activities of the vehicles. In order to evaluate the environmental impact of ATL in greater detail, a microscopic emissions model for vehicles is used to reflect the contribution of the smoother transportation. In this paper, the vehicles are passenger cars and trucks. All of the passenger cars have gasoline engines, while all of the trucks have diesel engines. The CO_2 emissions model for trucks can be expressed as follows (Baba 2007).

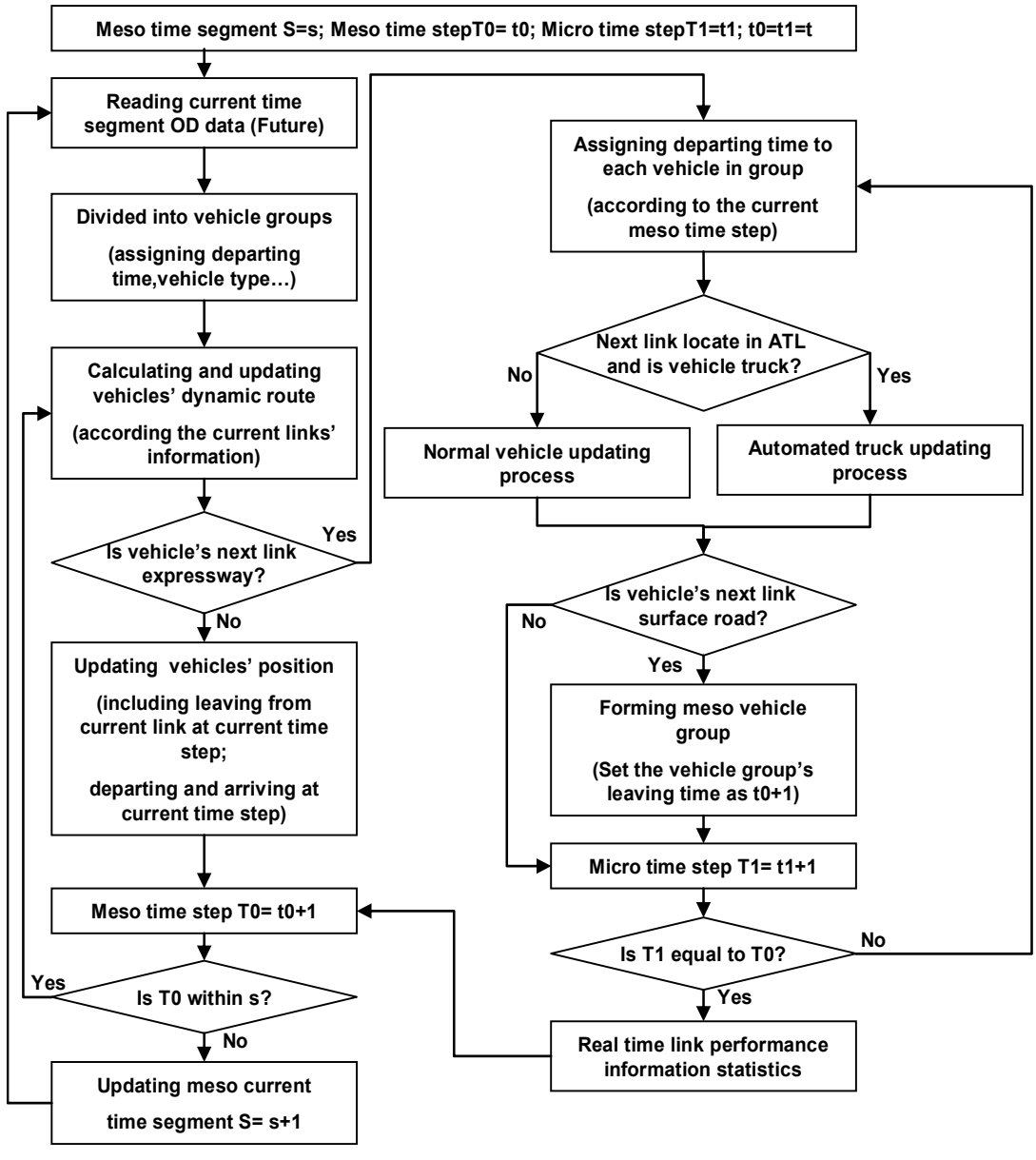


Figure 2. Flow chart of vehicle updating process for integrated simulator

$$E_{truck} = 3.50 \cdot v + 2.67 \cdot 10^{-4} \cdot v^3 + 3.64 \cdot a \cdot v + 113 \quad .9$$

where E_{truck} is the amount of CO_2 emissions (g/sec), v is the speed of the vehicle (km/h), and a is the acceleration of the vehicle (km/h/sec).

The CO_2 emissions model for cars can be expressed as follows (Oguchi 2006).

$$E_{car} = K_c \cdot (0.3 \cdot T + 0.028 \cdot D + 0.056 \sum_{k=1}^{K} \delta_k (v_k^2 - v_{k-1}^2))$$

where, E_{car} is the amount of CO_2 emissions (g), K_c is the coefficient of transformation (g- CO_2/cc-gasoline), T is the vehicle travel time (seconds), D is the vehicle travel distance (meters), δ_k is 1 when accelerating and 0 in all other cases, k is

the number of measured periods (k = 1,2,· ·K) and v_k is the vehicle speed for period k (m/sec.).

Basic Experiments

The experiment is conducted on a six km-long link with three lanes on the New Tomei-Meishin Expressway. Three cases are evaluated.

Mixed Lanes: This strategy allows trucks to travel in all lanes, which increases the number of car-truck interactions.

Dedicated Truck Lane: In this strategy, trucks are allowed to travel in the inner-most lane and cars are not allowed to travel in this dedicated truck lane, which reduces the number of car-truck interactions.

Automated Truck Lane: In this strategy, the inner-most lane is an automated truck lane for use by platoons of trucks.

Simulation Parameters. Traffic demand: Different traffic demands are used to test the cases studied. The traffic demand for a single lane varies from 500 vph to 2000 vph.

Percentage of trucks in traffic demand: The percentage of trucks can vary from 10% to 50% during peak periods. Hence, five percentages (10%, 20%, 30%, 40% and 50%) are utilized for the analysis.

Vehicle Parameters. Driver type: DSF (driver sensitivity factor) is a FRESIM parameter that reflects driver aggressiveness. The value of DSF ranges from 0.6 to 1.5. The car-following and lane-changing logics in the FRESIM freeway simulator are chosen as the base model based on insights from Aycin and Benekohal study (1999), while the extended truck-following model is used to reflect car-truck interactions.

Preferred speed: Each vehicle in the simulation is assigned a preferred speed under free-flow conditions. The preferred speed is a normal distribution ranging from 90 kmph to 120 kmph for passenger cars and from 90 kmph to 100 kmph for trucks. This represents the speed differences among vehicle types. The ATL expressway link has a speed limit of 100 kmph.

Case Comparisons. The performance data for the alternative scenario analyses is the CO_2 emissions ratio per vehicle kilometer and is obtained from five simulation runs. Figure 3 shows the aggregated results under a traffic demand of 5000 vph with trucks comprising 25% of the total traffic flow, which is close to the capacity of the expressway. Figure 3 shows that the ratio of CO_2 emissions for the ATL is the lowest among the three cases for all truck percentages above 10%. The simulation results under other traffic demands show a similar pattern.

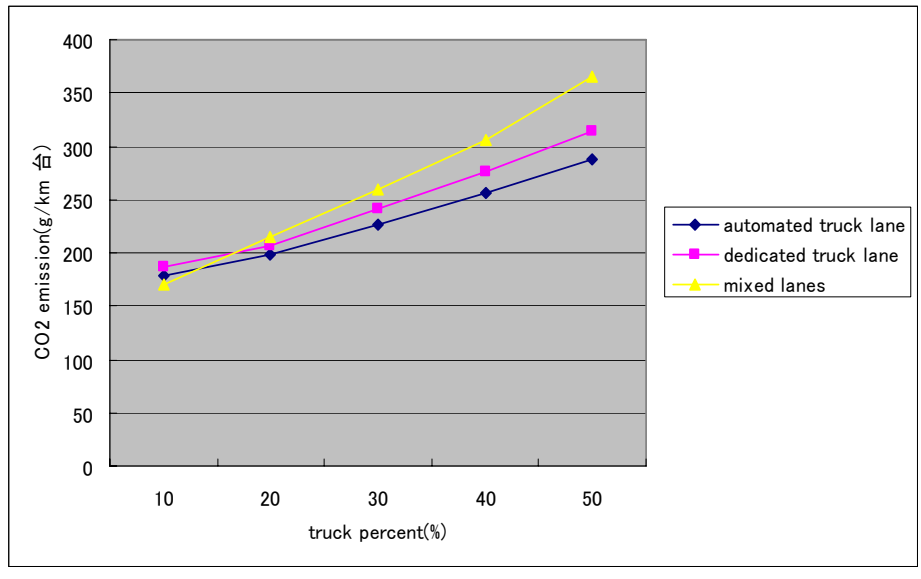


Figure 3. Overview of CO_2 emissions ratios for different scenarios and lane types under a traffic load of 5000 vph.

Simulation Experiment on an Intercity Expressway

In this experiment, the analysis focuses on the environmental impact of ATL systems on intercity expressways. The simulation compares three scenarios for the New Tomei-Meishin Expressway. The base case is the initial design case with four to six lanes serving two-way traffic, while case 1 has six lanes serving two-way traffic. Both the base case and case 1 allow trucks and cars in all lanes, and the maximum speed of the expressway is 100 kmph. Case 2 allows automated trucks to travel in the inner two lanes and normal trucks and cars to travel in the other four lanes. The main task is to compare the amounts of CO_2 emissions for the three cases and determine the impact on CO2 emissions of introducing ATL on intercity expressways.

Table 1 shows the simulation results for intercity expressways. Introducing ATL on the New Tomei-Meishin Expressway lowers the CO_2 emissions not only on the expressway itself, but also on the expressway networks and total road networks, as compared with the base case and case 1.

Table 1. Co_2 emissions for different cases for new tomei-meishin expressway.

	Total road network		Total expressway network		New Tomei-Meishin Expressway	
Base case (4-6 lanes 100 km/h)	90.3	-	33.7	-	11.7	-
Case 1 (6 lanes 100 km/h)	88.5	–1.8	32.5	–1.2	12.6	+0.9
Case 2 (4 lanes+2 automated truck lanes)	86.7	–3.6	31.1	–2.6	11.2	–0.5

Units: 1000 t- CO_2.

Table 2. CO_2 emissions ratios for different cases for New Tomei-Meishin Exp.

	Vehicle distance (km·vehicle)	Truck travel distance percent (%)	CO_2 emissions (kg)	CO_2 emissions ratio (g-CO_2/km·vehicle)
Base Case	29,748,025	0.7007	11,722,821	394.07
Case 1	35,689,631	0.7036	12,573,077	352.29
Case 2 (ATL)	35,806,337	0.7001	11,240,934	313.94

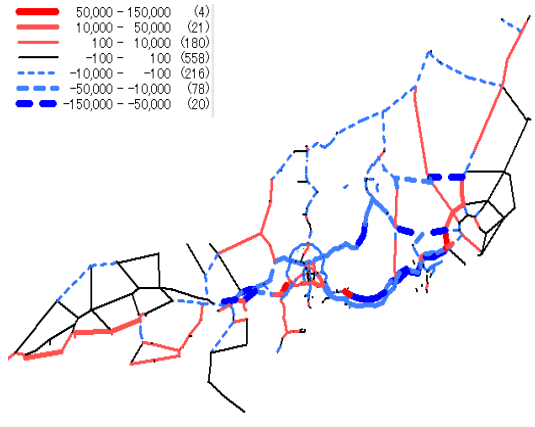


Figure 4. Difference in CO_2 emissions between case 2 and the base case

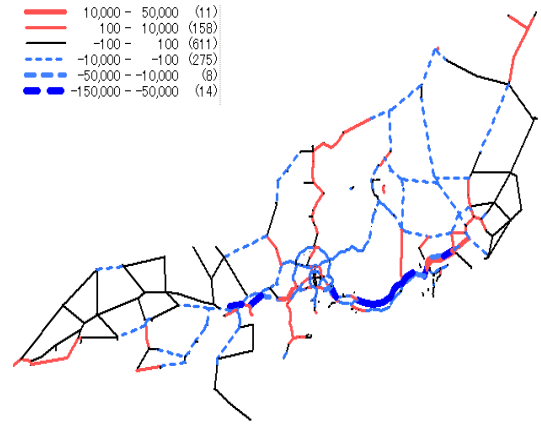


Figure 5. Difference in CO_2 emissions between case 2 and case 1

The CO_2 emissions ratios for the three cases for the New Tomei-Meishin Expressway are compared by calculating the CO_2 emissions for the per vehicle travel distances listed in Table 2. The CO_2 emissions ratio for the ATL case is 313.94 g-CO_2/km·vehicle, which is approximately 80% that of the base case, and 89% that of case 1. The lower environmental impact of the ATL is due to the steady travel of the automated truck platoons, which utilize fuel more efficiently than individual trucks.

The ATL system also changes CO_2 emissions across a large area. Figures 4 and 5 show that CO_2 emissions also clearly decrease in the alternate routes to the New Tomei-Meishin Expressway (i.e., Tomei and Chuou Expressways). The smoother travel on the New Tomei-Meishin Expressway improves the traffic flow from the Chuou and Tomei Expressways (dot dash lines). The result indicates that an ATL system reduces the environmental impact of the expressway on which it is installed as well other associated expressways. A more precise evaluation of the ATL system will require analysis of the system on large traffic networks.

Conclusions

An integrated simulation framework is used to study ATL system on an intercity expressway. Fundamental experiments comparing alternate strategies reveal the basic characteristics of the environmental impact of ATL systems. The results indicate that

the ATL system offers the best performance among the three cases for almost all traffic demands and truck percentages. The analysis on large traffic networks indicates that the ATL system has the lowest environmental impact among the three cases, with the lowest CO_2 emissions not only on the expressway itself but also on associated expressway networks and the total road networks. The ATL system also clearly decreases CO_2 emissions on the alternate routes to the New Tomei-Meishin Expressway (i.e., Tomei and Chuou Expressways). The results illustrate that we should analyze ATL system on large traffic networks, to make a more precise evaluation of it.

References

Nishida, M., *et al.* (1998). "An Automated Freight Transport System." *Transport Policy Studies' Review*, 1(2), 2-11.

Ishizaka, H. (2006). "An Advanced Freight Transport System." *Transport Policy Studies' Review,* 8(4), 12-18.

Garber, N. J., and R. Gadiraju. (1990). "Effects of Truck Strategies on Traffic Flow and Safety on Multilane Highways." In Transportation Research Record: *Journal of the Transportation Research Board*, No. 1256, TRB, National Research Council, Washington, D.C., 49-54.

Peeta, S., *et al.* (2005). "Behavior-based analysis of freeway car-truck interactions and related mitigation strategies." *Transportation Research Part B*, 39, 417-451.

Oguchi, T., *et al.* (2002). "Carbon Dioxide Emission Model In Actual Urban Road Vehicular Traffic Condition." *Journal of the Japan Society of Civil Engineers,* No. 695/Ⅳ-54, 125-136.

Baba, T., *et al.*(2007). "A Study on Vehicle Emission Factor in View of the Driving Conditions." *IBS Annual Report,* 85-87.

Aycin, M. F. and Benekohal, R. f. (1999). "Comparison of car-following models for simulation" *Transportation Research Record*, 1678, 116-127.

Yoo, H., and Green, P., (1999). "Driver behavior while following cars, trucks, and buses." *Project report prepared for University of Michigan Transportation Research Institute*, UMTRI-99-14, Ann Arbor, MI.

Determining Design Speed of Truck Escape Ramp Based on Back Analysis Method

Hongqi Liu[1], Tao Shen[2], Fang Wang[1], and Weihan Zhang[1]

[1] Road Safety Research Center and Research Institute of Highway, MOT, China, 100088 PH: +86 10 62355709, FAX: +86 10 62356515, Email: hq.liu@rioh.cn
[2] Yunnan Highway Development & Investment Co., Ltd, 650200. PH: +86 13987128106, Email: xmglst@126.com

Abstract

Research on design speed of TER (truck escape ramp) was carried out in order to make it more cost-effective and safer. The data of 9 TERs of Yuanmo Expressway has been collected. Back analysis method was brought forward and adopted to analyze the speed of out-of-control vehicles entering arrester bed. According to distribution of vehicle speed data, the speed of vehicles losing braking capabilities that met a probability were obtained, which could be used to determine TER design speed. The result also showed that 99 percent of vehicles' speed was lower than 102.2 km/h, which was much lower than the speed of 130 km/h introduced internationally. The TERs designed by using this result has opened to traffic for one year. No traffic accidents occurred yet caused by improper settings of TER.
Key Words: TER; design speed; traffic accidents; back analysis method

Introduction

The goal of this research was to investigate how to determine design speed of TER (emergency escape ramp) to make it more cost-effective and safer. Since the first TER was set up in the United States in 1956, it developed quickly and was adopted in worldwide (Witheford 1992). The first ramp in China was set up in Beijing Badaling Expressway in 1998. Although it appeared late in China, it developed fast because of the obvious role in reducing the accident rate caused by out-of-control vehicles, particularly trucks that have lost their brakes. More than a dozen provinces and municipalities have applied TER. However, with few studies conducted and a lack of experience on settings and application in China, vehicles frequently crashed on the ramp end or flew off the TER which created a very negative impact on this

driving group. Some drivers did not even dare to use TER. Therefore, improving safety of TER is very necessary.

According to TER design theory, design speed, which has a close relationship to arrester bed length and location, is the parameter most critical to TER security. If the numerical value of design speed is too low, it is easy for vehicles that lose breaking power to crash the ramp end or fly out of the TER and cause accidents. If the numerical value is too high, it will greatly increase the construction costs and add the difficulty of site selection, which may cause drivers to avoid the ramp due to the limitations of roadside.

While overseas scholars have carried out a lot of theoretical research and experimental research on material selection, and have developed technical specifications and guides, there is no clear consensus on TER design speed determining method. In the United States, the AASHTO Green Book (a policy on geometric design of highways and streets 2001) proposed 128.7 km/h as the minimum numerical value of design speed. Some states took energy cumulativeness iterative computation to determine the design speed of TER. There were fewer studies in China. Three ways were often used in design speed value selection, the first one was learned from the U.S. Green paper and suggested that the numerical value be 130 km/h (Cheng 2005); the second one was called the equivalent gradient component of forces (Bo 2007), assuming that the brake totally failed, vehicle gravity was equal to the sum of rolling resistance and air resistance; the third one was the conservation of energy, assuming that all the vehicles' potential energy converted into kinetic energy (Liu 2008). The three methods, or directly referred to foreign standard or calculated according to theoretical formulas, had no practicability in actual design. Local TER experience is still subjective determination, which is a serious threat to the safety of TER. Overall, it is very necessary to research on rational determining methods of TER design speed.

Back analysis value method is proposed in this research based on survey data of nine TERs on Yunnan Yuanmo Expressway. We collected the data of the arrester bed length and data of vehicle crashing the ramp-end, back analyzed the speed of a vehicle losing its breaks before it enters arrester bed by using arrester bed calculation theory and then determined losing braking vehicle speed which satisfied one probability required by adopting statistical analysis method. This method is simple and practical, and has been applied in Xiaomo Expressway in Yunna Province, no traffic accident casualties occurred since it applied more than one year.

Data Collection

By field observation and data collection of Yuanmo Expressway, we obtained alignment, TER design parameters, traffic volume, traffic composition, traffic accident records, and TER operation and maintenance records. Among the data that researchers studies were the lengths of a vehicle losing its breaking capabilities

before entering into the arrester bed. Accident numbers were the emphases of the data collection. The data we obtained provided a basis for further studies.

Survey Objection. Yuanjiang-Mohei Highway (Yuanmo Expressway) is located in Yunnan Province. The total length is 147.23 km. It has four lanes in two directions. There are many continuous long downgrades, a single continuous downgrade length is up to 24 km, and the largest longitudinal grade is 6%. Since it opened to traffic on December 28, 2003, accidents caused by losing braking trucks happened from time to time. To this end, nine TERs were installed on the whole line during 2004 and 2005. In 2005 the traffic flow increased 24%, but accident rate on long downgrade decreased 32%. The obvious effect shows it is a suitable research object although some casualties have taken place in arrester bed.

TER Data. Collected location and structure size of nine TERs on the Yuanmo Expressway was shown in Table 1.

From the field observation we knew that the TER arrester bed gradient and the length on the Yuanmo Expressway were basically determined by terrain of roadside. There was no uniform standard that was why some casualties took place on TER during the operation.

Accident Data. By communicating with management staff of the Yuanmo Expressway, we obtained data of losing braking vehicles entering into TER in recent two consecutive years. The data included distance of vehicle entering into arrester bed, vehicle damages, and casualties. Fifty-six pieces of data were valid excluding incomplete data records as well as the vehicles involved in accidents without asking for help and leaving TER by themselves. All the trucks entering into TER were overloaded and the majority had two axes (82%). Most of them could stop steadily in an arrester bed, but also a small number of them crashed the ramp-end (14%).

Table 1. Location and structure size of nine TERs on the Yuanmo Expressway.

Bollard No.	Approach Lane Length (m)	Percent Gradient (%)	Arrester Bed Length (m)	Arrester Bed Width (m)	Surfacing Material	Aggregate Depth (cm)
K225+600	49.17	6	56.1	8.5	Pea Gravel	60
K228+300	60	3	93	8.5	Pea Gravel	60
K232+000	30	15	62.27	8.5	Pea Gravel	60
K235+200	40	6	60	9.5	Pea Gravel	60
K281+900	110	3	50	7	Pea Gravel	60
K293+050	42.55	10	106.65	8.3	Pea Gravel	60
K297+020	38	10	73	8.5	Pea Gravel	60
K310+000	50	12+20	67.8	9	Pea Gravel	60
K317+000	25	8	60	7	Pea Gravel	60

Data Treatment

Traditional Method. It is very difficult to observe the speed of vehicles losing their braking capabilities. All the speed is predicted (so far the speed has not yet been observed at home). Several methods of calculating truck downgrade speed are available and are described briefly in the following:

Approximate method (Abdelwahab 1997). This method assumes that the loss of potential energy is completely converted into kinetic energy, that is:

$$V_x = \sqrt{V_i^2 - 2gh_x}$$

where

V_x = speed at distance x from the tope of grade, m/s;

V_i = initial speed (at top of grade), m/s;

g = acceleration of gravity, 9.81 m/s^2; and

h_x = difference in the elevation at distance x (negative) in meters.

While it is a gross simplification, it represents the upper bound for the speed at any distance along the downgrade.

Calculator program. Stanley (1978) developed the following equation to calculate the speed of a truck along any downgrade:

$$V_X = 8.82 \times \sqrt{0.01291 \times V_i^2 - 3.28 \times h_x - 3.28 \times KX - 3.3 \times 10^{-5} \times V_m X - \frac{0.0499 \times FXV_n^2}{W}}$$

where

V_X = speed at distance x from the top of grade, km/h;

K = constant incorporating surface friction and speed-independent portion of mechanical loss (equal to 0.01675 for pavement surface and 0.26125 for gravel bed);

V_m = average of V_i and V_X, km/h;

V_n^2 = mean value of V_i^2 and V_X^2, km^2/h^2;

F =frontal area of trucks, m^2;

W =truck GVW, kg; and

X = grade distance computed from stationing and decimal fraction grade.

where, V_m and V_n are related to V and to be solved by using iterative method.

Stanley (1978) verified his model by comparing observed and predicted downgrade speeds from field observations. The difference between the average observed and predicted speeds was within 5% of the observed speeds.

Simulation models. Several simulation models have been developed in the literature to predict the performance of heavy vehicles on grades. The main objective of these studies has been the development of speed-distance curves from which speed-reduction curves were derived to analyze the need for truck climbing lanes. Examples of these studies include one by the Society of Automotive Engineers (Truck 1987) and by Khan et al. (1990).

In method (1), all the potential energy is converted to vehicle kinetic energy, without considering air resistance, rolling resistance, etc. The forecast speed is obviously higher than the real speed of vehicles. While method (2) takes into account air resistance, it is very complicated. It needs to calculate every slope, iterative solution, and correct velocity in turn. Simulation methods call for the engineering familiar with the complicated models. So these methods are not suitable for application in practice.

Back Analysis Method. Because the above theoretical analysis methods do not have a practical value in engineering, and it is difficult to properly assess the forecast accuracy. After intensive research, we proposed an out-of-control speed back analysis method.

First of all, the speed of vehicles entering in an arrester bed was obtained by back analysis based on the braking length and TER parameter. Then, maximum entering speed meeting the requirements of probability can be obtained by statistics analysis. The analysis flowchart was shown in Figure 1.

According to kinematics formulas and lots of overseas practice, the following equation can be used to calculate the required length of TER arrester bed. (Witheford 1992)

$$L = \frac{V^2}{254(i + D_f)} \qquad (1)$$

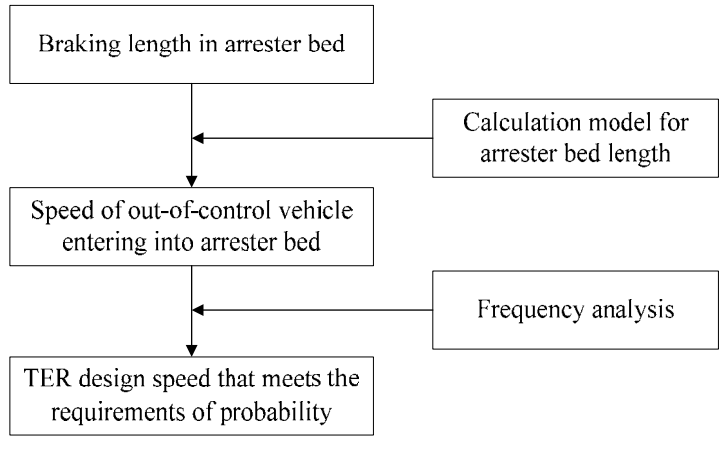


Figure 1. Analysis flowchart.

According to kinematics formulas and lots of overseas practice, the following equation can be used to calculate the required length of TER arrester bed. (Witheford 1992)

$$L = \frac{V^2}{254(i + D_f)} \tag{1}$$

where,

L = length of arrester bed, m;

V = entering velocity, km/h;

i = percent grade divided by 100; and

D_f = rolling resistance, expressed as equivalent percent gradient divided by 100.

After transforming Eq. 1 we can see:

$$V = \sqrt{254(i + D_f)L} \tag{2}$$

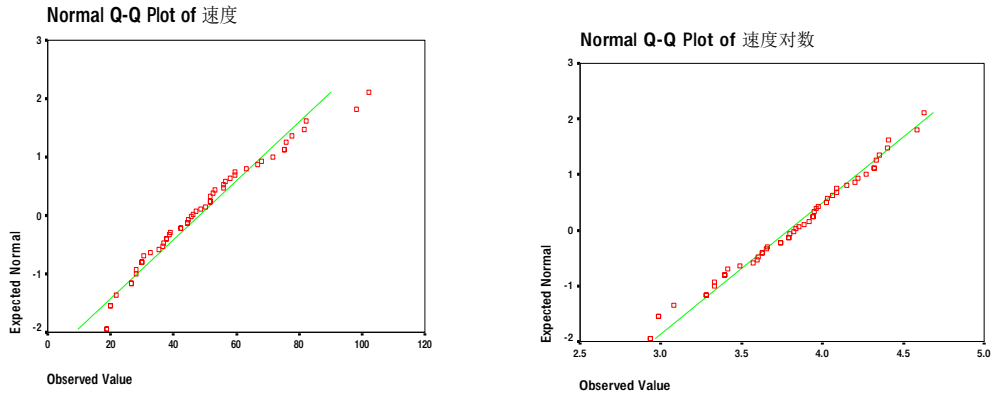
According to accident data, some of the vehicles crashed ramp-end. It indicated that vehicles still had kinetic energy at the end of an arrester ramp. It brought a certain degree of difficulty to predict the entering speed. According to overseas TER design experience, it is fatal to drivers when the speed of vehicle crashing ramp-end is higher than 40 km/h (Witheford 1992), so we make the following amendments to the original forecast data.

(1) If vehicles crashed the ramp-end causing death, the forecast speed increased to 40 km/h; (2) If vehicles crashed the ramp-end causing seriously injury, the forecast speed increased to 30 km / h; (3) If vehicles crashed the ramp-end causing minor injury, the forecast speed increased to 20 km / h; (4) If vehicles crashed the ramp-end causing no injury the forecast speed increased to 10 km/h.

Therefore, the speed of vehicles that lose their breaking capabilities could be determined based on TER design parameters, entering length, and casualties.

Result. Fifty-six pieces of speed data from vehicles that lost their breaking capabilities were obtained by back analysis. The sample data distributed approximate normal according to frequency analysis, the skewness was 0.676, Kolmogorov-Smirnov test passed the normal test (P value = 0.2), Shapiro-Wilk test failed passing the normal test (P value = 0.037), and the Q-Q analysis diagram showed that the data values had gradually deviated from the linear trend. So we took the logarithm of the sample data and then took a normal test, not only Kolmogorov-Smirnov (P value = 0.2) methods but also Shapiro-Wilk (P value = 0.385) passed the normal tests, the Q-Q analysis diagram showed that the data distributed around the straight line had no sign of deviation. So, speed data from vehicles that lost their breaking capabilities obeyed the log-normal distribution.

Speed lognormal regression analysis showed that the logarithm of speed of losing braking vehicles obeyed X ~ N (3.79,0.19), therefore we could calculate the speed of losing braking vehicles that met the probability of the following.



Before math.log After math.log

Figure 2. Q－Q chart.

Table 2. Correlation of P-value and velocity.

Probability	60%	70%	80%	90%	95%	99%
Velocity logarithmic (km/h)	3.94	4.02	4.18	4.34	4.43	4.63
Velocity (km/h)	51.64	55.78	65.28	76.53	84.34	102.22

Conclusion

According to data distribution, the speed of 102.2 km/h meets the 99% probability condition. Therefore, 105 km/h can meet the application requirement. The numerical value is much lower than 130 km/h that is usually mentioned internationally. Thus, the TER design speed should be determined by data from vehicles that lose their breaking capabilities but not adopt 130 km/h.

For highways on which TERs have been installed, it is very simple to decide design speed by adopting the back analysis method introduced by this research when there are new TERs or rebuilt old ones.

For new building highways that have no TER data, design speed can be obtained by adopting data of other highways that have the same channel because of the same traffic composition, vehicle load, and approximate driver level. If there is no data of adjacent highway on the same channel, the data from the same region also has a certain reference.
The research results have been applied on the Xiaomo Expressway, which belongs to the same international channel with the Yuanmo Expressway. No traffic accidents occurred yet caused by improper settings of TER.

Acknowledgment

This research was funded by Xiaomo Highway Construct Headquarters, Department of Communications of Yunnan Province.

References

"A policy on geometric design of highways and streets (Green Book)." (2001). *American Association of State Highway Transportation Officials (AASHTO)*, 259-268.

Abdelwahab, W., and Morral, J. (1997). "Determining need for and location of truck escape ramps." *Journal of Transportation Engineering*, 350-356.

Bo, Chen. (2007). "Study of rescue lane design method on the mountainous expressway downhill path." *Southwest Jiaotong University Maser Degree Thesis*, 19-32.

Huang, X., Tian, X. and Sun, Z. (2005). "Safety evaluation of truck escape ramp of Yuegan Expressway." *Highway*, 102-105.
pectives, 169-177.

Khan, A. M., Wong, J. Y., and Rastogi, M. (1990). "Heavy vehicle performance on grade and climbing lane criteria." Rep. No. *TDS-90-11,* Res. and Devel. Branch, Ontario Ministry of Transp. and Hwy., Downview, Canada.

Liu, X. (2008). "Determining technology of the location of escape ramp."*Journal of Highway and Transportation Research and Development*, 20-26.

Truck ability prediction procedure, SAE recommended practice. (1987).
Soc. of Automotive Engrs., Warrendale, Pennsylvania.

Witheford, D. (1992). "Truck Escape Ramps. NCHRP Synthesis of Highway Practice 178. " *Transportation Research Board*, 6-10.

Study of the Effectiveness of Nighttime and Truck Speed Limits

Dazhi Sun[1]

[1]Southwest Jiaotong University, Chengdu, China, Email: sbz2000@swjtu.edu

Abstract

The repealing of the national minimum speed limits in the U.S. gave states the freedom to set their own speed limits on non-interstate highways. Texas adopted different speed limits for cars and trucks dependent on the time of day. The purpose of this study is to evaluate, if the drivers are following the different speed limits during day and night time on various types of highways. Speed data were collected on the following highways: Interstate Highway-I37, Multilane Divided Highway-US 77, Farm Road-FR 665 and FR 70. Various statistical methods were applied to analyze the collected speed data. It was found that the most of car drivers on highways and truck drivers on farm roads do not follow the posted speed limits. The 85th percentile of traveling speed on the studied highways and farm road is 70mph ±2mph with irrespective of type of vehicle and time of the day.

Introduction

History of Speed Limits To combat the national energy crisis Congress passed the Emergency Highway Conservation Act in 1973 limiting Interstate highways maximum speeds to 55mph (Nicole Novotny, 2005). Also surface transportation and uniform relocation assistance act (STURAA) permitted the state agencies to increase the maximum speed limit on rural interstates from 55mph to 65mph in 1987. In 1995, states transportation agencies got permission to set their own speed limits for all non interstate highways due to the repeal of national maximum speed limit (NMSL) of 55mph.

Based on these laws, all the states of USA adopted the maximum speed limit of 55 mph for Interstates in 1974, as speed limit laws have been the responsibility of the individual states since 1901.Most of states agencies increased the speed uniformly for both cars and trucks from 55 to 65mph for rural interstates, however few states increased only the speed limits for cars leaving the truck speed as 55mph. And also

most of the state transportation agencies set the speed limits according to the type of the road and followed uniform speed limit (USL) and few states adopted differential speed limits and posted the speed limits according to the type of the road, vehicle class and time of the day since 1995. 32 states have raised speed limits to 70 mph or higher on some portion of their roadway systems as on March 2009.Table 1 gives the information about the states which follow the specific truck speed limit and day/night time speed limit.

It can be learnt from the above table that some states reduced posted truck speeds limits to 55 ~ 65mph, which is 5 ~ 15 mph less than the posted car speed limit. States like Idaho, Illinois, Montana, Oregon, and Washington posted 10mph less than car speed limit, Arkansas and Indiana posted 5mph less than cars speed limit, California posted 15mph less than car speed limit, Michigan posted 10-15mph and Texas posted 5-15mph less than car speed limits.

Table 1. States With Specific Truck Speed Limit and Nighttime Speed Limit (6).

State	Rural Interstates	Urban Interstates	Other Limited Access Roads	Other roads	Effective Date of Limits	
	Passenger Vehicles (mph)	Passenger Vehicles (mph)	Passenger Vehicles (mph)	Passenger Vehicles (mph)	Rural Interstates	Urban Interstates
Arkansas	70 Trucks:65	55	60	55	12/08/95	No action
California	70; Trucks:55	65 Trucks:55	70	65	01/08/96	No action
Idaho	75; trucks: 65	75	65	65	05/01/96	05/01/96
Illinois	65; trucks: 55	55	65	55	01/25/96	no action
Indiana	70; trucks: 65	55	60	55	07/01/05	no action
Michigan	70 (trucks 60); <70 (trucks55)	65	70	55	08/01/96	08/01/96
Montana	75 Trucks:65	65	Day:70 Night:65	Day:70 Night:65	05/28/99	05/28/99
Oregon	65 Trucks:55	55	55	55	06/07/87	No action
Texas	cars: 75 day; 65 night trucks: 70 day; 65 night	day: 70; night: 65	day: 75; night: 65	day: 60; night: 55	no action	09/01/99
Washington	70; trucks: 60	60	60	60	03/15/96	03/15/96

Truck Speed Limit. Whatever the type of speed limit, the main objective is to increase the road safety. Even though there are many factors which affect the road safety, speed limit is one of the major factors. Coming to truck speed limit it is implemented in 10 out of 50 states with respect to type of road. Truck speed limits are introduced by considering its size and heavy mass. Previous research found that "if a large truck is involved in a crash there is 20% to 40% increase in fatality rate" (Aerial Photograph and Google Earth, 2009). The statistical data from NHTSA shows that "trucks involved in 6.9% motor vehicle crashes, 4.9% motor vehicle injuries, 12% motor vehicle fatalities, and nearly half of the fatal crashes occurred at night times. Texas has the highest truck involved fatal crashes of 9.9%" (Chester G. Wilmot and Mandar Khanal, 1999).

Night Time Speed Limit. Texas is the only state in USA which posts the day and night time speed limits for both passenger cars and trucks with irrespective of type of the road. This type of speed limit is adopted by considering visibility and safety. During night time the visibility will be poor and it is hard for drivers to understand the road alignment and signs when compared to day time conditions, especially in winter season the visibility range fall severely due to the fog. "Another reason is the occurrence of accidents is high due to the driver fatigue, animals across the road (particularly on rural roads), and drunken drivers" (D. Alfred Owens and Michael Sivak, 1993).

*"*Research work done by ezio c. cerrelli found that there is 91% greater risk of fatal crash during night time than day time" (Aerial Photograph and Google Earth, 2009). The severity of accidents increases during night time including fatal pedestrians and bicycle crashes (Aerial Photograph and Google Earth, 2009, Chester G. Wilmot and Mandar Khanal, 1999, D. Alfred Owens and Michael Sivak,1993 ,Ezio C. Cerrelli, 1997). To avoid accidents and to increase the safety, posted speed limits are decreased during night time. The slower speed limits gives the driver sufficient time to react and take decision which helps in avoiding accidents. DOT'S are providing the information and warnings to drivers like lane markings, posting suggested speed limits, speed humps, rumble strips at the edge of the road and other to avoid accidents. "the research work done by USDOT-1996 found that there is a reduction in speed of about 3mph to 6mph and accidents by 36% by providing edge strips, widening visual markings and spacing of road side markings" (Maximum Posted Speed Limit, 2009). Various studies determined that night time driving conditions increase the severity of accidents including fatal pedestrians and bicycle crashes (Aerial Photograph and Google Earth, 2009, Chester G. Wilmot and Mandar Khanal, 1999, D. Alfred Owens and Michael Sivak,1993 ,Ezio C. Cerrelli, 1997).

Research Methodology

The research was conducted to study the effectiveness of the posted speed limits and influence of the type of the road on the speed of the vehicles using statistical analysis methods. The work is dividing into two major parts, field data collection and statistical analysis. Field data collection is further divided into site selection and data

collection. Coming to the data collection, speed of the vehicles is collected on 2 types of roads (highway and farm roads) during day & night time and detailed information is shown in table 2. The reason behind considering the road type is there will be difference in posted speed limit and geometry of the road section which affects the speed of the vehicle. The types of roads we selected are interstate highway, multilane divided highway, farm road with shoulder and without shoulder. The posted speed limits on interstate and multilane divided highway are same for trucks and passenger cars during day (70mph) and night time (65mph). And on farm roads there is different speed limit for trucks and passenger cars during day (T-60mph, C-70mph) and night time (T-55mph, C-65mph).

Field Data Collection

Site Selection. Speed measurements of the vehicles were taken on the 4 roads as said above, which are on the south of Texas. Sites are selected based on the criteria, that straight section of the road with adequate sight distance and suitable location for speed measurement.

Following are the sites selected for the study.

I37 N- Interstate Highway The first site is located on I37 which is one of the major interstate highways in Texas. It has 3 lanes with regular shoulder in each direction. The data was collected near the city of Corpus Christi where posted speed limit for cars and trucks is 70mph during day time and 65mph during night time.

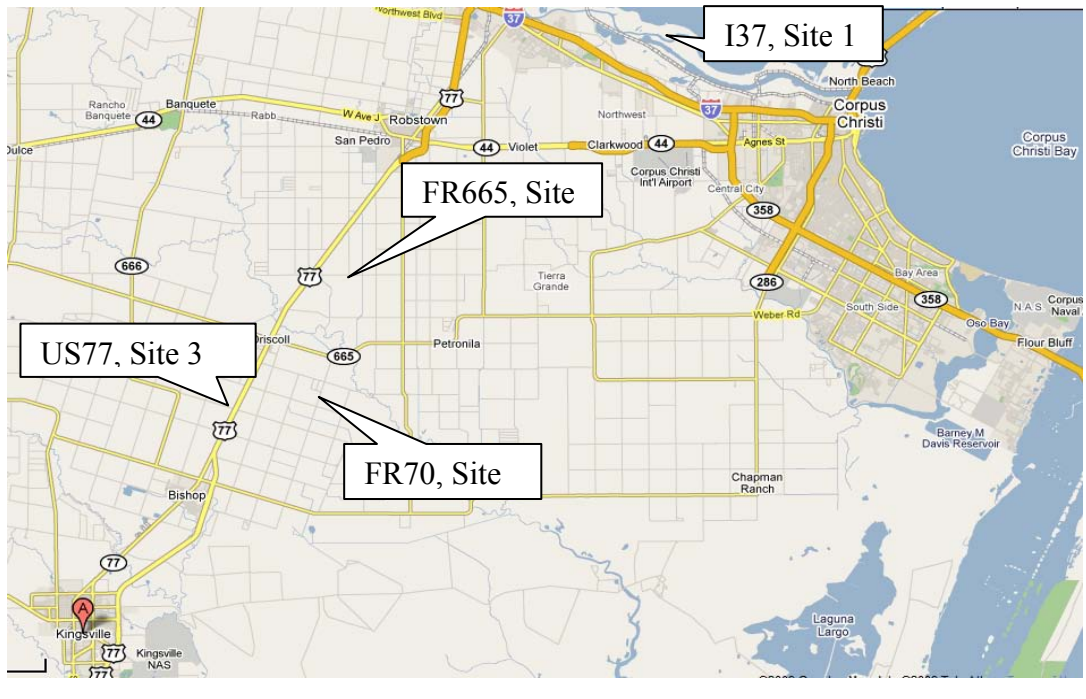


Figure 1. Location of the data Collection Sites.

FR 665- 2 Lane 2 Way Highway (With Shoulder) The second site is located on FR665 which is an undivided two-lane two-way highway with 6~8ft shoulder of each direction. The data was collected near the city of Driscoll where posted speed limit is different for cars and trucks. For cars it is 70 mph during day time and 65 mph during night time and for trucks 60mph during day time and 55mph during night time.

US 77- Multilane Divided Highway The third site is located on US77 which is a multilane divided highway. It has 2 lanes with regular shoulder in each direction. The data was collected near the city of Driscoll where posted speed limits for cars and trucks are 70 mph during day time and 65mph during night time.

FR 70- 2 Lane 2 Way Highway (Without Shoulder) The fourth site is located on FR70 which is an undivided two-lane two-way highway without shoulder. The data was collected on the way to Chapman ranch where speed limits are same as FR665.

Data Collection

Data was collected at a suitable location on one section of the road, where it is easy to collect the data and at the same time not disturbing the traffic. The speed of the vehicles was measured using Bushnell velocity Radar Gun.

500 readings were taken on I37, US77 and FR70 during day time except on FR665 (234) readings were taken due to less volume. Night time readings are taken on roads US 77& FR70. More detailed information is shown in Table 2.

Data Analysis

During the data analysis process, the mean and 85th percentile of traveling speed were identified for various situations and T-test was applied to compare the posted

Table 2. Information about Data Collection.

Road	Posted speed limit			No. of observations		No. of cars	No. of trucks
	Cars "MPH"		Trucks "MPH"	Day time	Night time		
I 37	70(day)	65(night)	70(day) 65(night)	500	----	384	116
US 77	70(day)	65(night)	70(day) 65 (night)	500	362	342(day) 294(night)	158(day) 68(night)
FR 665	70(day)	65(night)	60 (day) 55(night)	234	----	166	68
FR 70	70(day)	65(night)	60(day) 55(night)	500	200	462(day) 189(night)	38(day) 11(night)

speed limit with observed traveling speed. 85th percentile speed is the speed at which 85% of the drivers will be driving at or below. It is commonly used to set the speed limit of the roads. T-test method is used to study the effectiveness of speed limits. In this research, independent T-Test method is used to analyze two independent speed data sets. Whereas paired T-Test is used to analyze two interdependent data sets, e.g. the speed limit and observed traveling speed on the same road.

In this research T-Test is used to compare the observed vehicle speeds with posted speed limits, Trucks and cars speed during day and night time, day time with night time speed limits. This t-test analysis can be done using t-test formula (i.e. hand calculation) or computer software packages.

To analyze the effect of specific nighttime speed limit and truck speed limit, a series of T-test were done to compare the difference between the observed traveling speed and posted speed limit, and between truck speed and car speed.

In this study, we suppose the speed of the vehicle is V(i, j, s). While $i \in T$ (c, t), T(c, t) is the type of the vehicles, which including the cars and trucks. Besides, $j \in TI$ (d, n). TI(d, n) is when the data collected, including day time and nighttime. And, $s \in$ Site (1, 2, 3, 4). Site (1, 2, 3, 4) is the site of the data, including the four sites and the I-37 is site 1, FR665 is site 2, US77 is site 3 and FR70 is site 4. For each site, we compared the V(c, d, s) and V(t, d, s), and in two sites we also compare the V(c, n, s) and V(t, n, s). In this way we try to figure out that whether the cars and trucks travel at the same speed in different time and on different road.

Our hypothesis is $\mathbf{V(c, d, s) \neq V(t, d, s)}$, as well as $\mathbf{V(c, n, s) \neq V(t, n, s)}$.

In this study, we also tried to figure out if the vehicles drive at different speeds between the day and night time. So, for two sites, we compare the V(c, d, s) and V(c, n, s), also compare the V(t, d, s) and V(t, n, s). In this way, we can find whether the nighttime speed limit will affect the real driving speed.
Our hypothesis is $\mathbf{V(c, d, s) \neq V(c, n, s)}$, as well as $\mathbf{V(t, d, s) \neq V(t, n, s)}$.

Analysis Results
I-37 (Site 1)
The speed limit for I-37 is 70 mph for both cars and trucks during the day time. We got total 500 data during the day time, which includes 384 cars and 116 trucks.

From the data of table 3, we can see that cars and trucks have very similar driving speed and speeding percentage of the I-37 during the day time. Also the results of T-test show there are not significant difference between the actual traveling speeds of cars and trucks.

In I-37 we found that cars and trucks have almost the same speed during the day time. Also they have similar speeding percentage and 85% speed, which may also caused by the same speed limit of cars and trucks.

Table 3. Day Time Speeding Percentage of I-37.

	Total number (Daytime)	85% speed (mph)	Day time speed limit (mph)	Average Speed (mph)	Speeding percentage	
Car	384	71	70	66.1	71mph – 75mph	15.4%
					76+ mph	1.6%
Truck	116	71	70	66.4	71mph – 75mph	17.2%
					76+ mph	1.7%
Relative difference		0%	0%	0.5%		

Table 4. T-test Result Between the Day Time Data of Cars and Trucks on I-37.

hypothesis	V(c, d, 1) ≠ V(t, d, 1)
df	238
t Stat	1.8749521
P(T<=t) Single	0.03101197
t Single	1.65128116
P(T<=t) Double	0.06202394
t double	1.96998148

FR 665 (Site 2)

The speed limit for FR665 is 70 mph for cars and 65 for trucks during the day time. We got total 234 data during the day time, which includes 166 cars and 68 trucks.

Table 5. Day Time Speeding Percentage of FR 665.

	Total number (Daytime)	85% speed (mph)	Day time speed limit (mph)	Average Speed (mph)	Speeding percentage	
Car	166	71	70	68.6	71mph – 75mph	16.3%
					76+ mph	2.6%
Truck	68	70	60	66.9	61mph – 65mph	42.6%
					66+ mph	25%
Relative difference		1.4%	16.7	2.5%		

Table 6. T-test result between the day time data of cars and trucks on FR 665.

hypothesis	V(c, d, 2) ≠ V(t, d, 2)
df	137
t Stat	3.2610498
P(T<=t) Single	0.0007004
t Single	1.6560521
P(T<=t) Double	0.0014007
t double	1.9774312

The data of table 5 shows that cars and trucks have different driving speed and speeding percentage of the FR665 during the day time, and the relative difference is about 1.4% and 2.5% for the 85th speed and average speed, but as high as 16.7% for the posted speed limits. From the results of T-test, we can see that the difference between the actual traveling speeds of cars and trucks is statistically significant.

In addition, we can see that the speeding percentage of the trucks is much higher than cars, so in some ways, the truck speed limit is not working as much as people expect.

US 77(Site 3)

Table 7. Speeding percentage of US 77.

		Total number (Daytime)	85% speed (mph)	Day time speed limit (mph)	Average Speed (mph)	Speeding percentage	
Day Time	Car	342	72	70	68.6	71mph – 75mph	29.5%
						76+ mph	2.3%
	Truck	158	71	70	66.9	71mph – 75mph	16.5%
						76+ mph	1.9%
	Relative difference		1.4%	0%	2.5%		
Night Time	Car	294	70	65	66.0	66mph – 70mph	41.1%
						71mph – 75mph	10.5%
						76+ mph	2.4%
	Truck	68	68	65	64.2	66mph – 70mph	38.2%
						71mph – 75mph	5.9%
						76+ mph	0%
	Relative difference		2.9%	0%	2.8%		

The speed limit for US77 is 70 mph during the day time and 65 mph during the nighttime for both cars and trucks. We got total 500 total data during the day time, which includes 342 cars and 158 trucks; as well as 362 total data during the nighttime, which includes 294 cars and 68 trucks.

From the data of table 7, we can see that there is a little difference between the average speed of trucks and cars during both day and night time. This difference also shows in the speeding percentage and 85% speed, where trucks have fewer speeding percentage and lower 85% speed. The relative differences for day and night time are very small as well. However, the results of T-test show that the differences of the actual traveling speeds between cars and trucks are significant for both day and night time. Another T-test was conducted to compare the actual speeds between day and nighttime. And the results show that the difference is significant for both cars and trucks.

In all, truck speed of US77 is slower than car speed for both day and night time. Trucks also have lower speeding percentage. But the difference is only 1-2 mph.

Table 8. T-test result of US 77.

	Day time		Nighttime	
Comparison between cars and trucks	hypothesis	V(c, d, 3) ≠ V(t, d, 3)	hypothesis	V(c, n, 3) ≠ V(t, n, 3)
	df	316	df	97
	t Stat	5.1158897	t Stat	2.9390336
	P(T<=t) Single	2.714E-07	P(T<=t) Single	0.0020574
	t Single	1.6496899	t Single	1.6607146
	P(T<=t) Double	5.428E-07	P(T<=t) Double	0.0041147
	t double	1.9674995	t double	1.9847231
Comparison between day time and nighttime	Cars		Trucks	
	hypothesis	V(c, d, 3) ≠ V(c, n, 3)	hypothesis	V(t, d, 3) ≠ V(t, n, 3)
	df	624	df	116
	t Stat	7.507099	t Stat	3.599421
	P(T<=t) Single	1.0482E-13	P(T<=t) Single	0.00023517
	t Single	1.64729922	t Single	1.65809574
	P(T<=t) Double	2.0965E-13	P(T<=t) Double	0.00047035
	t double	1.96377287	t double	1.98062594

FR 70(Site 4)

The speed limit for FR70 is 70 mph for cars and 65 for trucks during the day time; and 65 mph for both cars and trucks during the nighttime. We got total 500 data during the day time, which includes 462 cars and 38 trucks; as well as 200 total data during the nighttime, which includes 189 cars and 11 trucks.

For FR70, it was found that the relative difference of the speed limits for truck and car is about 16.7%, the difference of the observed 85th percentile speed is about 2.9% and the difference of the mean speed is about 9.7%. And during night time, although the relative difference of the speed limits is as high as 18.2%, the difference of 85th speed and the average speed between cars and trucks is very small.
The data shows that cars and trucks have different driving speed and speeding percentage during the day time. But they have very similar driving speed and speeding percentage during the nighttime. When using T-test to compare the actual speed between cars and trucks, the difference is significant for the day time but not significant for the nighttime. Another T-test was also conducted to compare the actual speed between day and night time. And the results show that the difference is

Table 9. Speeding percentage of fr 70.

		Total number (Daytime)	85% speed (mph)	Day time speed limit (mph)	Average Speed (mph)	Speeding percentage	
Day Time	Car	462	72	70	65.4	71mph – 75mph	17.3%
						76+ mph	3.5%
	Truck	38	70	60	59.6	61mph – 65mph	21.1%
						66+ mph	23.7%
	Relative difference		2.9%	16.7%	9.7%		
Night Time	Car	189	68	65	61.8	66mph – 70mph	25.9%
						71mph – 75mph	7.9%
						76+ mph	1.1%
	Truck	11	68	55	61.4	56mph – 60mph	18.2%
						61mph – 65mph	27.3%
						66+ mph	36.4%
	Relative difference		0%	18.2%	0.7%		

significant for cars, but not significant for trucks. From these results, we found the day time speed between cars and trucks are significantly different. But during night, their speeds are almost the same and the difference is not significant. They also have the samiliar 85% speed during night time. When comparing the data between day and night time, the data of cars shows significant difference. But the data of trucks

Table 10. T-test result of FR70.

Comparison between cars and trucks	Day time hypothesis	V(c, d, 4) ≠ V(t, d, 4)	Nighttime hypothesis	V(c, n, 4) ≠ V(t, n, 4)
	df	41	Df	11
	t Stat	4.44514	t Stat	0.185848
	P(T<=t) Single	3.27E-05	P(T<=t) Single	0.427974
	t Single	1.682878	t Single	1.795885
	P(T<=t) Double	6.54E-05	P(T<=t) Double	0.855947
	t double	2.019541	t double	2.200985
Comparison between day time and nighttime	Cars hypothesis	V(c, d, 4) ≠ V(c, n, 4)	Trucks hypothesis	V(t, d, 4) ≠ V(t, n, 4)
	df	309	Df	17
	t Stat	6.020606918	t Stat	0.6533733
	P(T<=t) Single	2.46147E-09	P(T<=t) Single	0.2611277
	t Single	1.649799827	t Single	1.7396067
	P(T<=t) Double	4.92294E-09	P(T<=t) Double	0.5222553
	t double	1.96767082	t double	2.1098156

Table 11. Result of All T-tests for Significance.

Site	Hypothesis	Level of significance (α)	T value (two tails)	Reject the Null hypothesis?
1	V(c, d, 1) ≠ V(t, d, 1)	5%	1.96998148	Reject
2	V(c, d, 2) ≠ V(t, d, 2)	5%	1.9774312	Fail to Reject
3	V(c, d, 3) ≠ V(t, d, 3)	5%	1.9674995	Fail to Reject
	V(c, n, 3) ≠ V(t, n, 3)	5%	1.9847231	Fail to Reject
	V(c, d, 3) ≠ V(c, n, 3)	5%	1.96377287	Fail to Reject
	V(t, d, 3) ≠ V(t, n, 3)	5%	1.98062594	Fail to Reject
4	V(c, d, 4) ≠ V(t, d, 4)	5%	2.019541	Fail to Reject
	V(c, n, 4) ≠ V(t, n, 4)	5%	2.200985	Reject
	V(c, d, 4) ≠ V(c, n, 4)	5%	1.96767082	Fail to Reject
	V(t, d, 4) ≠ V(t, n, 4)	5%	2.1098156	Reject

shows no significant different between day and night time. This specific result may be caused by the specific driving condition of FR70, which is an undivided two-lane two-way highway with very narrow shoulder. So vehicles are more likely to stay in a constant speed.

Conclusions and Recommendations

On both highways I-37 & US-77, 85th percentile speed for cars and trucks during day and night time is nearly 70mph. Observations and a series of statistical analysis on the data collected at US-77 highway, indicate that during night times 54% of the cars and only 44 % of the trucks are not following the speed limit. On farm road FR665 & FR 70 the percentage of trucks moving over the speed limit is very high compared to cars during day and night time and the average speed of trucks is almost equal to the cars. Also the T-test analysis for the same data for these locations suggest that there is no significant difference between day and night time traveling speed. Another finding is that there is no significant difference between the average speeds of trucks during day and night time. Percentage of cars exceeding the posted speed limit is much higher during night time than day time and the 85th percentile speed for both cars and trucks is nearly equal to 70mph. Finally the results show that the specific truck and night time speed limit is not working as much as people expected in Texas and there is a need to revise the posted speed limits as most of the car drivers on highways and truck drivers on farm roads are not following them and 85th percentile speed for highways and farm roads is 70mph ±2mph with irrespective of type of vehicle and time of the day.

In order to make a more comprehensive study in the future, more data will be collected on different types of highways, because only two Interstate highways and two rural roads were considered in this study. In addition, some other factors would be included in future study, such as weather conditions, day of the week, time of the day etc which also affect the traffic speeds and were not considered in earlier studies.

Acknowledgement

Authors are very thankful to the graduate students of transportation engineering laboratory of Texas A & M University -Kingsville, for their wonderful job during field data collection.

References

*"*Aerial photograph, google earth*."*,(2009) http://www.googleearth.com/, Accessed

Chester G. Wilmot, Mandar Khanal,(1999). "Effects of speed limits on speed and safety: a review." *Transportation Reviews,* 19(4), 315-329

D. Alfred Owens, Michael Sivak,(1993). "The role of reduced visibility in night time road fatalities." *Transportation Research Institute*, No.UMTRI-93-33.

Ezio C. Cerrelli,(1997). "Fatal crash involvement-what are the odds?" *NHTSA, US Department of transportation, Research Report*.

Maximum Posted Speed Limit,(2009) Insurance institute for highway safety, http://www.iihs.org/laws/speedlimits.aspx.

Texas safety facts (2007) data, publication DOT HS 810 993. NHTSA's Nation Center for Statistics and Analysis.

US Department of Transportation (USDOT), (1996). FHWA study tour for Speed Management and Enforcement Technology.

Study on Unsignalized Access Points Using Microscopic Traffic Simulation

Guoqiang Zhang and Jian Lu

Department of Traffic Engineering, Transportation College, Southeast University, 2 Si-Pai Lou, Nanjing City, Jiangsu Province, 210096, China; PH:(86)25-83790776; FAX(86)25-83795352; Email: flyingfeather_zhang@yahoo.com.cn, jianjohnlu@sina.com.cn

Abstract

Faced with severe problems of traffic congestion and safety, the United States of America has carried out research and application of access management techniques, aiming to improve urban road transportation, which has produced very good social and economic benefits and can be adopted by other countries. In China, urban arterial roads often have too many access points, which is one of the main causes of traffic congestions and traffic accidents. The paper has primarily studied unsignalized access points on arterial roads with two lanes in each direction, using a microscopic traffic simulation. By analyzing average delay time of the minor road approach in many situations, standards concerning traffic volumes of the intersection approaches have been given. When traffic volumes are beyond the standards, higher levels of access control should be considered. Results of the paper are very helpful for application of access management techniques.
Key Words: access management; unsignalized access points; microscopic traffic simulation; analysis of average delay time

Introduction

Access management is the systematic control of the location, spacing, design, and operation of driveways, median openings, interchanges, and street connections to a roadway (Committee on Access Management, 2003). It also involves roadway design applications, such as median treatments and auxiliary lanes, and the appropriate spacing of traffic signals. The purpose of access management is to provide vehicular access to land development in a manner that preserves the safety and efficiency of the transportation system (Eisele and Toycen, 2005). Access management programs seek to limit and consolidate access along major roadways,

while promoting a supporting street system and unified access and circulation systems for development. The result is a roadway that functions safely and efficiently for its useful life as well as a more attractive corridor.

Concepts and techniques concerning access management have already been accepted and adopted by Chinese traffic engineers (Ge et al., 2008). Some of them have been used in many metropolitan areas of China. However, many of the applications of an access management technique lack systematic analysis, and study and the overall benefits are not very obvious. From 2003 to 2007, under the leadership of Professor Lu of Southeast University, a reach project, sponsored by China Department of Transportation, was carried out to study traffic safety technique for highway Intersections. In the research, access management has been systematically studied, and some meaningful conclusions are drawn for practical application of access management in China (Lu et al., 2007). Results of the research project have set up the basis for future study on access management in China. This paper is the extension of the research project.

In China, one of the most serious problems that urban transportation systems are faced with is that arterial roads have too many access points (Lu et al., 2007). The paper will mainly discuss unsignalized access points on arterial roads with two lanes in each direction, which exist widely in urban transportation networks of China. Using microscopic traffic simulation, all kinds of traffic situations of access points are analyzed to find out meaningful suggestions for access control measures. The researchers mainly studied yield-controlled access points. Stop-controlled access points can also be studied in the same way.

Study Framework

Field survey of arterial roadways shows that, for most unsignalized access points, long waiting time on approaches of the minor roads almost always cause drivers to force an entry into the conflicting traffic stream of the major roads, which is very dangerous and hampers movements of the major road vehicles (Zhang and Chen, 2009). This type of action is the main cause of traffic accidents and traffic congestions and should be eliminated (Zhang et al., 2008). Therefore, access management measures should be taken when delay time of the minor roads is unbearably long. The paper analyzed average delay time on the minor road approach of the arterial road access point, in different traffic conditions, using microscopic traffic simulation. For a given traffic volume of the minor road, traffic volumes of the arterial road were increased gradually from lower levels to higher levels and during the process, average delay time on the minor approach would increase. When average delay time surpassed the allowable maximum value, the corresponding traffic volumes of the major road were regarded as the reference volumes that higher access control measures must be taken. Otherwise, serious traffic problems such as traffic crashes or traffic congestions would appear.

The Highway Capacity Manual 2000 specifies that for unsignalized intersections, more than 50 seconds average control delay corresponds to a Level of Service F. Average delay time is very close to the concept of average control delay and can be used as a substitute for average control delay. Compared with cities of the United States, metropolitan areas of China are more crowded, and people usually wait much longer to get through an intersection safely. Therefore, we used 60 seconds as the allowable maximum value for average delay time in our analysis. Access points in our analysis are three-way intersections, which have one minor road approach and two major road approaches. With fewer pedestrians and bicycles on the road, our research will concentrate on vehicles.. Motor vehicles that turn into the minor road from the major roads are also assumed to be very few and have no influence upon the operation of the access point. Therefore, it is the traffic volumes of the minor road approach and the major road approaches that decide the criteria for higher degree of access control.

Three scenarios, which represent three modes of access design and are widely used in China, were analyzed separately, and they are as follows: (1) left turns and right turns from the minor road approach are permitted, and only one lane is provided for both of left turns and right turns; (2) Left turns and right turns from the minor road approach are permitted and a left turn lane and a right turn lane are provided separately for left turns and right turns; (3) Only right turns from the minor road approach are permitted, and a right turn lane is provided for right turns.

According to discussions above, access design for one of the three scenarios was made. Traffic volume of the minor road approach was set to different levels. Then, for each volume level of the minor road approach, traffic volumes of the major road approaches were increased gradually, and microscopic traffic simulation was used to analyze average delay time of the minor road approach until the time when average delay time of the minor road approach exceeded 60 seconds. Then, the traffic volumes of the minor road approach and the major road approach in the last simulation were the maximum volumes below which the unsignalized intersection could operate reasonably. In this way, standards for more restrictive access management on arterial roads could be obtained. Thus, study framework for analysis of access points on arterial roads can be shown by Figure 1.

Simulation Modeling

To facilitate the study, microscopic traffic simulation is used to simulate all kinds of situations that exist in the real world. The use of microscopic traffic simulation models can provide a quick and cost-effective way to evaluate the average delay time of the minor road approach in all kinds of situations (Jha et al., 2004). Microscopic models feature the calculation and prediction of the state of individual vehicles in continuous or discrete time-space and offer detailed descriptions of both road and traffic characteristics, traffic control and management strategies and techniques and so on. Therefore, in this paper we adopt one of the microscopic traffic simulation models, TSIS (Traffic Software Integrated System), as our study tool.

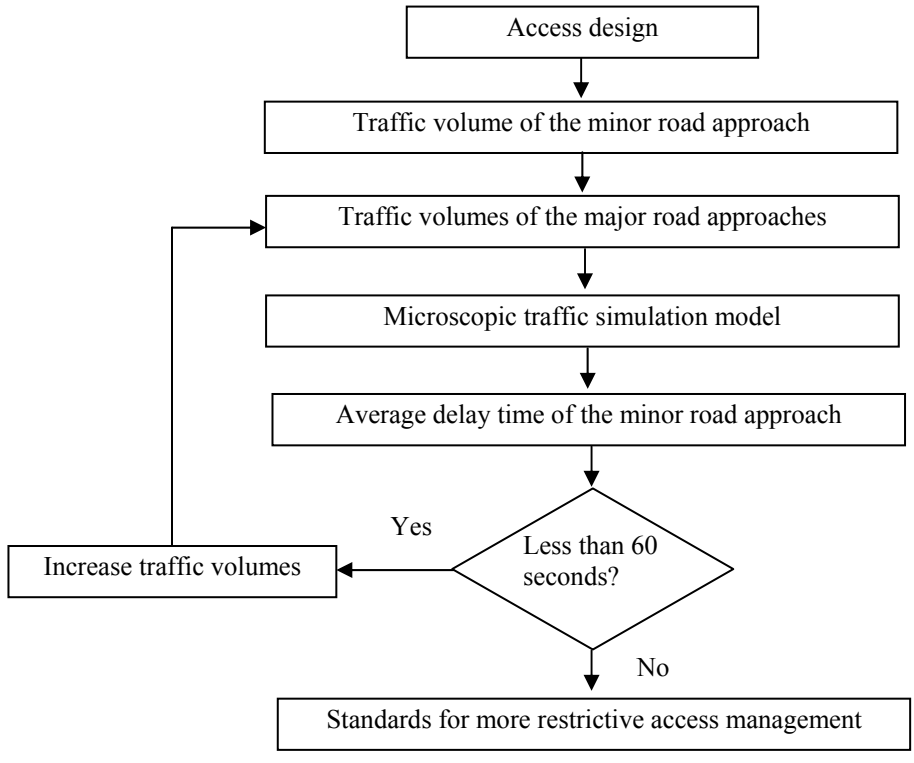


Figure 1. Study framework for analysis of access points on arterial roads.

TSIS is developed in the United States of America by ITT Industries, Inc., Systems Division under the direction of the Federal Highway Administration (FHWA). It is a toolbox that contains tools that allow the user to define and manage traffic analysis projects, define traffic networks and create inputs for traffic simulation analysis, execute traffic simulation models, and interpret the results of those models. The pre-configured tools include a GUI (graphical user interface) based network and simulation input editor (TRAFED), the CORSIM traffic simulation, and the TRAFVU animation and simulation analysis tool.

CORSIM is the pivotal component, which applies time step simulation to describe traffic operations. A time step is 1 second. Each vehicle is a distinct object that is moved every second. Each variable control device (such as traffic signals) and each event are updated every second. CORSIM is a stochastic model, which means that random numbers are assigned to driver and vehicle characteristics and to decision-making processes. CORSIM accumulates data every time step. At the end of each time period the accumulated data is used to produce MOEs (measures of effectiveness), which include vehicle speed, vehicle stops, delays, vehicle-hours of travel, vehicle-miles of travel, fuel consumption, and pollutant emissions. The physical environment is represented as a network comprised of nodes and unidirectional links. The links generally represent road sections or freeway sections, and the nodes generally represent intersections or points at which a geometric property changes (such as a change in grade or a major midblock traffic generator).

Using TSIS, we have built up microscopic traffic simulation models for the three scenarios of the yield-controlled access point on the arterial road with two lanes in each direction. Pedestrians and bicycles were not simulated in the models. The percentage of motor vehicles that turn into the minor road from the major road approaches are 3% when traffic volumes of the major road approaches are below 1000 vehicles per hour and are 1% when traffic volumes of the major road approaches are above 1000 vehicles per hour. The percentage of trucks is 10%. To simplify analysis, traffic volumes of the two approaches of the major road are given the same value. Besides, traffic volume of left turns from the minor road approach and traffic volume of right turns from the minor road approach are also given the same value for Scenario 1 and Scenario 2, where left turns and right turns from the minor road approaches are both permitted.

In order to calibrate the microscopic traffic simulation models, data have been collected from some cities of Jiangsu Province. Based upon careful analysis of the data, key parameters of the simulation models are set. Maximum deceleration in case of emergency is 15 feet per second squared and maximum deceleration in case of non-emergency is 8 feet per second squared. The mean value of start-up lost time is 2 seconds, and the mean queue discharge headway is 1.8 seconds. The length of cars is 14 feet, and the length of trucks is 57 meters. Vehicles from the minor road have to wait for adequate headways in order to perform right turns or left turns. For right-turn vehicles, the threshold is 4.3 seconds, and for left-turn vehicles, the threshold is 8 seconds. The data of average delay time are collected every minute by the software and since there are great variations in the beginnings of the simulation, we only used the data of the last 30 minutes. For each situation, 10 times of simulation have been run with different random seeds, and the values of average delay time from the runs are then averaged to get rid of the variations caused by different random seeds.

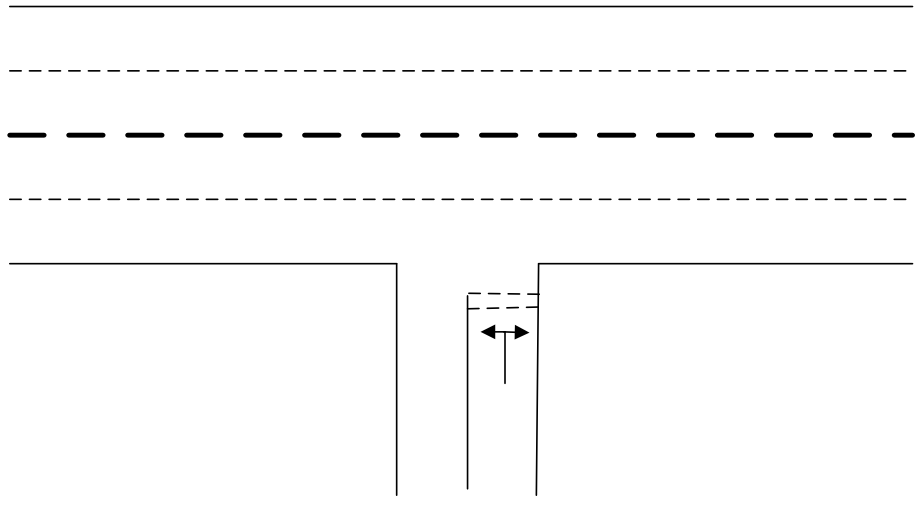


Figure 2. Intersection layout for Scenario 1.

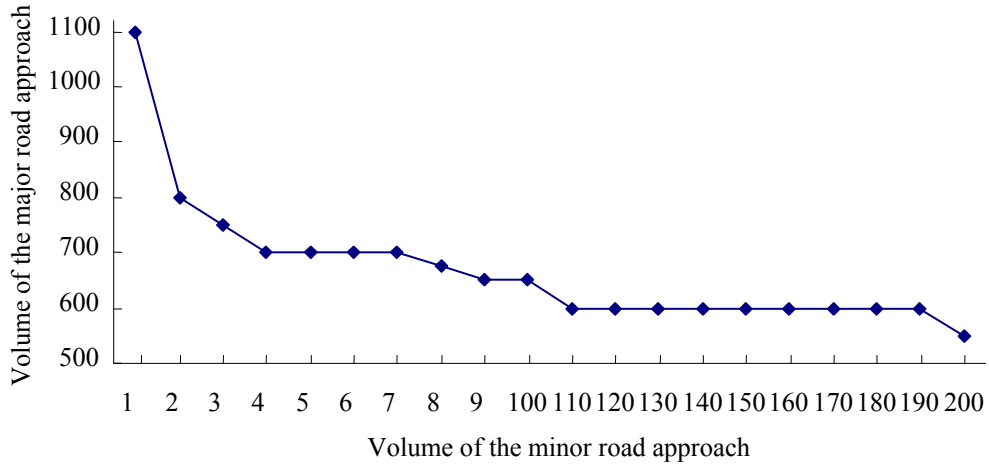


Figure 3. Access control standards of traffic volumes for Scenario 1.

Analysis of Three Scenarios

***Scenario* 1.** For Scenario 1, left turns and right turns from the minor road approach are both permitted. Only one lane is provided on the minor road approach for both the right-turn vehicles and the left-turn vehicles. Intersection layout for Scenario 1 is shown in Figure 2. Traffic volumes of the two major road approaches are equal. According to the study framework, the traffic volume of the minor road approach is given a series of values from 10 vehicles per hour to 200 vehicles per hour. For a traffic volume of the minor road approach, the maximum traffic volume of the major road approach is fine above which average delay time will become unbearably long. Therefore, traffic volumes obtained in this way are the standards for higher levels of access control. The maximum volume of the major road approach for each of the 20 volumes of the minor road approach is shown in Figure 3, which gives the standards for access control. Traffic volumes of access points should be below the curve. Otherwise, some measures should be taken on the minor road approach, such as channelization of left turns and right turns or prohibition of left turns.

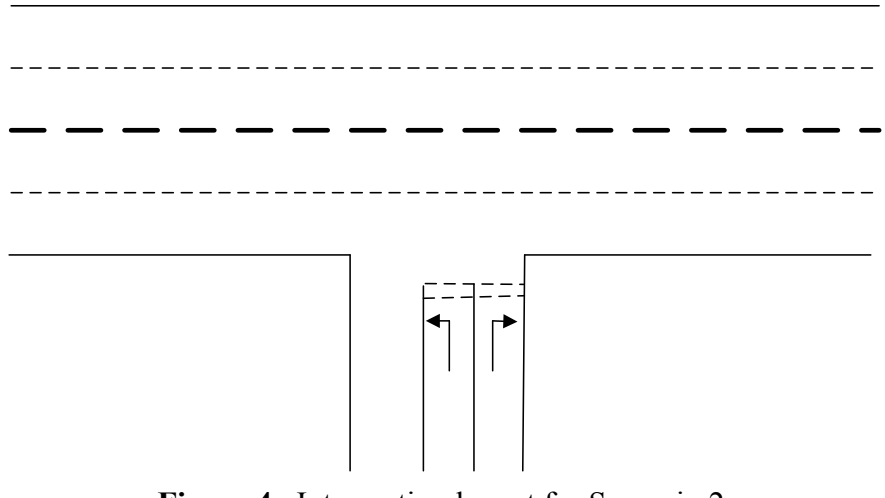


Figure 4. Intersection layout for Scenario 2.

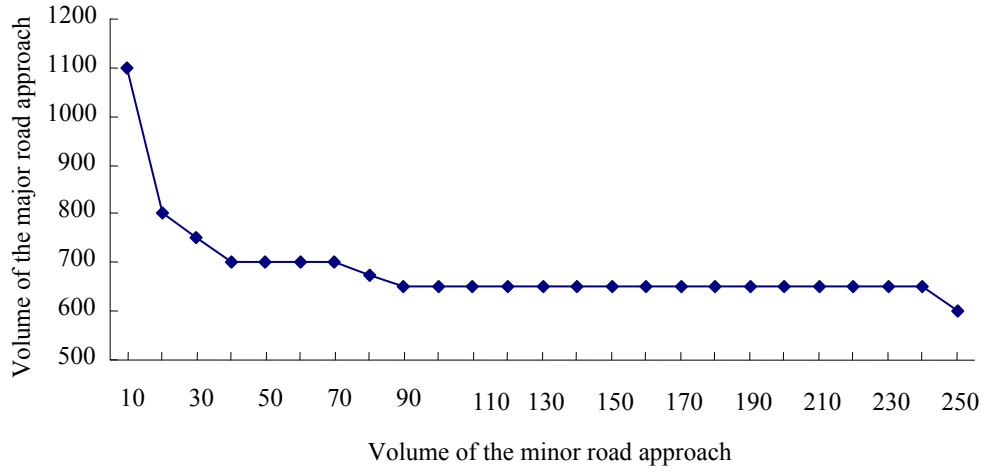


Figure 5. Access control standards of traffic volumes for Scenario 2.

***Scenario* 2.** For Scenario 2, left turns and right turns from the minor road approach are both permitted. Besides, a left-turn lane and a right-turn lane are provided separately for left turns and right turns. Intersection layout for Scenario 2 is shown in Figure 4. Because right turns now use their own right-turn lane, they are not to be impeded by left turns. As a result, average delay time of right turns is much shorter than that of left turns. Therefore in our analysis of Scenario 2, average delay time of left turns of the minor road approach were used to find the standards of traffic volumes, instead of the overall average delay time of the minor road approach. Results are shown in Figure 5, which gives the standards for access control of Scenario 2. Traffic volumes of access points should be below the curve. Otherwise, some measures should be taken on the minor road approach, such as prohibition of left turns or closure of the minor road approach.

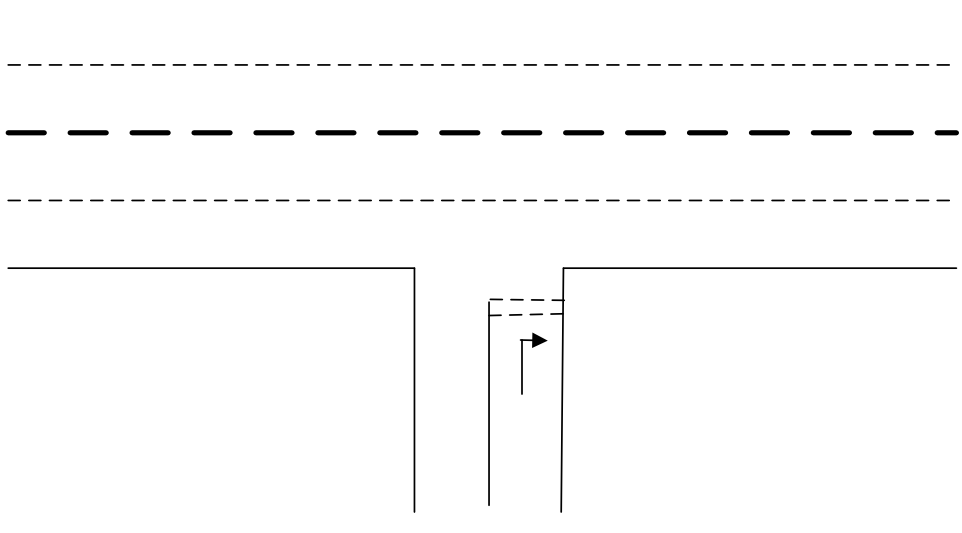


Figure 6. Intersection layout for Scenario 3.

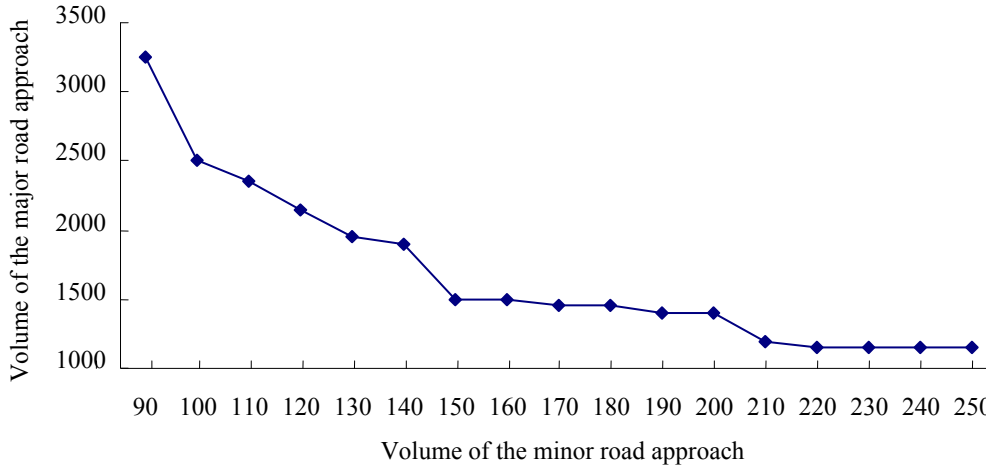


Figure 7. Access control standards of traffic volumes for Scenario 3.

***Scenario* 3.** For Scenario 3, only right turns from the minor road approach are permitted, and a right turn lane is provided for right turns. Intersection layout for Scenario 3 is shown in Figure 6. Since only right-turn traffic exists on the minor road approach, it was average delay time of right turns that were analyzed to set up the standards for access control, which are shown in Figure 7. Access design of Scenario 3 is reasonable if traffic volumes are below the curve. Otherwise, the access point will have to be signalized or closed down.

Conclusions

Using microscopic traffic simulation, the paper has set up standards of traffic volumes for access control of arterial access points. Although higher levels of access control should be considered when traffic volumes are above the standard, well-being of those travelers, impacted adversely, should be protected as much as possible. For example, when left turns are prohibited, traffic signs should be installed to indicate alternative paths. The study is based upon some assumptions, and there exist some other conditions that have not been considered in the analysis. Further study is needed to expand and modify the results of the paper.

Acknowledgments

This work is substantially supported by the National Natural Science Foundation Council of China (Project Number 50778040), the Western Transportation Construction Science and Technology Project from the Ministry of Transport of China (Project Number 20083180008), and the Research and Development Project of the Ministry of Housing and Urban-Rural Construction (Project Number 2009-k5-12).

References

Committee on Access Management (2003). "Access management manual", *Transportation Research Board*, Washington, D.C.

Ge, X., Lu, J., Xiang, Q. J., and Wang, P. Y. (2008). "Study on driving behavior and traffic conflict at highway intersection", *2008 Workshop on Power Electronics and Intelligent Transportation System*, 492-495.

Jha, M., Gopalan, G., and Garms, A. (2004). "Development and Calibration of A Large-scale Microscopic Traffic Simulation Model", *Calibration and Validation of Simulation Models*, Issue: 1876, 121-131.

Lu, J., Xiang, Q. J., and Zhang, G. Q. (2007). *Research on Traffic Safety Technique for Highway Intersections in China*. Unpublished.

Lu, J., Yuan, L., and Zhang, G. Q. (2007). "Diagnostic Approach for Safety Performance Evaluation and Improvements of Highway Intersections", *Annual Meeting of Transportation Research Board,* Washington D.C., CD-ROM (07-1291).

Eisele, W. L., and Toycen, C. M. (2005). *Identifying and quantifying operational and safety performance measures for access management: micro-simulation results* (Report 167725-1), Texas Transportation Institute.

Zhang, G. Q., Lu, J., Wang, W., and Xiang, Q. J. (2008). "Safety performance evaluation of highway intersection based upon traffic flow theories" *2008 IEEE International Conference on Service Operations and Logistics, and Informatics*, 1459-1463.

Zhang, G. Q., and Chen J. (2009). "Study on saturation flow rates for signalized intersections", *2009 International Conference on Measuring Technology and Mechatronics Automation*, 598-601.

Effects of Freeway Exit Ramp Types on Traffic Operations

Yongfeng Ma[1], Junping Xie[2], Jian Lu[3] and Chuan Lu[4]

[1] Lecturer, [3] Prof., School of Transportation, Southeast University, Nanjing 210096, China, Email: mayongfeng@sohu.com, Email: jianjohnlu@sina.com
[2] Lecture, School of Automobile and Traffic Engineering, Jiangsu University, Zhenjiang 212013, China, Email: junpingshe@gmail.com
[4] Assistant Engineer, Anhui Highway Transportation Administration Bureau, Hefei 230001, China, Email: andylu8153@126.com

Abstract

Some proposals on classifying freeway exit ramps in China are put forward. Google Earth was used to identify the contributing factors on exit ramp classification by surveying more than 500 samples in China, USA, Japan, and Europe. The factors determined are configuration of deceleration lanes, number of deceleration lanes, lane changes in the mainline upstream and downstream segments, and number of ramp lanes. With the classification, four freeway exit ramp types are proposed. Furthermore, traffic operational performances of different types of freeway exit ramps were evaluated by studying traffic operational speeds (in the upstream segment, downstream segment, and exit ramp segment) and capacities by TSIS package and field observations. Thus, reasonable exit ramp types are proposed in order to satisfy the different operational requirements.
Key Words: exit ramp; type; freeway; traffic operation; operational speed; capacity

Introduction

As the development of the economy in China, the construction of freeway speeds up. The freeway mileage in China has reached 60,300 kilometers in 2008, and ranked second in the world in terms of total freeway miles (MOC 2009). On the other hand,

the problems which are posed by freeways are rather serious. During 2003 and 2007, there were 105,687 crashes on China freeways, which resulted in 30,588 fatalities, 77,505 injuries, and 2.62 billion RMB direct costs totally (web news 2009). Freeway exit ramps are the sites with many driving behaviors and conflicts, which could result in higher traffic crashes compared with other freeway segments or sites. Inadequate design could result in reduced capacity on exit ramps, which could result in traffic "spill out" to main freeway lanes and significant reduction in main segment capacity. Besides, inadequate design may cause high crash frequency in the impact areas of exit ramps. Statistical data have shown that freeway exit ramps could result in about double number crashes compared to freeway entrance ramps (McCartta 2004).

In the past years, the researchers in many western countries have studied on ramp classification, ramp location determination, spacing between ramps, and traffic control at the ramp terminals (Lundy 1965, Chen 2007, TRB 2000, AASHTO 2004, Zhou 2004, Lu 2007). In China, the most research has focused on ramp capacity analysis, characteristics of merging and diverging traffic on ramps, and spacing between ramps (MOC 2006, Li 2004, Zhang 2002, Long 2005). For exit ramp classification research in USA, the classification is mainly based on ramp geometric characteristics, including the way how traffic diverges from the main traffic stream upstream exit ramps, number and length of the deceleration lanes, number of lanes of exit ramps, the characteristics of merging and diverging paths on ramps, and other factors. For the traffic capacity analysis of different exit ramp types, empirical methods based on regression analysis and traffic simulation methods have been widely applied (Batenhorst 2000, Li 2003).

Currently, there are many urgent issues related to traffic operations and safety on exit ramps in China. Particularly, there is no specific classification method that has been adopted in freeway design in China, which could result in difficulty in selecting adequate exit ramp types for given conditions (Wang 2009). This paper presents a part of a research project that aimed to developing safety design procedures for freeway exit ramps. More specifically, the results presented in the paper focus on the systematical classification of exit ramps and evaluation of the impacts of the different exit ramp types on traffic operations.

Classification of Exit Ramps

The main purpose to perform exit ramp classification study is to help the planners or designers for the determination of exit ramp types for given conditions so that the operational performance at exit ramp sites can be improved.

Table 1. Main factors considered in exit ramp classification.

No.	Factor
1	Configuration of deceleration lanes
2	Number of deceleration lanes
3	Changes in number of lanes on upstream and downstream of main traveling lanes
4	Number of exit ramp lanes

Survey of Existing Exit Ramp Configurations. In the survey, Google Earth was used to survey more than 500 exit ramps in China, USA, Japan, and Europe. It was concluded from the survey that the main factors to be included in exit ramp classifications are shown in Table 1.

Exit Ramp Types From the Google Earth survey and field observation, it was found 4 types of the exit ramps mainly in China, which are shown in Figure 1 to Figure 4. The differences among the four types are discussed as follows:

Type A: Single-lane exit with a tapered deceleration lane is illustrated in Figure 1. In this type, the number of deceleration lane is single and the configuration of deceleration lane is the tapered design. The number of the ramp lane is also single.

Type B: Single-lane exit with a parallel deceleration lane is illustrated in Figure 2. In this type, the number of deceleration lane is single and the configuration of deceleration lane is the parallel design. The number of the ramp lane is also single.

Type C: Two-lane exit with two parallel deceleration lanes is illustrated in Figure 3. In this type, the number of deceleration lane is double and both of the configurations of deceleration lanes are the parallel designs. The number of the ramp lanes is also double.

Type D: Two-lane exit with two successive deceleration lanes is illustrated in Figure 4. In this type, the number of deceleration lane is double and both of the configurations of deceleration lanes are the parallel designs. The outer deceleration lane is usually shorter than the other one to shape two successive deceleration lanes. The number of the ramp lanes is also double.

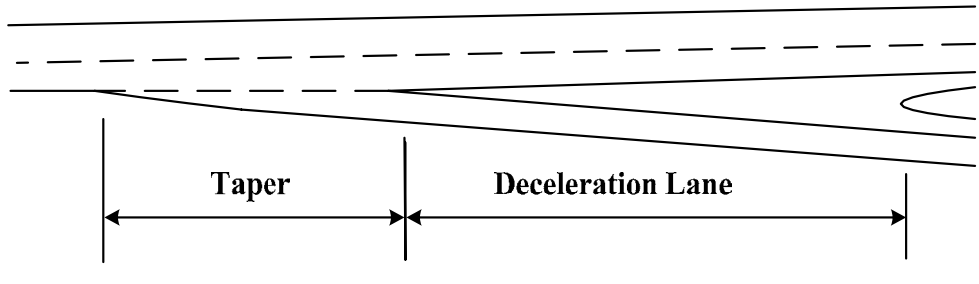


Figure 1. Type A: Single-lane exit ramp with tapered deceleration Lane.

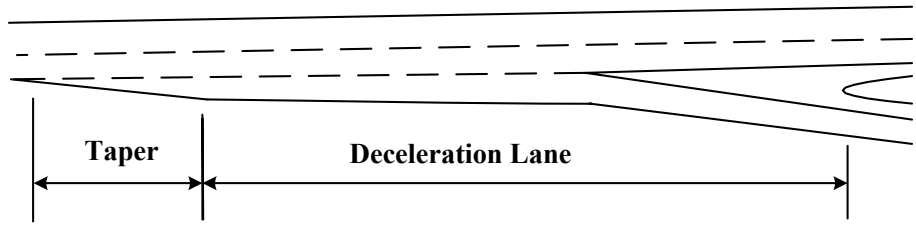


Figure 2. Type B: Single-lane exit ramp with parallel deceleration lane.

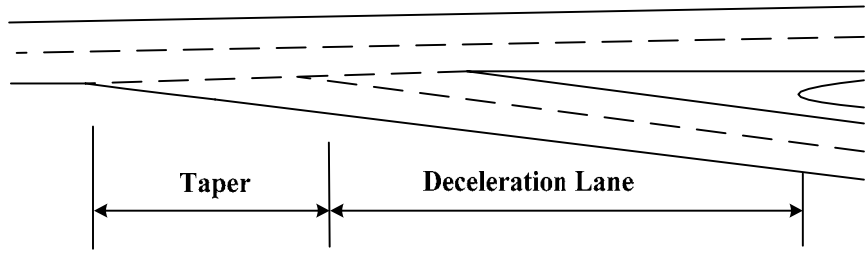


Figure 3. Type C: Two-lane exit ramp with tapered deceleration lanes.

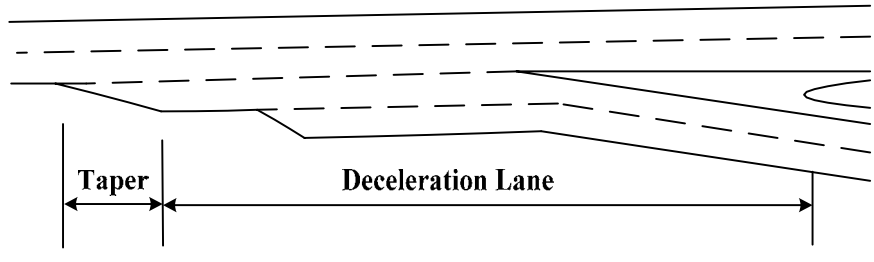


Figure 4. Type D: Two-lane exit ramp with successive deceleration lanes.

To validate the classification of exit ramps, 270 samples of exit ramps in Beijing, Shanghai, Jiangsu, Zhejiang, Hunan, Guangdong, Shanxi, Sichuan, Yunnan and Chongqing in China have been classified according to the 4 exit ramp types by Google Earth. The results are shown in Table 2. The percentages of different types are shown in Figure 5. The total number of 4 types occupies about 96.3% of the total samples surveyed, which proves the rationality of the classification.

Table 2. Number of different types of exit ramps surveyed.

Type	Sample
A	212
B	30
C	15
D	3
Others	10
Total	270

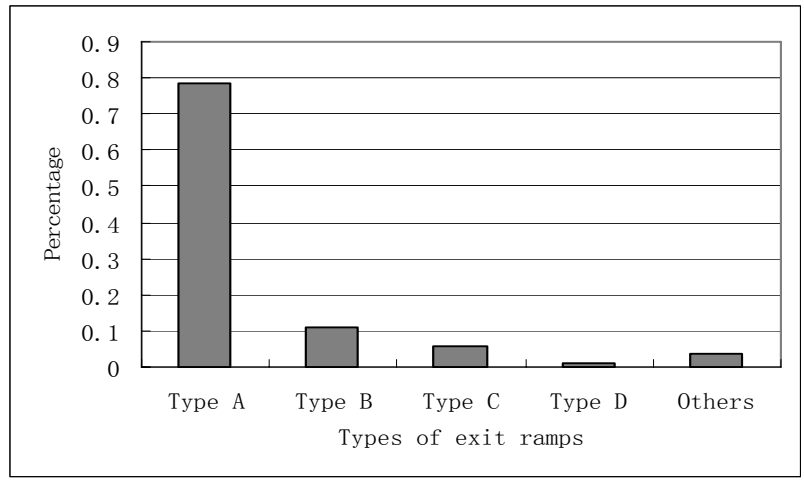


Figure 5. Percentage of different types of freeway exit ramps surveyed.

Impacts of Exit Ramp Types on Traffic Operations

Traffic operations of freeway exit ramps can be measured by operational speed of the vehicles traveling around exit ramp terminals and the capacity of exit ramps. In this study, the simulation package, TSIS, was used to analyze traffic operational performances of different exit ramp types with other given conditions. Basically, TSIS can simulate all kind exit ramp configurations so that the impacts of exit ramp types on traffic operations can be analyzed (SEH 2003). Based on a large scale field survey (through field observations and Google Earth survey), the parameters of TSIS-package were calibrated in order to be applied in China. Field survey and traffic observation using video recording methods were performed to obtain traffic data in the upstream and downstream segments of the exit ramps, including traffic volume, traffic composition, headway, speed distribution, exit traffic volume and composition, and other relevant field data. These field data are very important for simulation and traffic operational analysis.

TSIS Calibration. It is very critical that the TSIS models be calibrated to replicate existing conditions. To calibrate the parameters of TSIS, a prototype of North Yangzhou Exit on Yangli Expressway (eastward) was set up and the traffic data were collected using digital videos and Bushnell speed guns. The calibration steps are given as follows:

(1) Measure the average vehicle speeds and volumes in the mainline and ramp of the North Yangzhou Exit.

(2) Input the initial values of the parameters and the field traffic data collected to the TSIS prototype.

(3) Collect the simulated output of vehicle speeds and volumes in the mainline and ramp.

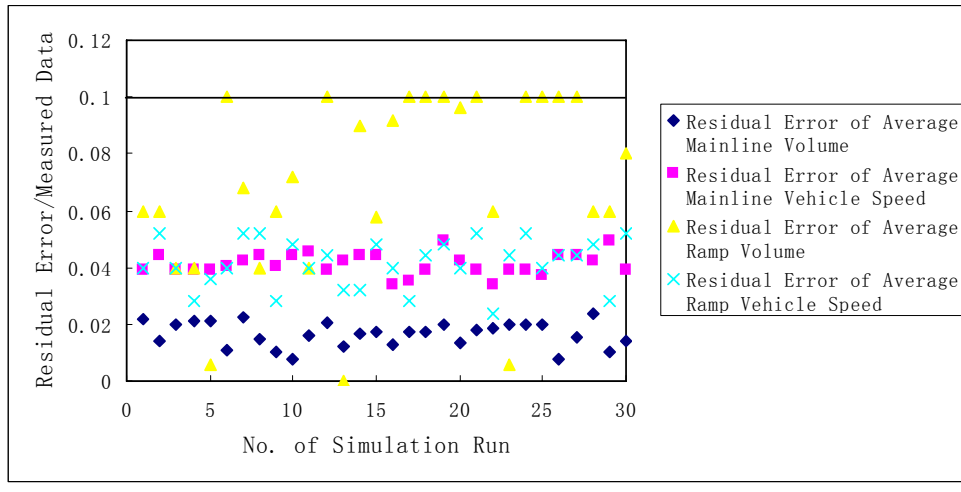


Figure 6. Residual error ratios (residual error / measured date) of average volume and vehicle speed in 30 continuing simulation runs.

(4) Calculate residual errors of vehicle speeds and volumes as:

$$RES = | SIM\text{-}DET |$$

where:
SIM = the simulated average volume(vph) or average vehicle speed (mph);
DET = the measured average volume(vph) or average vehicle speed (mph).
(5) If both of the residual errors of volumes in the mainline and ramp are within 10% of the measured data; and both of the residual errors of vehicle speeds in the mainline and ramp are within 20% of the measured data, the statistics are acceptable and the model is calibrated. If they are not, then modify the parameters until the statistics are acceptable.

The statistics of the TSIS model finally adjusted are shown as Figure 6, and 30 continuing Simulation Runs were carried out to reduce the random errors (SEH 2003). All of the residual errors are less than 10% of the measured average volumes and vehicle speeds on the mainline and ramp in Figure 6, which proves the validity of the TSIS model.

Measurements of Effectiveness There are many measurements that can be used to evaluate traffic operational analysis. However, in order to select adequate measures of effectiveness to evaluate traffic operational performance in the exit ramp area, the feasibility for obtaining reasonable data should be considered. In this research, two measurements were used as measures of effectiveness to quantify the impacts of different exit ramp types on traffic operations: 1) traffic operational speeds in the upstream segment, downstream segment, and exit ramp segment; 2) traffic capacities of different exit ramp types.

Impacts of Exit Ramp Type on Operational Speed Under the given conditions: 2 mainline lanes for one direction, 100 kph free-flow speed on mainline, 40 kph free-flow speed on ramp, 20% turning rate and 40% of heavy vehicle composition, the simulation analysis was conducted with a range of 0 vph to 5100 vph increasing by the pace of 300 vph. The main purpose for the simulation analysis is to determine the best exit ramp type under given conditions. As stated previously, the operational speeds on mainline upstream segment, mainline downstream segment, and exit ramp segment were used as the measures of effectiveness. 30 continuing runs of simulation analysis under same given conditions were performed and the average numbers were used to represent the simulation results.

Figure 7 presents the simulation results for operational speeds on the mainline upstream segment. It can be seen that exit ramp types B and D have the best performance, meaning the minimum impacts on operational speeds on mainline upstream segment with a traffic volume changing from 1500 vph to 3900 vph. Within the same traffic volume range, exit ramp types A and C could result in rapid speed reduction.

Figure 8 presents the speed change of traffic on mainline downstream segment. The simulation results suggest the similar conclusions as suggested by Figure 7. From Figure 8, it can been seen that exit ramp types B and D show the best performance with minimum speed reduction and exit ramp types A and C could result in significant speed reduction, which indicates that exit ramp types A and C may not be suitable if high traffic demand exits.

Figure 9 shows the simulation results for operational speed of traffic on exit ramp segment. The simulation results suggest that the types of the exit ramps have little impacts on the vehicle speeds running on the exit ramp segments.

From the simulation results presented in Figures 7, 8, and 9, it is clearly seen that exit ramp types A and C could result in significant reduction of operational speeds of traffic on mainline upstream and mainline downstream around the exit ramp terminal as traffic volume increases. This means that the lane configuration of exit ramp type could generate great impacts on exiting vehicles. These two types of exit ramps can not be suitable for the situation with high traffic speed demand.

By the field observations, two appearances were observed: (1) In all exit impact areas of 4 exit ramp types, similar operational speeds were observed; (2) Under all reasonable traffic demand conditions, traffic operational speeds on all segments in the impact areas are relatively stable with minimum speed reduction. The reason of the first appearance may be the volumes on the freeways observed were on a low level far from the congestion.

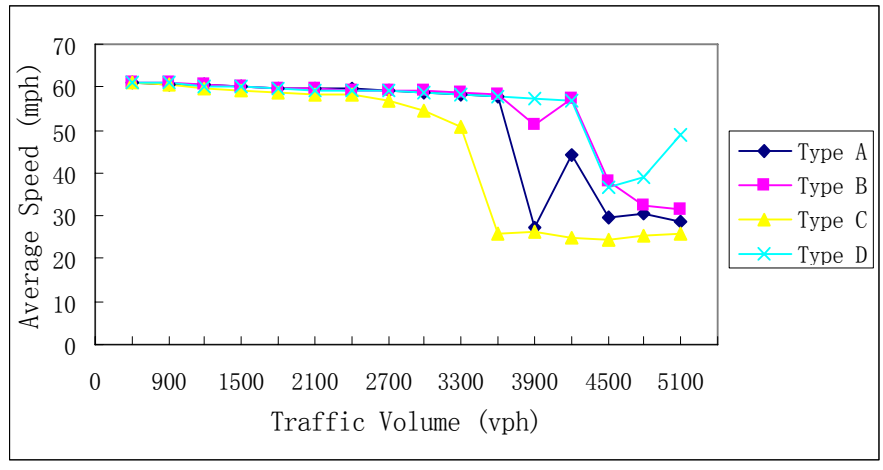


Figure 7. Change of operational speeds on mainline upstream segment.

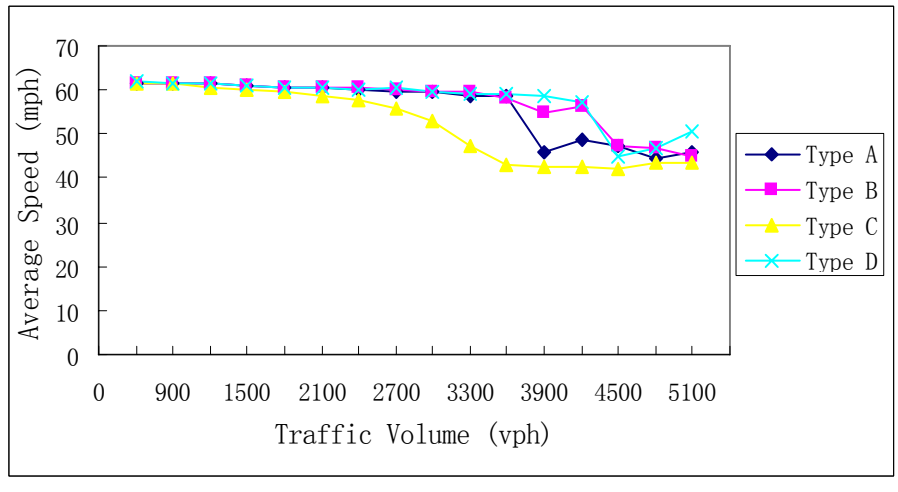


Figure 8. Change of operational speeds on mainline downstream segment.

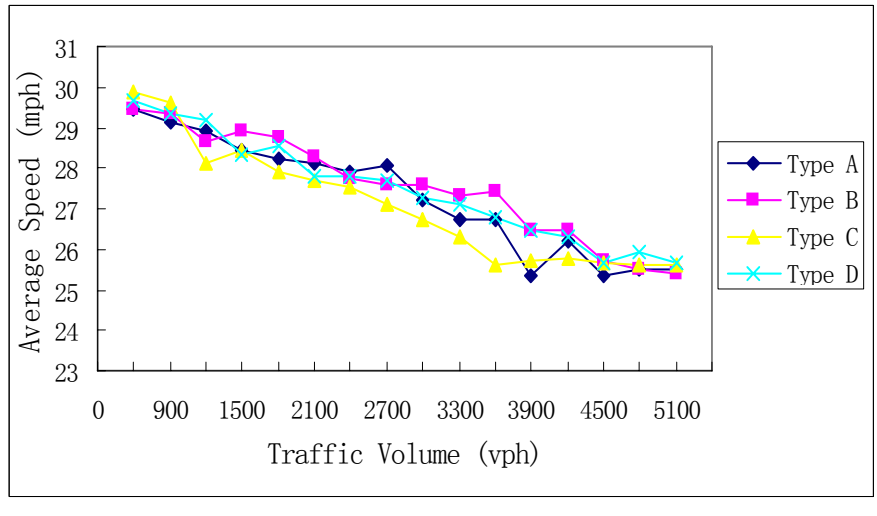


Figure 9. Change of operational speeds on exit ramp segment

Table 3. Capacity of freeway exit ramp under simulation conditions.

Type	Capacity (vph)
A	2000
B	2200
C	2900
D	3200

Impacts of Exit Ramp Type on Capacity To compare the capacities of the 4 types of exit ramps, the vehicle turning rate was raised up to 100% in TSIS package. With the volume inputted increasing by a pace of 100 vph, the vehicles running though the exit ramp terminal and into the exit ramp segment were increasing in number. When the package presented "packed up", the volume could be considered as the capacity of exit ramp under the simulation conditions. Table 3 shows the results of the capacities of different exit ramps.

It can be concluded from Table 3 that the types A and B have lower capacities than types C and D because of the number of deceleration lane and ramp lane; on the other hand, types B and D have higher capacities than types A and C, respectively, because of the configurations of the deceleration lanes. The latter finding should raise concerns to the researchers and planners.

Conclusions

With the wide investigation by Google Earth and filed observations, more than 500 freeway exit ramp samples were studied to summarize the main factors in freeway exit ramp classification. Based on configuration of deceleration lanes, number of deceleration lanes, lane changes in the mainline upstream and downstream segments, and number of ramp lanes, freeway exit ramps were classified into four groups which represent the most widely used ones in China.

In the research illustrated in this paper, traffic operational performances were evaluated with the use of TSIS package. It was concluded that under lower traveling demand, all types of exit ramps could provide reasonable traffic service in traffic operations. However, when traffic volume is higher, the planners should determine the number of the deceleration lanes and ramp lanes (single or double); and then type B and type D are recommended for single-lane and two-lane exit ramps, respectively, as the little reduction in operational speeds and high capacities.

The research presented in this paper could provide certain guidelines for selecting exit ramp types with the considerations of operational performances. However, the

main results were reached under the conditions of the computer simulation in this paper. In fact, many factors could affect the selection of exit ramp types. More researches should be carried out to evaluate contributing factors that affect traffic operational performances.

References

Ministry of Transport of the People's Republic of China (MOC) (2009). "The Roadway and Waterway Transportation Development Gazette-2008." http://www.moc.gov.cn/zhuzhan/tongjixinxi/fenxigongbao/tongjigongbao/200904/t20090429_577812.html.

"Traffic Accidents Analysis for Recent 5 Years in China"(web news 2009). http://bbs.auto.ifeng.com/thread-31231-1-1.html.

McCartta, A. T., Northrupb, V. S., and Retting, R. A. (2004). "Types and Characteristics of Ramp-related Motor Vehicle Crashes on Urban Interstate Roadways in Northern Virginia." *Journal of Safety Research*, Vol. 59, 107–114.

Lundy R. A. (1965). "The Effect of Ramp Type and Geometry on Accidents." *California Department of Public Works, Division of Highways,* Sacramento, California.

Chen, H. Liu, P., Behzadi, B., and Lu, J. (2007). "Impacts of Exit Ramp Type on the Safety Performance of Freeway Diverge Areas." In Pavement Management CD-ROM. *Transportation Research Board of the National Academies*, Washington, D.C.

TRB (2000). "Highway Capacity Manual". Transportation Research Board, National Research Council, Washington, D.C.

AASHTO (2004). "A Policy on Geometric Design of Highways and Streets." American Association of State Highway and Transportation Officials. Washington, D.C.

Zhou, H. (2004). "Computer Simulation Analysis of the Impacts Exit Ramp Length on Exit Ramp Delay." Center for Urban Transportation Research, University of South Florida. Tampa, Florida.

Lu, J. Liu, P., and Behzadi, B. (2007). "Safety Evaluation of Freeway Exit Ramp." *Operational Effects of Geometric* CD-ROM. Transportation Research Board of the

National Academies, Washington, D.C.

Ministry of Transport of the People's Republic of China (MOC). (2006). "Specifications for Highway Geometric Design (JTG D20-2006)".

Li, W., Wang, L., and Wang, W. (2004). "Modeling for Capacity in the Merging Areas of Freeway Ramps." *Journal of Transportation Engineering*, 4(2), 80-84.

Zhang, P. (2002). "Analysis of Types and Function of Ramp." *Journal of Hebei Institute of Architecture Engineering*, 4, 22-24.

Long, K. J., Yang, X. G., Wang, Y. H., and et al (2005). "Models for determining minimum distance of ramps on urban expressway." *Journal of Traffic and Transportation Engineering*, 5(1), 106-110.

Batenhorst, R. A., and Gerken, J. G. (2000). "Operational Analysis of Terminating Freeway Auxiliary Lanes with One-Lane and Two-Lane Exit Ramps: A Case Study." *Mid-Continent Transportation Symposium 2000 Proceedings*.

Li, J. and Fang, X. (2003). "Safety Studies of Freeway Exit and Entrance Ramps". *Journal of Central Southern China Highways*. 2(28), 31-34.

Wang, P. Y., Lu, J., Ma, Y. F., and et al. (2009). "Research Status and Prospects of Freeway Off-ramp Geometric Characteristics and Control Modes". *Journal of Transportation Engineering and Information*, 7(1), 104-110.

SEH. (2003). "Advanced CORSIM Training Manual (Ver. 11). *Minnesota Department of Transportation*.

Capacity Analysis of Type-A Weaving Section by Using Cellular Automata Model

Xingang Li, Ziyou Gao and Bin Jia

MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China; Email: lsinban@gmail.com

Abstract

In this paper, the capacity of the weaving section, in which an on-ramp and an off-ramp are joined together by an auxiliary lane, influenced by the inflow rates of drive-in and drive-out vehicles is investigated. The drive-in vehicle should merge into the main road while the drive-out vehicle should diverge from the main road. When the number of drive-out vehicle is larger than the number of drive-in vehicle, the total flux can be reduced to lower values. The drive-out vehicle is difficult to enter into the off-ramp then it queues up at the weaving section and blocks the following traffic. Thus the flux on the off-ramp is much lower than the inflow rate of drive-out vehicle and the traffic on the main road is greatly disturbed. As to the length of the weaving section, it should be better larger than 150 m.
Key Words: capacity; weaving section; cellular automata

Introduction

Weaving is defined as the crossing of two or more traffic streams traveling in the same general direction along a significant length of highway without the aid of traffic control devices (HCM2000). Weaving areas are formed when a merge area is closely followed by a diverge area, or when an on-ramp is closely followed by an off-ramp and the two are joined by an auxiliary lane. The traffic behaviour around the weaving section is complex. The drive-in vehicle, which comes from the on-ramp, should merge into the main road. And the drive-out vehicle, which enters into the off-ramp, should diverge from the main road. The mandatory lane-changing behaviour performed by the drive-in vehicle and the drive-out vehicle brings serious disturbance to the traffic flow and causes capacity drop. The weaving section is a bottleneck to the freeway. The capacity of a weaving section varies as a function of

the prevailing proportion of weaving and non-weaving flows. As the proportion of weaving traffic increases, the amount of lane changing also increases, which in turn lowers the capacity of the weaving segment. But the exact relationship between the capacity and the proportion of weaving and non-weaving flows is difficult to estimate. There is not commonly accepted method for estimating the capacity of weaving section.

In the early 1970s through the early 1980s, the freeway weaving analysis procedures are based on estimating the speeds for the weaving and non-weaving traffic streams (Pignataro et.al, 1975; Roess and Ulerio, 2000). And this is the method used in HCM 2000. Subsequent research efforts have shown that the method's ability to predict the operation of a weaving section is limited (Rakha and Ahn, 2004), which is most probably due to the outdated and limited database that was utilized to develop these models. Since 1980s, several software have been developed to estimate the capacity of weaving section, such as WEAVSIM (Zarean and Nemeth, 1988), INTRAS (Skabardonis et.al 1989), INTEGRATION (Stewart et.al 1996), FORSIM (Vermijs 1998), etc. Now INTEGRATION has become an professional and useful tool for estimating the capacity of weaving section (Rakha and Zhang, 2006; Zhang and Rakha, 2009). The gap-acceptance method for estimating the capacity is also developed (Lertworawanich and Elefteriadou, 2002; Lertworawanich and Elefteriadou, 2003). In this paper, the simulation method is used to estimate the capacity of weaving section, and the cellular automata (CA) model is adopted.

CA model is an excellent tool for simulating real traffic flow, because its efficient and fast performance when used in computer simulations. Nagel and Schreckenberg (1992) proposed the well-known NaSch model for single-lane traffic flow. Later the two-lane models were presented by introducing lane-changing rules (Rickert et.al, 1996; Chowdhury et.al, 1997). Nowadays, CA model has been widely used to simulate bottleneck traffic, which includes on-ramp, off-ramp, lane closing, etc (Jiang et.al, 2002; Jia et.al, 2003, 2004, 2005,). Several models for weaving section has also been proposed (Huang, 2006; Lei et.al, 2006). However, those models are mainly focus on the traffic dynamics around weaving section. Recently, we proposed a cellular automaton model for weaving section to investigate the capacity influenced by the weaving vehicle (Jia et.al, 2010). In this paper, the capacity of weaving section is systematically studied by using the model in our previous work.

The weaving section commonly exists on the expressway system in Beijing, such as the 3rd and 4th ring road. And most of the weaving section is of type-A, in which the weaving movement need once and only once lane change. So here only the type-A weaving section is simulated and analyzed by using the CA model.

This paper is organized as follows: In section 2, the model simulating the weaving section is introduced. In section 3, the simulation results are discussed. At last, the conclusion is given.

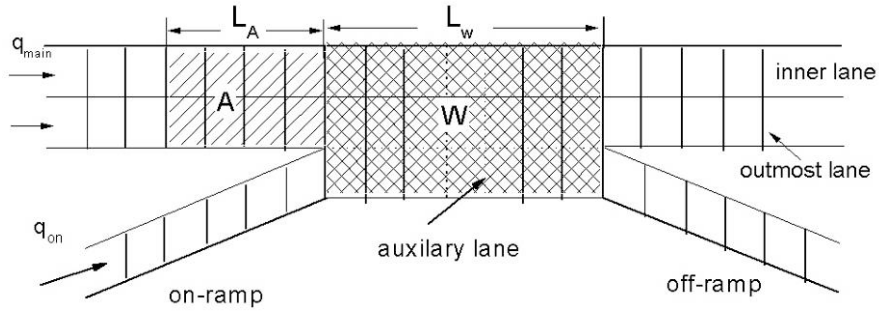


Figure.1. The schematic of the weaving section in the case of the main road contains two lanes.

Model

In this work, the CA model for weaving section (Jia et.al, 2010) is used. Next, the model is briefly reviewed. The refined NaSch model is used to describe the forward movement of the vehicle. The updating rules of the refined NaSch model contains 4 sub-steps: (1) acceleration: $v_n(t+1/3) \rightarrow \min(v_n(t)+1, v_{\max})$; (2) deceleration: $v_n(t+2/3) \rightarrow \min(v_n(t+1/3), d_n)$; (3) randomization: $v_n(t+1) \rightarrow \max(v_n(t+2/3)-1,0)$ with probability p; (4) position update: $x_n(t+1) \rightarrow x_n(t)+v_n(t+1)$. Here $v_n(t)$ and $x_n(t)$ denote the velocity and the position of the nth vehicle respectively; $v_{\max}$ is the maximum velocity; $d_n = x_{n+1}(t) - x_n(t) - l$ denotes the number of empty cells in front of the nth vehicle, and l is the length of the vehicle, i.e., the number of cells that taken up by the vehicle; p is the randomization probability. In the original NaSch model, l=1 is selected. Here l=3 is used.

The schematic of the weaving section is shown in Fig.1. One can see that an on-ramp and an off-ramp are joined together by an auxiliary lane, which is adjacent with the main road. Here only the case that the main road contains two lanes is plotted. While the cases that the main road contains three and four lanes are also studied in this paper. As one can see the on-ramp and the off-ramp have one lane. The weaving section W has a length of L_W cells. There is a section A with a length of L_A on the main road, and the control measures in section A are as follows: (1) the drive-out vehicle must change to the outmost lane before it reaches the right edge of section A; (2) the vehicle, which drives to the right part of the main road, is not allowed to change to the outmost lane in section A. In the weaving section W, the drive-out vehicle on the main road must change to the auxiliary lane before it reaches the right edge of section W; the drive-in vehicle, which comes from the on-ramp and drives into the main road, must change to the outmost lane before it reaches the right edge of section W. To avoid the gridlock that both the drive-out vehicle and the drive-in vehicle reach the right edge of W but have not completed the lane-changing, it is assumed that the drive-out vehicle is not allowed to enter into the last $2l$ cells at the right end of section W.

According to the motivation of lane-changing, the lane-changing rules can be classified into two kinds : discretionary and mandatory. Discretionary lane changes

are those which are only desired when the driver wishes to move from a slower lane to a fast moving one. Mandatory lane changes arise due to the vehicle approaching the entrance which it must pass through to continue the journey. The discretionary lane changes on the two-lane main road are modeled by the following symmetric rules (Rickert et.al 1996):

$$d_n < \min(v_n+1, v_{\max}) \text{ , } d_{n,other} > d_n \text{ and } d_{n,back} > v_{\max}. \quad (1)$$

Here $d_{n,other}$, $d_{n,back}$ denote the number of free cells between the nth vehicle and its two neighbour vehicles on the other lane at time step t. $d_{n,back} > v_{\max}$ is the security criterion and $d_n < \min(v_n+1, v_{\max})$, $d_{n,other} > d_n$ is the incentive criterion.

On the auxiliary lane, the drive-in vehicle is eager to change to the main road; on the outmost lane in section W, the drive-out vehicle has to change to the auxiliary lane; in section A, the drive-out vehicle is inclined to run on the outmost lane for the convenience to exit and the driving ahead vehicle is inclined to run on the inner lane. Such behaviour is mandatory lane change and the lane-changing rule is described as follows:

$$d_{n,other} \geq l \text{ and } d_{n,back} \geq v_{n,back}. \quad (2)$$

Here $v_{n,back}$ is the velocity of the vehicle behind on the destination lane. It means that the vehicle will change lane as long as the conditions on the desired lane is convenient for driving and the gap to the following vehicle is large enough. One should note that if an drive-out vehicle on the inner lane of the main road is not able to change to the outmost lane when it reaches the right edge of section A, it will stop and wait for a chance to change to the outmost lane. The following traffic can be seriously disturbed by the stop vehicle. In order to alleviate the disturbance induced by the stop vehicle, the rule that the vehicle on the right lane makes room for the stop vehicle is adopted. That is to say, when a vehicle find out an drive-out vehicle is stopping on the adjacent left lane ahead, it will not pass the stop vehicle until the stop vehicle complete the lane change.

The number of lanes on the main road is N. The inflow rate of each lane on the main road is α_m veh/h. The inflow rate of on-ramp is α_1 veh/h. And the inflow rate of drive-out vehicle, i.e, from main road to off-ramp, is α_2 veh/h. The vehicle enters into the system at the right boundary. At each time step, a vehicle can enter into each lane of the main road (on-ramp) with probability $\alpha_m/3600$ ($\alpha_1/3600$) as long as the first $v_{\max}$ cells of the lane are not occupied by any vehicles. And the position of the new vehicle is $\min(x_{last}- v_{\max}, v_{\max})$, where x_{last} denotes the position of the last vehicle on the lane. The vehicle entering into the main road is set as an drive-out one with probability $\alpha_2/3600/N$. At the right end of the main road and the off-ramp, the vehicle is removed if its position is larger than the length of the road.

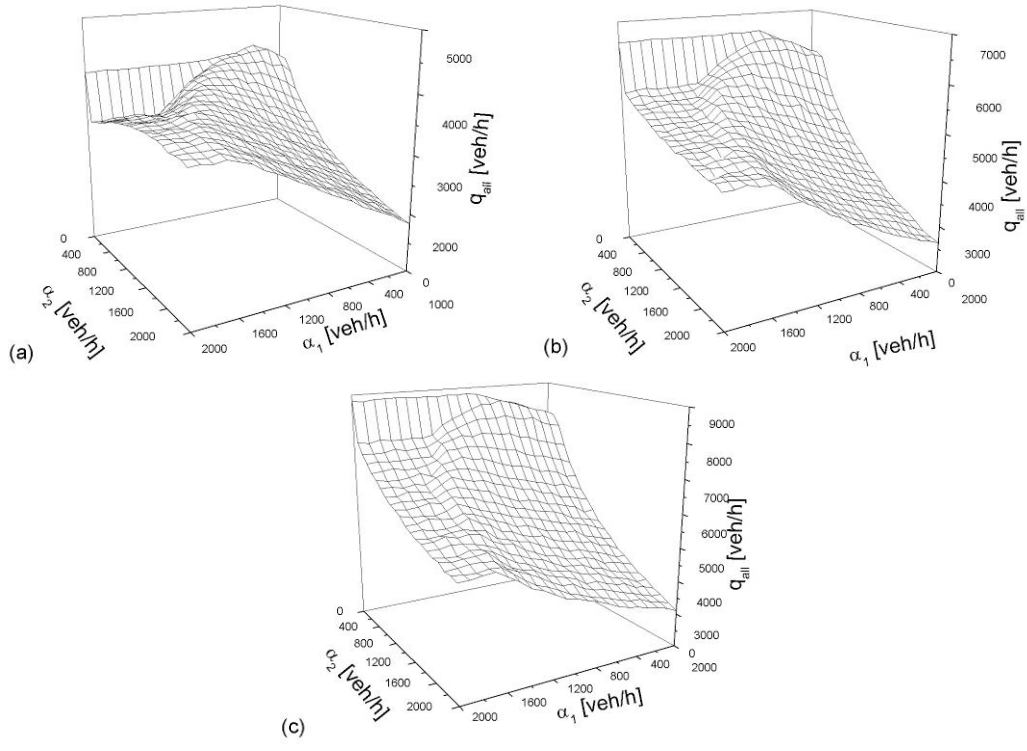


Figure 2. The total flux q_{all} as a function of α_1 and α_2. (a) N=2; (b) N=3; (c) N=4.

Simulation Results

In this section, the simulation results are presented. In the simulations, the main road is divided into $N \times L$ (L=12000) cells, and the weaving section is at the center of the road. The on-ramp, the auxiliary lane and the off-ramp are joined together as one lane with the length of L. And we call this lane as ramp lane. Each cell corresponds to 2.5 m and a vehicle has a length of three cells. One time step corresponds to 1 second. The model parameters $v_{\max}$=15 and p=0.1 are used. L_A=60 and L_W=60 are chosen if it is not specifically mentioned. Virtual detectors are installed at upstream and the downstream of the weaving section. Then the flux on each lane can be recorded. The flux on the main road is q_m veh/h, that on the on-ramp is q_{on} veh/h and that on the off-ramp is q_{off} veh/h. The total flux of the weaving section q_{all} veh/h. The first 10000 time steps are discarded to let the transient time die out. The results are obtained in 100000 time steps.

Firstly, the inflow rate of the main road α_m is set as 2000veh/h. And the total flux of the weaving section q_{all} depending on the inflow rate of on-ramp α_1 and the inflow rate of drive-out vehicle α_2 is investigated. Figure 2 shows the value of q_{all} as a function of α_1 and α_2 with different N. One can see that when α_1 is fixed to certain values, q_{all} will first decrease then increase as α_2 increases, and when α_2 is fixed to certain values, q_{all} will first decrease then increase as α_1 increases. Thus a ridge is formed. The total flux q_{all} can be reduced to lower values in the case of $\alpha_2 > \alpha_1$. The lowest value of q_{all} is obtained when α_2=2000 veh/h and α_1=0 veh/h. Such results

indicate that the drive-out vehicle brings much disturbance to the traffic on the main road than the drive-in vehicle. As we know that it is believed that the drive-out vehicle and the drive-in vehicle play the same role in reducing the capacity of weaving section in previous work (Lertworawanich and Elefteriadou, 2003; Rakha and Zhang, 2006). That is to say q_{all} is symmetric to the line $\alpha_1=\alpha_2$. Our results is quite different to the previous work. The total flux in the case of $\alpha_1>\alpha_2$ is higher than that in the case of $\alpha_1<\alpha_2$. We argue that the drive-out vehicle plays an important role in determine the capacity of weaving section.

Figure 3 shows the flux on the on-ramp and the off-ramp as a function of α_1 and α_2. One can see that q_{on} increases linearly with α_1 before it reaches saturated value. While q_{off} grows slowly with α_2 and it is much lower than α_2. So the drive-out vehicle is difficult to enter into the off-ramp then it has to queue up on the main road. The traffic on the main road can be greatly disturbed by the drive-out vehicle and this is the reason why q_{all} is very low when α_2 is large. Such phenomenon has been observed at the weaving section on the 3rd ring road in Beijing. Figure 4 shows the typical traffic congestion caused by weaving vehicles on the 3rd ring road from Sitong bridge to Suzhou bridge. One can see that the drive-out vehicle queues up at the weaving section and blocks the following traffic.

Last the total flux q_{all} depending on the length of weaving section L_W is considered. Here only the case that the main road contains 3 lanes is studied. The inflow rate of the main road is set as α_m=2400 veh/h, so as to the traffic on the main road is in saturated state. α_1 and α_2 are fixed to certain values. The total flux q_{all} as a function of L_W is shown in Fig.5. It shows that q_{all} will be improved as L_W increasing. q_{all} in the case of $\alpha_1=\alpha_2$=500 veh/h is higher than that in the case of $\alpha_1=\alpha_2$=1000 veh/h. Higher proportion of weaving vehicle brings more disturbance into the system, thus the flux is lower. The results also show that q_{all} drops to very small values when L_W is lower than 150 m. In order to guarantee the weaving section having a high capacity, its length should not be shorter than 150 m.

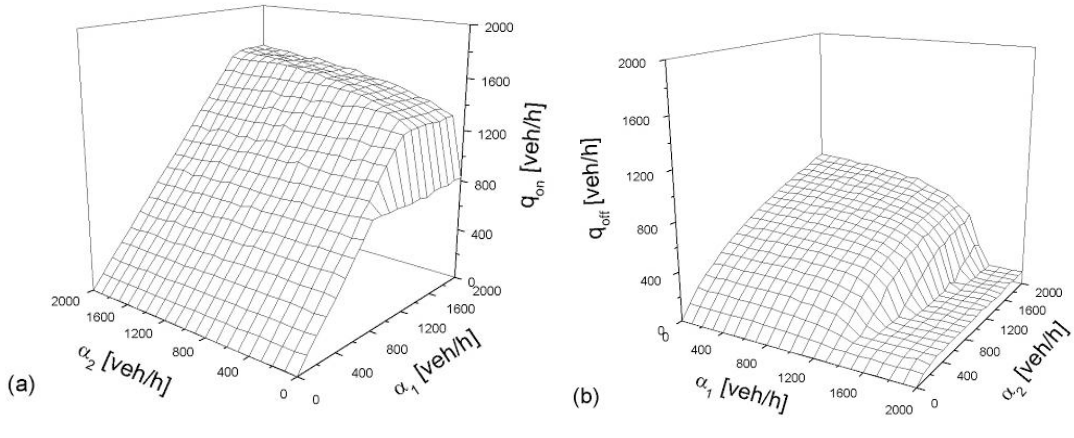


Figure 3. q_{on} (a) and q_{off} (b) as a function of α_1 and α_2.



Figure 4. The typical congestion state at the weaving section on the 3rd ring road from Sitong Bridge to Suzhou bridge in Beijing.

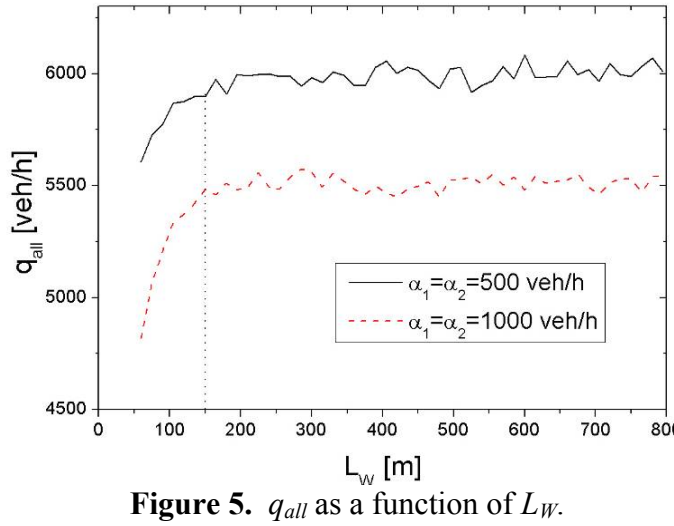


Figure 5. q_{all} as a function of L_W.

Conclusions

In this paper, the capacity of the weaving section depending on the inflow rate of drive-in and drive-out vehicle is investigated. The cellular automata model is used to simulate the traffic dynamics around weaving section.

The simulation results show that the total flux in the case of $\alpha_1>\alpha_2$ is higher than that in the case of $\alpha_1<\alpha_2$. Higher inflow rate of drive-out vehicle depress the total flux to lower values. The drive-out vehicle plays an important role in determine the capacity of the weaving section. It is also shown that the length of the weaving section should not be shorter than 150m.

In our future work, the capacity of other types of weaving section will be investigated by the cellular automata model. We are going to find relationship

between the capacity and the proportion of weaving vehicle.

Acknowledgements

This work is partially supported by National Basic Research Program of China under No. 2006CB705500, the National Natural Science Foundation of China under Key Project No.70631001, Program for New Century Excellent Talents in University (NCET-07-0057), and Beijing Natural Science Foundation (9093020).

References

Chowdhury, D., Wolf, D. E. and Schreckenberg, M. (1997) "Particle hopping models for two-lane traffic with two kinds of vehicles: effects of lane changing rules." *Physica A*, 235: 417-439.

Highway Capacity Manual (HCM), 2000. Transportation Research Board, Special Report 209, Washington DC.

Huang, D. W. (2006). "Ramp-induced transitions in traffic dynamics." *Physical Review E* 73(1): 016123.

Jia, B., Jiang, R., Wu, Q. S. (2005). "The effects of accelerating lane in the on-ramp system." *Physica A* 345(1-2): 218-226.

Jia, B., Jiang, R., Wu, Q. S. (2004). "Traffic behavior near an off ramp in the cellular automaton traffic model." *Physical Review E* 69: 056105.

Jia, B., Jiang, R., Wu, Q. S. (2003). "The traffic bottleneck effects caused by the lane closing in the cellular automata model." *International Journal of Modern Physics C* 14(10): 1295-1303.

Jia, B., Li, X. G., Jiang, R., Gao, Z. Y. (2010). " Traffic behavior around the weaving section in cellular automata model." *International Journal of Modern Physics C*, inpress.

Jiang, R., Wu, Q. S., Wang B. H. (2002). "Cellular automata model simulating traffic interactions between on-ramp and main road." *Physical Review E* 66(3): 036104.

Lei, L., Dong, L.Y., Song, T., Dai, S.Q. (2006). "Study on the traffic flow of weaving section in elevated road system with cellular automaton model." *Acta Physica Sinica* 55: 1711.

Lertworawanich, P., and Elefteriadou, L. (2002), "Capacity estimation for type-b weaving areas based on gap acceptance." *Transportation Research Record* 1776, 23-34.

Lertworawanich, P., Elefteriadou, L., (2003), "A methodology for estimating capacity at ramp weaves based on gap acceptance and linear optimization." *Transportation Research Part B*. 37(5), 459-483.

Nagel, K., Schreckenberg, M., (1992), "A Cellular automaton model for freeway traffic. " *Journal de Physique I (France).* 2, 2221-2229.

Nassab, K., Schreckenberg, M., Boulmakoul, A., Ouaskit, S., (2006). "Effect of the lane reduction in the cellular automata models applied to the two-lane traffic." *Physica A* 369: 841-852.

Pignataro, L.J., McShane, W.R., Roess, R.P., (1975). "Weaving areas design and analysis." *NCHRP Report* 159.

Rakha, H., and Ahn, K., (2004), "INTEGRATION modeling framework for estimating mobile source emissions." *Journal of Transportation Engineering*, 130(2), 183-193.

Rakha, H., Zhang, Y., (2006), "Analytical procedures for estimating capacity of freeway weaving, merge, and diverge sections." *Journal of Tranportation Engnieering.* 132(8), 618-628.

Rickert, M., Nagel, K., Schreckenberg, M., Latour, A., (1996). "Two lane traffic simulations using cellular automata." *Physica A*, 231: 534-550.

Roess, R. and Ulerio, J., (2000), "Weaving area analysis in year 2000 highway capacity manual" *Transportation Research Record*, 1710, 145-153.

Skabardonis, A., Cassidy, M.J., May, A.D., Cohen, S., (1989), "Application of simulation to evaluate the operation of major freeway weaving sections." *Transportation Research Record*. 1225, 91-98.

Stewart, J., Baker, M., Aerde, M.V., (1996), "Evaluating weaving section designs using INTEGRATION." *Transportation Research Record.* 1555, 33-41.

Vermijs, R., (1998), "New dutch capacity standards for freeway weaving sections based on micro simulation." *Third International Symposium On Highway Capacity.* 1065-1080.

Zarean, M., Nemeth, Z.A., (1988), "WEAVSIM: a microscopic simulation model of freeway weaving sections." *Transportation Research Record*. 1194, 48-54.

Zhang, Y., Rakha, H., (2009), "Systematic analysis of capacity of weaving sections." Al-Qadi, I. L., Sayed, T., Alnuaimi, N., Masad, E., (Editors) *Efficient Transportation and Pavement Systems: Characterization, Mechanisms, Simulation, and Modeling.* 151-164.

Length of Acceleration and Deceleration Lane on Urban Expressway

Xinsheng Song[1], Aizeng Li[2], Binghua Wu[3] and Lihua Liu[4]

[1] Associate Professor, [2] Lecturer, [3] Lecturer, [4] Lecturer, Department of Traffic Engineering, Henan University of Urban Construction, New District, Pingdingshan City, Henan Province, China 467044; PH: 86-375-2089318E-mail: sxs@hncj.edu.cn, liaizeng@163.com, wuiceflower@126.com, liulihuaalicia@126.com

Abstract

To ensure traffic flow on urban expressway running safely and efficiently and perfect the designing standard of interchange spacing, it is necessary to study the length of acceleration and deceleration lane on urban expressway. Based on the vehicles' running characteristics, the acceleration lane was divided into three sections, namely acceleration section, waiting merging section and width transition section. Applying dynamics theory, the models calculating the length of acceleration section and width transition section were got. By combining probability theory and differential method and taking main lane traffic volume into consideration, the model calculating the length of waiting merging section was established as well. Then, the twice deceleration theory put forward by AASHTO(American Association of State Highway and Transportation Officials) was adopted to analyze the vehicles' running process on deceleration lane. The deceleration lane was divided into three sections, too, i.e. width transition section, the first and second deceleration section. Applying the dynamics theory again, the model calculating the length of width transition section was got. Applying dynamics theory and taking drivers' driving comfort into consideration, the models calculating the length of the first and second deceleration section were established as well. And then, taking mainline lane vehicles' time headway distribution and mainline lane and ramp vehicles' velocity combination into consideration, the length values of acceleration and deceleration lane on urban expressway were calculated. The calculation results indicate that the length of

acceleration and deceleration lane increases with the increase of mainline lane vehicles' velocity and decreases with the increase of ramp vehicles' velocity.
Key Words: urban expressway; length of acceleration lane; length of deceleration lane; dynamics theory; probability theory; differential method; driving comfort

Introduction

Urban expressway is the arterial road in a city, which is designed for the vehicles with requirements of long distance and quick movement. Urban expressway has been built in many big cities in China, for example Beijing, Shanghai, Guangzhou, and so on, to alleviate the traffic congestion. However, the existing literatures about the expressway are mostly about freeway, there are few about urban expressway not only in aspect of basic theory such as traffic flow characteristics and capacity (Zhang 2005) but in other aspects such as the length of acceleration and deceleration lane. The existing literatures studying the length of acceleration and deceleration lane are mostly about freeway, it can't fit for the traffic flow running characteristics on urban expressway very well. Studying the length of acceleration and deceleration lane on urban expressway will lay a foundation for the traffic flow running safely and effectively. About the existing literatures, He (2005) analyzed the vehicles' running characteristics on acceleration and deceleration lane, researched the length of it from the viewpoint of safety, and established relative calculating models. Shao (2008) pointed out that the velocities of mainline lane vehicles and merging vehicles can be described by its relative time headway, then established models to calculate the length of the acceleration lane. Zhi (2009) studied traffic volume and minimum time headway on freeway mainline lane and acceleration lane, then provided a optimized method to analyze the length of acceleration lane. By analyzing ramp vehicles' merging rate and mainline lane service traffic volume, Zhao (2004) calculated the recommended length values of acceleration lane. Based on the method of probability theory, Shi (2000) researched the length of acceleration lane, and established calculation model. Applying accelerating and waiting theory, Li (2000) put forward a method to calculate the length of acceleration lane on freeway. Chen (2003) analyzed the vehicles' running process on deceleration lane, divided the length of it into three sections, and researched the length of each section. Bared (1999) and Alfredo (2006) analyzed the length of acceleration and deceleration lane from the viewpoint of safety, and established relative calculation models. In addition, Abishai (1987) and El-Basha (2007) analyzed the vehicles' running characteristics on acceleration and deceleration lanes as well.

The literatures above are mostly about freeway. Because of the difference of traffic flow running characteristics between freeway and urban expressway, the results above can't be applied to urban expressway directly. In this paper, the vehicles'

running characteristics on urban expressway will be analyzed, and the length of acceleration and deceleration lane will be studied. It will lay a foundation for designing acceleration and deceleration lane, and will provide a basis for determining the minimum interchange spacing on urban expressway as well.

Position of Acceleration and Deceleration Lane

Acceleration and deceleration lane are passageways for vehicles driving in and departing from an urban expressway. It lies not only in the end of an interchange ramp but in the place where the road connect with the urban expressway.

Length of Acceleration Lane

Acceleration lane is an important part of an interchange, its functions mainly show in two aspects: firstly, it provides a location where the merging vehicles can accelerate smoothly from the ramp velocity to the mainline lane velocity; secondly, it also provides a location where the ramp vehicles can wait for an opportunity to merge into the mainline lane when driving. In addition, there is a width transition section on the end of an acceleration lane. Based on the analyses above, the acceleration lane can be divided into three sections, namely acceleration section, waiting merging section and width transition section, as shown in Figure 1.

Let L_a be acceleration lane length, L_1 be acceleration section length, L_2 be waiting merging section length and L_3 be width transition section length. Then, $L_a=L_1+L_2+L_3$.

Acceleration Section Length L_1. Acceleration section length is the minimum length that the ramp vehicles need to accelerate from the ramp velocity to the mainline lane

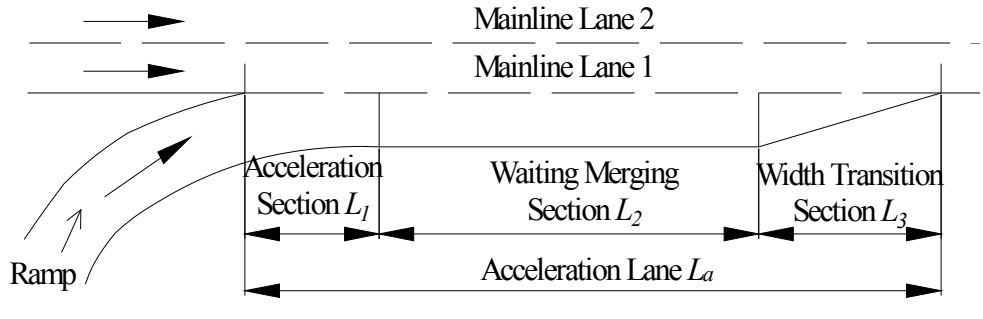


Figure 1. Sketch-map of acceleration lane

velocity. Let v_1 be mainline lane velocity(km/h), v_2 be merging velocity(km/h) and v_r be ramp velocity(km/h). Then, applying dynamics theory, $L_1 = (v_2^2 - v_r^2)/(25.92a)$,

where a = acceleration(m/s^2), generally, let a = 1.2m/s^2. According to investigation, the merging vehicles' velocity is smaller than the mainline lane vehicles' velocity, by surveying the velocity difference on-the-spot on freeway in Spain, Holand, Germany and Danmark, it is 20km/h (Zhao 2002). Let mainline lane vehicles' velocity be 60, 80, and 100km/h, taking the same reduction factor, the merging vehicles' velocity on urban expressway is v_2=v_1-10, v_2=v_1-13, and v_2=v_1-17 respectively.

Waiting Merging Section Length L_2. At the spot where the ramp vehicles may merge into the mainline lane in the earliest time, namely the end of the acceleration section and the beginning of the waiting merging section, let x=0. At the end of the waiting merging section, let x=L_2. Let $p(l)$ be the successfully merging probability of the ramp vehicles from x=0 to x=l($0<l\le L_2$), obviously when x<0, $p(x)$=0. In the following analysis, the relationship between the merging probability $p(l)$ and the driving distance l will be studied.

To obtain $p(l)$, it is needed to know the probability of the acceptance gap appearance in the expressway mainline lane 1 from x=0 to x=l($0<l\le L_2$) where the acceptance gap begin to appear. Because on the spot where x=l, the probability of the acceptance gap appearance is stochastic, it follows a certain distribution, let its distribution function be

$$F(t) = P\{h \ge t\} \tag{1}$$

Let $p(l+\Delta l)$ be the successfully merging probability of the ramp vehicles on the spot where $x = l + \Delta l$ ($\Delta l \to 0$), then,

$$p(l+\Delta l) = p(l) + [1 - p(l)]\Delta t F(t_c(l)) \tag{2}$$

where Δt = ramp vehicles' driving time in distance Δl, because the ramp vehicles drive in velocity v_2 when waiting for an acceptance gap, $\Delta t = \Delta l / v_2$.

Because Δl is very short, the values of the acceptance gap at the spot where $x = l + \Delta l$ is equal to the values at the spot where x=l, namely $t_c(l+\Delta l) = t_c(l)$. Then, the Equation 2 can be transformed into Equation 3,

$$\frac{p(l+\Delta l) - p(l)}{\Delta l} = -p(l)\frac{F(t_c(l))}{v_2} + \frac{F(t_c(l))}{v_2} \tag{3}$$

Let $Q(l) = F(t_c(l))/v_2$, then the Equation 3 can be transformed into Equation 4,

$$\frac{p(l+\Delta l) - p(l)}{\Delta l} = -p(l)Q(l) + Q(l) \tag{4}$$

Let $\Delta l \to 0$, taking limit on both sides of the Equation 4, the following differential

equation can be got:

$$p'(l) = -Q(l)\,p(l) + Q(l) \tag{5}$$

Let $t_c(l)$ be mean value of the ramp vehicles' critical gap, $Q = \overline{Q}(l)$ be mean value of the distribution function $Q(l)$, then the general solution of the Equation 5 is as follows:

$$p(l) = e^{-Ql}(e^{Ql} + c) \tag{6}$$

Because Equation 6 is a continuity equation, moreover, when l<0, $p(l)$=0, take left limit on both sides of it for l=0, then c= -1, so the Equation 7 can be got:

$$p(l) = 1 - e^{-Ql} \tag{7}$$

Let $l=L_2$, then, the Equation 8 can be got:

$$p(L_2) = 1 - e^{-QL_2} \tag{8}$$

The $p(L_2)$ is namely the probability that the acceptance gap isn't appear from x=0 until at the end of the waiting merging section. It is also the probability that the ramp vehicles can successfully merge into the mainline lane until at the end of the waiting merging section. Let $p(L_2)$ be greater than or equal to the probability value α (to make the ramp vehicles merge into mainline lane successfully, α should be nearly close to 1), namely:

$$p(L_2) = 1 - e^{-QL_2} \geq \alpha \tag{9}$$

Then, the Equation 10 can be got:

$$L_2 \geq -\ln(1-\alpha)/Q \tag{10}$$

Let
$$L_2 = -\ln(1-\alpha)/Q \tag{11}$$

L_2 is namely the waiting merging section length.

Width Transition Section Length L_3. Width transition section is designed to make the vehicles smoothly change a lane from the acceleration to mainline lane. The length of it can be calculated by equation $L_3 = v_1 t/3.6$, where t = traverse time in which the vehicles can smoothly change a lane from the acceleration to mainline lane. Base on related research results and practical applications, generally, t=3~4s, in the paper let t=3.5s. Because the vehicles have arrived at the end of the acceleration lane, let its velocity be equal to the mainline lane velocity v_1.

Length of Deceleration Lane

When analyzing vehicles' decelerating process on deceleration lane, the twice deceleration theory put forward by AASHTO(American Association of State Highway and Transportation Officials) was generally adopted. According to the theory, the decelerating process was divided into two stages, the first stage is by engine, and the second is by brake. In addition, on the beginning of the acceleration lane, there is a width transition section. Based on the analysis above, the length of deceleration lane can be divided into three sections, i.e. width transition section, the first and second deceleration section, as shown in Figure 2.

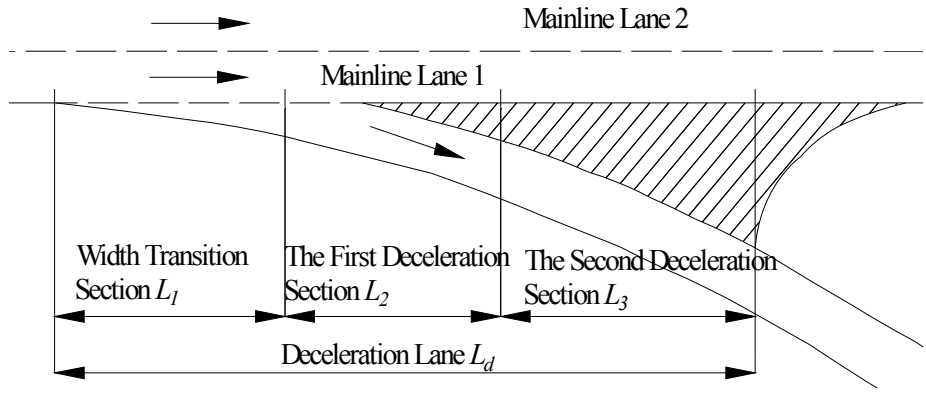


Figure 2. Sketch-map of deceleration lane

Let L_d be deceleration lane length, L_1 be width transition section length, L_2 and L_3 be the first and second deceleration section length, respectively. Then, $L_d=L_1+L_2+L_3$.

Width Transition Section Length L_1. The calculation method of width transition section length on deceleration lane is similar to it on acceleration lane. So, $L_1 = v_1t/3.6$, where v_1=mainline lane velocity(km/h); and t= traverse time, let t=3.5s.

The First Deceleration Section Length L_2. About the deceleration section length, Li (2006) calculated it by equation $D_3 = (v_1^2 - v_2^2)/[2g(f+G)]$, where v_1 and v_2 are mainline lane and ramp velocity(km/h), respectively. He (2005) calculated it by equation $L' = L_1 + L_2 = \frac{v_1 t}{3.6} - \frac{1}{2}a_1 t^2 + \frac{1}{25.92a_2}|(v_1 - 3.6a_1 t)^2 - v_2^2)|$, where a_1 and $a_2$are engine and brake deceleration(m/s^2), respectively. Based on the vehicles' performance and decelerating process in China, in the paper, the deceleration section length will be calculated as follows. Let a_1 be the first deceleration, namely the

engine deceleration(m/s^2), v_4 be velocity after this deceleration(km/h). The engine deceleration is carry out by reducing accelerograph, with the felling of comfort as the premise. According to the Automobile Drivers Manual in China, the comfort deceleration range is from -1.0 to -1.5m/s^2, let a_1=-1.2 m/s^2. Generally speaking, the velocity difference between two adjacent gears is about 15km/h, let $v_4 = v_1$-15. Then, $L_2 = (v_4^2 - v_1^2)/(25.92a_1)$.

The Second Deceleration Section Length L_3. By engine deceleration, if the vehicles' speed is still high, the drivers may carry out the second deceleration, namely the brake deceleration, it is also with the felling of comfort as the premise. Let v_r be ramp velocity(km/h), a_2 be brake deceleration(m/s^2). According to the Automobile Drivers Manual in China, in this deceleration process, the deceleration range is from -1.5 to -2.0 m/s^2, let a_2=-1.8 m/s^2. Then, $L_3 = (v_r^2 - v_4^2)/(25.92a_2)$.

Length Values Calculation

By analyzing above, to calculate the length of waiting merging section, it is necessary to know the time headway distribution of the mainline lane vehicles. The related research results are as follows (Zhang 2005):

(1) When the mainline lane traffic volume is less than 500 pcu/h/ln, the time headway is subjected to shifted negative exponential distribution. The distribution function is as follows:

$$P(h \geq t) = e^{-\lambda'(t-t_m)}, \quad \text{for} \quad t \geq t_m$$

where λ' = parameter of shifted negative exponential distribution, $\lambda' = 1/(\bar{t} - t_m)$; $\bar{t}$ = average time headway(s); and t_m = minimum time headway(s).

(2) When the mainline lane traffic volume is in the range of 500~1000 pcu/h/ln, the time headway is subjected to second-order Erlang distribution. The distribution function is as follows:

$$P(h \geq t) = \sum_{i=0}^{1} (2\mu t)^i \frac{e^{-2\mu t}}{i!}$$

where μ = vehicles' average arrival rate(pcu/s)

(3) When the mainline lane traffic volume is in the range of 1000~1500 pcu/h/ln, the time headway is subjected to M3 distribution. The distribution function is as follows:

$$p(h \geq t) = \begin{cases} \rho e^{-\lambda(t-t_m)}, & \text{for } t \geq t_m \\ 0, & \text{for } t < t_m \end{cases}$$

where t_m= minimum time headway(s); and ρ = traffic flow proportion with its time headway greater than or equal to t_m.

$$\rho = \begin{cases} e^{-0.55(q-0.025)}, & \text{for } q \geq 0.025\, pcu/s \\ 1, & \text{for } q < 0.025\, pcu/s \end{cases}$$

where λ = scale parameter, $\lambda = \rho q/(1-t_m q)$; and q = flow rate(pcu/s).

On expressway weaving section and the joint where the ramp connects with the mainline lane, the traffic flow running conditions should be designed as secondary level of service (Wang 2000). Under this level of service, if the designing velocity of the road section is 60, 80 and 100km/h, the traffic volume of the road section should be in the range of 220～750, 500～1150 and 750～1550pcu/h/ln respectively (Zhang 2005). To make the ramp vehicles successfully merge into mainline lane, the traffic volume under the secondary level of service should take the upper part of the range. Corresponding to the mainline lane velocity 60, 80 and 100km/h, let the traffic volume under the secondary level of service be 700, 1100 and 1400pcu/h/ln respectively. Then, the mainline lane vehicles' time headway is subjected to the second-order Erlang, M3 and M3 distribution respectively.

By investigation, the acceptance gap on expressway merging section is in the range of 3~4s, to the car, let t_c=3s (Wang 2005). Let t_m=1s (Zhang 2005), and $\alpha = 97\%$. Applying the calculation models above and taking the mainline lane and ramp vehicles' velocity combination into consideration, the length values of the acceleration and deceleration lane under different designing velocities are shown in Table 1, in which the acceleration and deceleration lane are all single-lane.

Table 1. The length of Acceleration and Deceleration Lane (m).

Mainline Lane Velocity(km/h)	Length of Acceleration Lane					Length of Deceleration Lane				
	Ramp Velocity(km/h)					Ramp Velocity(km/h)				
	80	60	50	40	30	80	60	50	40	30
100	393	483	518			204	264	288		
80		269	304	333			161	185	204	
60				159	181				118	133

Conclusions

Based on the vehicles' running characteristics on acceleration lane, the length of it was divided into three sections, namely acceleration section, waiting merging section and width transition section. Applying dynamics theory, the models calculating the length of acceleration section and width transition section were got. By combining probability theory and differential method and taking main lane traffic volume into consideration, the model calculating the length of waiting merging section was established as well.

Considering vehicles' decelerating process on deceleration lane, based on the twice deceleration theory put forward by AASHTO, the length of deceleration lane was

divided into three sections, i.e. width transition section, the first and the second deceleration section. Applying dynamics theory, the model calculating the length of width transition section was got. Applying dynamics theory and considering drivers' driving comfort, the models calculating the length of the first and second deceleration section were established as well.

Taking mainline lane vehicles' time headway distribution and mainline lane and ramp vehicles' velocity combination into consideration, the length values of acceleration and deceleration lane were calculated. The results indicate that the length of the acceleration and deceleration lane increases with the increase of mainline lane vehicles' velocity and decreases with the increase of ramp vehicles' velocity .

When the acceleration and deceleration lane are two-lane, the traffic flow running conditions are more complex, mainly manifesting in vehicles' lane selection and mutual interference. About the length of acceleration and deceleration lane with tow-lane, according to the current criterion, it can be 1.5 times of the lane length with single-lane.

References

Abishai, P., and Moshe, L. (1987). "Comments on flow characteristics on acceleration lanes." *Transportation Research, Part A: General*, 21(1), 39-46.

Alfredo, G. G. and Mario, R. R. (2006). "Deceleration lane length influence in operation and safety." *Carreteras*, 4(145), 6-16.

Bared, J., Giering, G. L., and Warren, D. L. (1999). "Safety evaluation of acceleration and deceleration lane lengths." *ITE Journal(Institute of Transportation Engineers)*, 69(5), 50-54.

Chen, T. Z., and Wu, J. Y. (2003). "Discussion on length design of deceleration lane at interchanges." *Central South Highway Engineering*, 28(3), 56-58.

El-Basha, R. S., Hassan, Y., and Sayed, T. A. (2007). "Modeling freeway diverging behavior on deceleration lanes." *Transportation Research Record 2012*, Transportation Research Board, Washington D. C., 30-37.

He, X. G., Liang, H., Wang, J. P., and Xiang, J. J. (2005). "A study of determination of shift lane length for interchanges." *Journal of Wuhan University of Technology(Transportation Science & Engineering)*, 29(3), 370-373.

Li, S., and Zhang, Y. (2000). "An application study of merging theory on designing acceleration lane for expressway." *China Journal of Highway and Transport*, 13(2), 110-113.

Shao, C. Q., and Rong, J. (2008). "Research on model of computing acceleration lane length and analysis of its influence factors." *Journal of Beijing University of Technology*, 34(1), 72-75.

Shi, X. F., Li, W. Q., Li T. Z., and Chang, Y. L. (2000). "Method of designing acceleration lane length in freeway interchange." *Journal of Henan University(Natural Science)*, 30(4), 17-20.

Transportation Research Board 85th Annual Meeting. (2006). *A Supplement to Advance Guide Sign Placement Guidelines in MUTCD*, Li, Washington.

Wang, L. (2005). Research on Capacity of Freeway On-Ramp Junction. M.S. Dissertation, Southeast University, China.

Wang, W. and Guo, X. C. (2000). Traffic Engineering. Southeast University Press, China.

Zhang, Y. P. (2005). Research on traffic characteristics and capacity on urban expressway basic section. Ph.D. Dissertation, Harbin Institute of Technology, China.

Zhao, C. (2002). Research on Traffic Simulation and Evaluation of Merging-Area on Freeway. M.S. Dissertation, Southeast University, China.

Zhao, C., Deng, W., Zhou, R. G., and Wang, W. (2004). "The length design method for acceleration lane taking into account of traffic." *Journal of Highway and Transportation Research and Development*, 21(8), 103-107.

Zhi, Y. F., Zhang, J., and Shi, Z. K. (2009). "Research on design of expressway acceleration lane length and merging model of vehicle." *China Journal of Highway and Transport*, 22(2), 93-97.

Self-motion Speed Perception of Visual Information in Driving: Modeling and Validation

Hirofumi Yotsutsuji[1] and Hideyuki Kita[2]

[1] Doctoral Student, [2] Professor, Graduate School of Engineering, Kobe University, 1-1 Rokkodai, Nada-ku, Kobe, Japan, Email: yotutuzi0932@miracle.ocn.ne.jp, kita@crystal.kobe-u.ac.jp

Abstract

When considering the engineered traffic safety measures against speeding accidents, it is important to know the speed perception structure of each driver. However, how modeling of the speed perception structure can be achieved in theory has not necessarily been carried out in traffic engineering research up to now. If a speed perception structure model can be theoretically derived, a speed choice behavior model incorporating such a perception structure can be expected to be one of the powerful engineered analysis tools for analyzing the effect of traffic safety measures taken against speeding accidents caused by a driver's misperception of speed. The purpose of this paper is to suggest a theoretical model for driver perception of vehicle speed, and to verify the model by conducting in-house experiments with a driving-simulator. The model is built on the basis of knowledge of visual perception psychology research. In the model, a mathematical relationship between the respective time-rate-of-changes of the actual and perceived speeds can be derived, obeying a power law with respect to "sensitivity", which is defined as a new indicator in this paper. Then, the model shows that the sensitivity produces "over-perception of speed" in theory. Through statistical tests from the results of the in-house experiments, this paper shows that the model is statistically valid.
Key Words: self-motion speed perception; visual information; optic flow; visible distance perception; driving simulator; in-house experiment; model validation

Introduction

Driver misperception of one's own vehicle speed is a risk factor in causing speeding accidents. For example, in the area of traffic psychology research it is well-known that, in decelerating a vehicle without paying attention to the speedometer at an exit-

ramp on an expressway, after driving at a constant speed for a long time, the actual vehicle speed may be observed to be higher than the driver had expected. Then, although the driver judges the current speed as safe at the ramp section, a misperception of the speed exists. The discrepancy between the actual vehicle speed and the perceived speed may lead to speeding accidents, such as a collision with the road-side wall, even though the driver has no awareness of such a discrepancy, or even though the driver has made a rational judgment of speed choice, and there have been no mistakes in vehicle operation.

Considering the engineered traffic safety devices deployed on roads, serving to reduce speeding accidents caused by driver misperception of speed, such as distance-markers or delineators, in order to assess the effectiveness of such devices, an analytical framework incorporating both the speed choice behavior model and the speed perception structure model is required, because not only the behavioral response but also the perceptual response toward the devices must be analyzed. In the construction of such an analytical framework, it will be beneficial to incorporate a speed perception structure model into a driving behavior model based on the rationality of speed choice. It is believed that the speed choice behavior of a rational driver can be modeled on an assumption of the maximization of speed utility. On the other hand, although it is well-known that an empirical rule exists behind a driver's perception of speed, how modeling of the speed perception structure can be achieved in theory has not necessarily been carried out in the past in the area of traffic engineering research. This paper focuses on the knowledge of visual perception psychology research on speed perception, and constructs a theoretical model on the basis of a theory of self-motion speed perception, based on visual information called "optic flow".

The purpose of this paper is to suggest a theoretical model for driver perception of vehicle speed, and to verify the model by conducting in-house experiments with a driving-simulator. As will be described later in this paper, the speed perception structure model constructed on the basis of optic flow theory is formed so as to be easily incorporated into the speed choice behavior model. The model is derived from the geometric relationship between the actual distance and the perceived distance from an optical stimulus on the road surface that the driver can perceive in a frontal view while driving. The model is expressed by the mathematical relationship between the respective time-rate-of-changes in the oncoming speed and the perceived speed of the stimulus. As a result, the speed perception structure model turns out to be a power law with respect to "sensitivity", if a proportion of the time variation of the Weber fraction of the perceived speed to the time variation of the Weber fraction of the actual distance is defined as the "sensitivity". "Weber fraction" is a technical term in psychophysics research, as will be explained later. The theoretical model of the speed perception structure shows that the sensitivity produces "over-perception of speed" in theory, and also shows that an estimate of the sensitivity enables the visible distance perception structure of the driver to be verified. Finally, through several statistical tests from the results of two experiments,

i.e., the constant-speed-adaptation experiment and the speed-ratio-evaluation experiment, the validity of the model will be statistically checked.

Outlines of Modeling

Considering the engineered analytical framework of driving behavior, Kita (2000) proposed a modeling framework based on utility approach, and analyzed merging behavior at the entrance ramp on an expressway. Kita constructed a discrete choice model of driving behaviors, such as merging or passing, at the ramp section, on the assumption that the driver has an instantaneously formed utility in cases of choosing such behaviors. In this paper, it is assumed that, when choosing driving behaviors, such as accelerating or decelerating, the choice of a behavior depends on the choice of the subjectively perceived speed, v_S, and the driver chooses a speed, v_S^*, satisfying the maximization of the instantaneously formed utility, $U(v_S)$. Then, the speed choice model of the rational driver can be expressed as follows:

$$v_S^* = \arg\max U(v_S) \tag{1}$$

On the other hand, for driver perception, in general, it is believed that the sensory information that a driver uses in driving a vehicle is almost all visual (Sivak 1996). In this paper, for the speed perception of a driver, it is assumed that the sensory information that the driver uses to perceive one's own vehicle speed is mainly visual. Then we define the speed perception structure, φ, against an actual vehicle speed, v_O, satisfying the following equation:

$$v_S = \varphi(v_O, E, \varsigma) \tag{2}$$

where E denotes the factor affecting the sensory information other than the visual ones, such the as auditory or vibratory information from the vehicle body, and ς denotes the factor affecting the driver's attributes or driving skills.

In this paper, the influence of both E and ς on φ is separate from the influence of visual information. Marr (1982) introduced the idea that, in the modeling of human visual perception, a distinction between the detection and the cognition of visual information is needed. According to Marr, the detection of visual information represents a purely geometric computational processing on the retina at the back of the human eyeball. In this paper, on the basis of the visual information which the optic flow in driver's view can provide to the driver, the structure model of φ can be mathematically derived from the purely geometric relationship between v_S and v_O.

It may be easy to imagine that, when a driver drives a vehicle forward at a constant high speed for a long time on a straight section of an uncrowded expressway, the driver can perceive self-motion though the phenomenon by which the road scenery, which would have objectively remained stationary against the moving vehicle, seems to have flowed along in an inverse direction against the moving direction of the vehicle. According to Gibson (1966), a driver can perceive a visible distance of depth

to an optical stimulus, on the basis of an array of optical stimuli that has been already laid out on the road surface, such as the gradient of the density of road texture. Additionally, the driver can also perceive a self-motion speed, on the basis of the velocity field of the stimuli, i.e. the optic flow field, expanding all-round from a focal-point-of-view in the driver's field-of-view, along the inverse direction relative to the moving direction of the vehicle. In terms of an optic angle to an optical stimulus, this optic angle, in a precise sense, denotes the angle that is formed between a line to the focal-point-of-view and a line connecting the stimulus with the counter-image projected on the retina of the driver's eye. Lee (1980) mathematically translated the optic flow as the temporal differentiation of this optic angle. And through modeling the optic flow, a model of visible distance perception structure was derived, such that a visible distance from the focal-point-of-view to an optical stimulus increases in inverse proportion to the actual distance to the stimulus on the road surface. Lee's modeling framework is easy to handle due to its simplifications. Inspired by Lee's research, a number of models that followed his framework have been suggested. However, to our knowledge, these models have yet to focus on the discrepancy between the actual vehicle speed and the perceived speed.

Considering the speed of an optical stimulus approaching along an inverse direction, against the moving direction of the vehicle, it is believed that the driver can perceive a self-motion speed, v_S, on the basis of this oncoming speed, v_O, indicated in the equation (2). It is also believed that the driver can perceive self-motion on the basis of this self-motion speed. This is called "visually induced self-motion perception" or "vection" (Lappe et al., 1999). In this paper, the structure model of the self-motion speed perception can be mathematically derived through the modeling of the visible distance perception structure, using examples from Lee's framework.

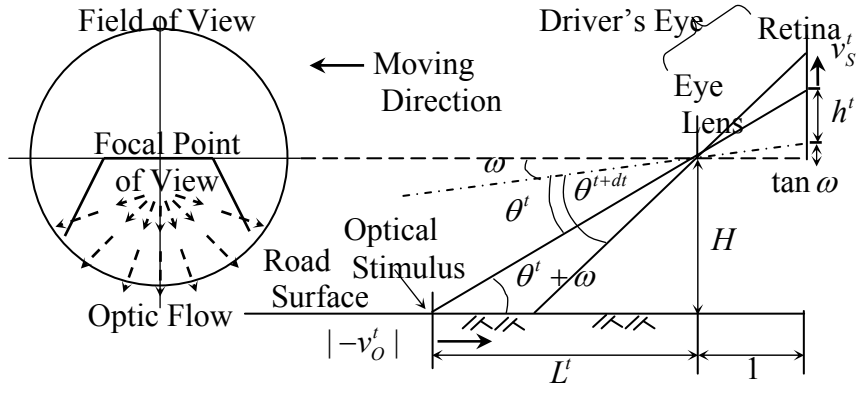
Modeling of Speed Perception Structure

Modeling Framework. The framework for modeling the speed perception structure is illustrated in Figure 1. Suppose that a driver is operating a vehicle on a straight road, maintaining a constant eye-level of *H*, and the driver is watching a focal-point-of-view at a constant angle of ω to the moving direction of the vehicle. The left side of Figure 1 shows the optic flow expanding from a focal-point-of-view in the driver's field-of-view. On the other hand, the right side of Figure 1 shows a lateral view illustrating the geometric relationship between an optical stimulus on the road surface and the counter-image projected on the retina of the driver's eye. Despite any longitudinal gradient of the road surface, the moving direction of the vehicle is always parallel to the road surface. The distance between the eye-lens and the retina of the driver's eye is standardized as 1. Assuming that the driver is operating without paying attention to the speedometer, the actual vehicle speed, v_O^t, at the time-period t is unknown for this driver. Based on an optical stimulus located at the actual distance of L^t, the driver perceives the oncoming speed, $|-v_O^t|$, along the inverse direction to the moving direction, while the counter-image projected on the retina at the visible distance of h^t from the focal-point-of-view stimulates the perceived

speed, v_S^t. Both of the stimuli can make the optic angle of θ^t along the direction of the focal-point-of-view. Then the temporal change of $\theta^t + \omega$ can form an instantaneous velocity field on the retina, namely the optic flow field. The driver can perceive a speed of v_S^t on the basis of the temporal change of $\theta^t + \omega$.

Visible Ddistance Perception Structure Model. Let's consider the geometric relationship among the visible distance of h^t, the actual distance of L^t, and the optic angle of $\theta^t + \omega$, in Figure 1. Now, for convenience of the latter perspective, L^t has a distortion of $\widetilde{L}^t$ which is defined as follows:

$$\widetilde{L}^t \equiv \frac{(L^t)^{\delta-1}}{\delta - 1} \tag{3}$$



- Legends -

L^t	: Actual Distance to an Optical Stimulus
h^t	: Visible Distance to an Optical Stimulus
$\lvert -v_O^t \rvert$	: Oncoming Speed of the Optical Stimulus
v_O^t	: Actual Vehicle Speed along the Moving Direction
v_S^t	: Perceived Speed of Optic Flow
θ^t	: Optic Angle Formed by the Optic Stimulus
H	: Height of Driver's Eye (Constant Level)
ω	: Optic Angle Formed by the Focal Point of View (Constant Level)

Figure 1. The framework for modeling the speed perception structure.

Then, the derivative of $\theta^t + \omega$ with respect to t can be expressed by the derivative of each of h^t and $\widetilde{L}^t$ with respect to t, respectively, as follows:

$$\frac{d(\theta^t + \omega)}{dt} = \frac{d}{dt}\tan^{-1}(h^t + \tan\omega) = \frac{1}{1+(h^t+\tan\omega)^2}\frac{dh^t}{dt} \tag{4}$$

$$\frac{d(\theta^t + \omega)}{dt} = \frac{d}{dt}\tan^{-1}\frac{H}{\widetilde{L}^t} = \frac{-H}{(L^t)^\delta\{1+(H/\widetilde{L}^t)^2\}}\frac{dL^t}{dt} \tag{5}$$

Consequently, the following equations are derived from both equations of (4) and (5).

$$dh^t = -\frac{H}{(L^t)^\delta}dL^t \tag{6}$$

$$h^t + \tan\omega = \frac{H}{\widetilde{L}^t} \tag{7}$$

Both equations of (6) and (7) illustrate the models of the visible distance perception structure of a driver. Now, let's consider how the Eq. (7) can be modeled, depending on the degree of the distortions of L^t.

First, the case that L^t has no distortion perfectly represents $\delta = 2$ in the equation (3), and $\widetilde{L}^t = L^t$ in the equation (7). Now, in the case that the minimal-change, dL^t, of L^t can be differentiable, the fraction of dL^t / L^t is called "Weber fraction of distance", and the empirical rule that the Weber fraction may be constant, namely $dL^t / L^t = k$ where k is constant, is called "Weber's law". As well, under the acceptance of Weber's law, the empirical rule that dh^t may be proportional to dL^t / L^t is called "Fechner's law". Fechner's law is a classical psychophysics rule from the 19th century. On the other hand, the empirical rule that $dL^t/(L^t)^\delta = k'$ where k' is constant is called "the near-miss-to-Weber's law" (McGill *et al*. 1968). This paper tentatively calls the fraction of $dL^t/(L^t)^\delta$ "the near-miss-to-Weber fraction". When the near-miss-to-Weber fraction of $dL^t/(L^t)^\delta$ is constant, because of $dL^t / L^t = k/(L^t)^{1-\delta}$, the Weber fraction depends on $(L^t)^{1-\delta}$. In a case of δ = 2 in the equation (6), on the assumption that $dL^t/(L^t)^2$ is constant, the minimal-change of the perceived distance, dh^t, is proportional to the near-miss-to-Weber fraction of distance, $dL^t/(L^t)^2$. Then, by equation (7), $h^t + \tan\omega$ is inversely proportional to L^t.

Secondly, the case when L^t has a distortion that perfectly represents δ = 1 in the equation (6), dh^t is proportional to the Weber fraction of distance, dL^t / L^t. Then, the functional relationship between $h^t + \tan\omega$ and L^t becomes a logarithmic decrement, because of $h^t + \tan\omega = H/\{\ln L^t + 1/(\delta - 1)\}$.

Accordingly, the functional relationship between $h^t + \tan\omega$ and L^t takes either an inverse proportion form if δ = 2 and the near-miss-to-Weber fraction of distance is constant, or a logarithmic decrement form if δ = 1 and the Weber fraction of distance is constant. If the degree of the distortion satisfies $1 \le \delta < 2$, then $L^t > \widetilde{L}^t$.

Table 1. The sign condition of Eq. (9) corresponding to the speed choice behaviors.

	$dv_S^{\tau+1} - dv_S^{\tau}$	$\lvert v_S^{\tau+1}\rvert - \lvert v_S^{\tau}\rvert$	$\frac{dv_S^{\tau+1}}{\lvert v_S^{\tau+1}\rvert} - \frac{dv_S^{\tau}}{\lvert v_S^{\tau}\rvert}$	$\frac{dv_O^{\tau+1}}{\lvert -v_O^{\tau+1}\rvert} - \frac{dv_O^{\tau}}{\lvert -v_O^{\tau}\rvert}$	$\delta\left(\frac{dL^{\tau+1}}{L^{\tau+1}} - \frac{dL^{\tau}}{L^{\tau}}\right)$
U. M.	≥ 0	> 0	> 0	$= 0$	> 0
U. A.	≥ 0	≥ 0	> 0	< 0	> 0
U. D.	≤ 0	≤ 0	≤ 0	> 0	≤ 0
	≥ 0	≥ 0	≥ 0	> 0	≥ 0

Annotation: U. M., U. A., and U. D. denote uniform motion, uniform accelerated motion, and uniform decelerated motion, respectively.

Whichever of the above is valid, the sensitivity of the speed perception structure model, as will be discussed later, acts as the indicator of validation.

Speed Choice and Speed Perception. Now, let a_S^t and a_O^t be defined as $a_S^t \equiv dv_S^t / dt \equiv d^2 h^t / dt^2$ and $a_O^t \equiv dv_O^t / dt \equiv d^2 L^t / dt^2$, respectively. The following equation is obtained by differentiating the equation (6) twice with respect to t.

$$\frac{\lvert a_S^t \rvert}{\lvert v_S^t \rvert} = \frac{\lvert -a_O^t \rvert}{\lvert -v_O^t \rvert} + \delta \frac{\lvert -v_O^t \rvert}{L^t} \tag{8}$$

Provided that discrete time is applied, the difference of the equation (8) for a given time-period, τ, is expressed as:

$$\frac{dv_O^{\tau+1}}{\lvert -v_O^{\tau+1}\rvert} - \frac{dv_O^{\tau}}{\lvert -v_O^{\tau}\rvert} = \frac{dv_S^{\tau+1}}{\lvert v_S^{\tau+1}\rvert} - \frac{dv_S^{\tau}}{\lvert v_S^{\tau}\rvert} - \delta\left(\frac{dL^{\tau+1}}{L^{\tau+1}} - \frac{dL^{\tau}}{L^{\tau}}\right) \tag{9}$$

Note that the time interval between τ and $\tau+1$ does not need to be coincident with dt.

Now, let's consider the relationship between speed choice behaviors and the Weber fraction of speed, employing the equation (9). Table 1 shows the sign condition for each term of the equation (9) corresponding to the speed choice behaviors, such as uniform motion, uniform accelerated motion, and uniform decelerated motion. In particular, the fourth column from the left side of Table 1, $dv_S^{\tau+1} / \lvert v_S^{\tau+1}\rvert - dv_S^{\tau} / \lvert v_S^{\tau}\rvert$, represents the relationship between the speed choice behaviors and how to perceive the optic flow expanding from the focal-point-of-view. In other words, the sign condition of the fourth column suggests that a driver who drives a vehicle at a constant speed or constant acceleration can perceive the optic flow on the rim of the field-of-view as flowing faster than the optic flow near the focal-point-of-view.

Speed Perception Structure Model. Now, define a new indicator, $\alpha_{\tau+1}$, satisfying the following equation.

$$\alpha_{\tau+1} \equiv \frac{dv_S^{\tau+1} / v_S^{\tau+1} - dv_S^{\tau} / v_S^{\tau}}{dL^{\tau+1} / L^{\tau+1} - dL^{\tau} / L^{\tau}} \tag{10}$$

The $\alpha_{\tau+1}$ defined by the equation (10) indicates a proportion of the variation of the Weber fractions of perceived speeds, to the variation of the Weber fractions of actual distances, at a time period of $\tau+1$. In this paper, $\alpha_{\tau+1}$ is called "sensitivity". Consequently, the following equation is derived from equation (9) by employing the sensitivity, $\alpha_{\tau+1}$. In the following equation, $c_{\tau+1}$ denotes a constant of integration.

$$\frac{|-v_O^{\tau+1}|}{|-v_O^{\tau}|} = c_{\tau+1}\left(\frac{|v_S^{\tau+1}|}{|v_S^{\tau}|}\right)^{1-\frac{\delta}{\alpha_{\tau+1}}} \tag{11}$$

The Eq. (11) illustrates a model of the speed perception structure of a driver. This model suggests that the functional relationship between the oncoming speed, v_O^{τ}, and the perceived speed, v_S^{τ}, of an optical stimulus can be represented by a power function with respect to the sensitivity, $\alpha_{\tau+1}$, in theory.

Now, let's consider the sign condition of $1-\delta/\alpha_{\tau+1}$. If $1-\delta/\alpha_{\tau+1}>1$ for any $c_{\tau+1}$, then the equation (11) illustrates that the actual vehicle speed is accelerating or decelerating more than the accelerated or decelerated speed that the driver perceived. On the other hand, if $0<1-\delta/\alpha_{\tau+1}<1$, then the actual vehicle speed is accelerating or decelerating less than the accelerated or decelerated speed that the driver perceived. This paper calls the former "under-perception" and the latter "over-perception". However, because of $\delta>0$, the speed perception structure produces not under-perception but over-perception of speed, in theory.

Distortion of Distance Perception and Sensitivity of Speed Perception. Because $1-\delta/\alpha_{\tau+1}$ in the equation (11) is nonnegative, the following condition is obtained.

$$\alpha_{\tau+1} \geq \delta \tag{12}$$

If the near-miss-to-Weber fraction of L^{τ} is constant, i.e., $\delta = 2$, then $\alpha_{\tau+1}$ meets the condition of $\alpha_{\tau+1} \geq 2$. If L^{τ} has a distortion with a degree of $1 \leq \delta < 2$, then $\alpha_{\tau+1} \geq 1$. In particular, if the Weber fraction of L^{τ} is constant, i.e., $\delta = 1$, then the inverse of the $\alpha_{\tau+1}$, $1/\alpha_{\tau+1}$, is zero, and the equation (11) has the power of 1.

Focal-point-of-view Direction and Sensitivity. Let's consider the relationship between the optic angle, ω, formed by the focal-point-of-view direction and the sensitivity, $\alpha_{\tau+1}$, of the speed perception structure, in a case in which the driver is driving while changing the optic angle of ω. For convenience, the sensitivity with respect to $\omega = 0$ describes $\alpha_{\tau+1}^{0}$, and also the sensitivity with respect to $\omega = \overline{\omega}$ for any $\overline{\omega}$ describes $\overline{\alpha}_{\tau+1}$. Then the relationship between $\alpha_{\tau+1}^{0}$ and $\overline{\alpha}_{\tau+1}$ can be represented as follows:

$$1-\frac{\delta}{\overline{\alpha}_{\tau+1}}=1-\frac{\delta}{\alpha^{0}_{\tau+1}}-\delta\frac{\tan\overline{\omega}}{H}\Delta_{\tau+1} \tag{13}$$

where,

$$\Delta_{\tau+1}\equiv\frac{dL^{\tau+1}-dL^{\tau}}{dv_S^{\tau+1}/v_S^{\tau+1}-dv_S^{\tau}/v_S^{\tau}} \tag{14}$$

Now, let's consider the relationship between the speed choice behaviors and the $\Delta_{\tau+1}$. For uniform motion, $\Delta_{\tau+1}$ meets the condition of $\Delta_{\tau+1} = 0$ because of $dL^{\tau+1} = dL^{\tau}$ so that $\alpha^{0}_{\tau+1}$ corresponds with the $\overline{\alpha}_{\tau+1}$. This result suggests that the driver's perception of one's own vehicle speed at a uniform motion does not depend on the focal-point-of-view direction, in theory. At uniformly accelerated motion, $\Delta_{\tau+1}$ meets the condition of $\Delta_{\tau+1} > 0$ because of $dL^{\tau+1} > dL^{\tau}$ so that the relationship between $\alpha^{0}_{\tau+1}$ and $\overline{\alpha}_{\tau+1}$ meets $1-\delta/\overline{\alpha}_{\tau+1} < 1-\delta/\alpha^{0}_{\tau+1}$. This result suggests that, if the driver, in driving at a uniformly accelerated motion, positions the focal-point-of-view direction to the optic angle of $\omega = \overline{\omega}$ for any $\overline{\omega}$, the speed perception of the driver is underestimated, as compared in a case of uniform motion with $\omega = 0$. That is, for example, when the actual vehicle speed doubles the speed perception of the driver does not become less than double, but the degree of the shortage becomes so high as to position the focal-point-of-view direction to the near-side of the field-of-view. Accordingly, when choosing the uniformly accelerated motion, it is important for the driver to look as far ahead as possible, from the viewpoint of traffic safety. Finally, for uniformly decelerated motion, the speed perception of the driver becomes complex, as can be seen from Table 1.

Model Validation though In-house Experiments

Environment of the Experiments. In order to verify whether the speed perception structure model of the Eq. (11) is applicable in practice, both a constant-speed-adaptation experiment and a speed-ratio-evaluation experiment were conducted, employing a driving simulator. For an on-road experiment using a vehicle, a long , straight road is needed in order to enable the driver to perceive the change of vehicle speed in a safe manner. In contrast, an in-house experiment using a driving simulator does not need such a safety constraint, and has the advantage of making it easy to reconstruct the same driving situation every time.

The driving simulation employs a Honda simulator, as shown in Figure 2, produced for driving safety training at the Suzuka Circuit Traffic Education Center in Japan. In the experiments, the driving situation involves a straight single-lane asphalt-paved road with a tree-lined vista, a width of 7 meters and no lane-markings, and without other vehicles around the driver's one. The speedometer equipped in this simulator is analog.

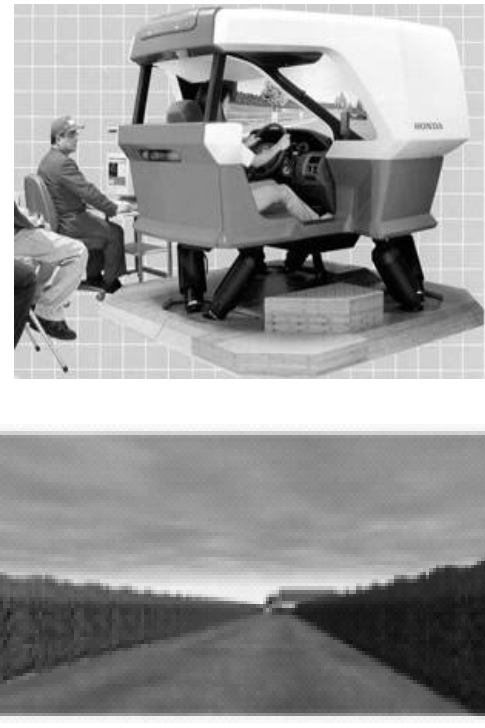


Figure 2. The driving simulator and the display monitor.

This simulator provides not only visual information to the driver while driving, but also auditory information through road noises and body-sensory information through seat vibrations. Meanwhile, the model of the Eq. (11) takes into account only visual information, despite the availability of other information. Accordingly, for the purpose of testing the validation of the Eq. (11), a driving environment under the auditory or body-sensory controlled conditions may be better, in the sense that various biases caused by other perceptual information except vision are eliminated. However, it is not sufficiently clear how these biases influence the driver's perception of speed. Therefore, as a first step toward the validation tests of the equation (11), the experiment is held in a driving environment exactly similar to the on-road situations, involving both auditory and body-sensory conditions.

Participants in the experiments. There are a total of 3 participants in both experiments: a non-expert driver who is a male in his 20s, called participant A, an expert driver who is a male in his 40s, called participant B, and a non-expert driver who is a male his in 40s, called participant C.

For the purpose of analyzing the relationship among speed-adaptation tendencies, driving skills, and speed preferences, speed utilities of all participants were estimated through a questionnaire survey after both experiments. In the questionnaire, the experimenter asked all participants the driving speeds equivalent to the utilities of situations in which the vehicle speeds change at 20 km/hr or 120 km/hr while driving.

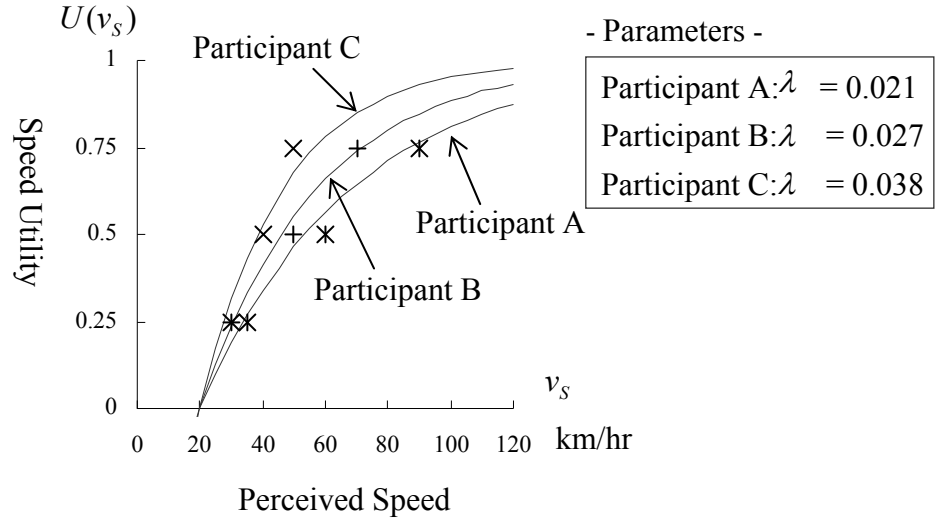


Figure 3. Speed utilities of all participants.

The speed utility functions of all participants were identified as shown in Figure 3, by employing the risk aversion utility function as follows:

$$U(v_S) = 1 - \exp[-\lambda(v_S - 20)] \tag{15}$$

Figure 3 shows that the perceived speeds at the same utility level were high in order of the participants A, B, and C. The utility function of the equation (15) assumes that speed preferences of all participants become so high as to drive at the high speed, because the environment of the experiments is simple.

Methods of the Experiments. In both experiments, the focal-point-of-view directions of all participants were set as either "a far-point" or "a near-point". The experimenter illuminates either point with a penlight on the screen, so that the participants perform while watching the point. In such driving situations, the influence of the focal-point-of-view directions on the speed perception structure is analyzed.

In both experiments, a participant operates the simulator vehicle, checking the speed on the speedometer, and after 20 seconds applies the brakes. The purpose of this pilot drive is to make the participant check the speed instructed by the experimenter and to recognize the optic flow field in the field-of-view at the instructed speed.

In the constant-speed-adaptation experiment, the experimenter instructs the participant to maintain speed at a baseline speed for 10 seconds, and hides the speedometer. The participant tries to maintain the speed. And then the experimenter monitors the change of actual speeds3 times with the passage of each 20 seconds.

In the speed-ratio-evaluation experiment, after the participant maintains a baseline speed for 10 seconds, the experimenter instructs the participant to gain speed with an instructed speed-ratio to the baseline speed, and hides the speedometer. The participant announces the moment that the participant subjectively evaluated the

speed with the instructed speed-ratio after the speed adjustment. Then the experimenter monitors the actual speed at the moment of the participant's announcement.

The Results of the Constant-speed-adaptation Experiment. The results are summarized in Figure 4. Figure 4 shows how the actual vehicle speed, v_O^1, was changing with the passage of time, after the experimenter instructed the participant to maintain speed at the baseline speed, v_S^0, and concealed the speedometer, under a condition in which the focal-point-of-view directions were set at either a far-point or a near-point. The baseline speed, v_S^0, was set at 20 km/hr intervals between 20 km/hr and 100 km/hr. The symbols of 0 and 1 for both v_S and v_O denote the ex-ante and the ex-post, respectively.

In the case of a far-point view in Figure 4, participant A, who was a non-expert driver, could not maintain the baseline speed of 80 km/hr, although he had the capability to maintain a baseline speed of 60 km/hr. Participant C, who was also a non-expert driver, tended to keep both baseline speeds of 60 km/hr and 80 km/hr close to 40 km/hr with the passage of time. Participant B, who was the expert driver, tended to maintain all baseline speeds from 20 km/hr to 80 km/hr. These results imply that speed-adaptation tendencies of the drivers with a far-point view may be related to their driving skills or speed preferences.

If the speed perception structure model is valid, the result of this experiment suggests that the temporal change of actual speeds to a baseline speed does not depend on whether the focal-point-of-view is far or near. Then the statistical test of significant differences between the far-point and near-point data groups for each baseline speed was implemented. After evaluating the correlation ratio,η^2, by a one-way analysis of variance, the null hypothesis H_0^1 that the far-point and near-point data groups for each baseline speed have no difference is tested by F-statistic. The results of the statistical tests indicated that the case of the middle and low levels of the baseline speeds was valid, but the case of the high level of the baseline speeds showed a significant difference. These results suggest that changing the focal-point-of-view directions suddenly, while driving at the high and constant speeds without watching the speedometer, in practice has the potential of a greater change of speed than expected. This point implies that it is important for a driver who is driving at a high speed to maintain the focal-point-of-view directions, from the viewpoint of traffic safety.

The Results of the Speed-ratio-evaluation Experiment. The results are summarized in Figure 5. Figure 5 shows the relationship between the perceived baseline speed, $\mathbf{v_s^0}$, and the adjusted speed, $\mathbf{v_s^1}$, at the moment when the participant evaluated the speed with the instructed ratio of $\mathbf{v_s^1/v_s^0}$. Figure 5 also shows the relationship between the actual vehicle speed, $\mathbf{v}_O{}^0$, against the perceived baseline speed, $\mathbf{v}_S{}^0$, and the monitored speed, $\mathbf{v}_O{}^1$, against the adjusted speed, $\mathbf{v_s^1}$.

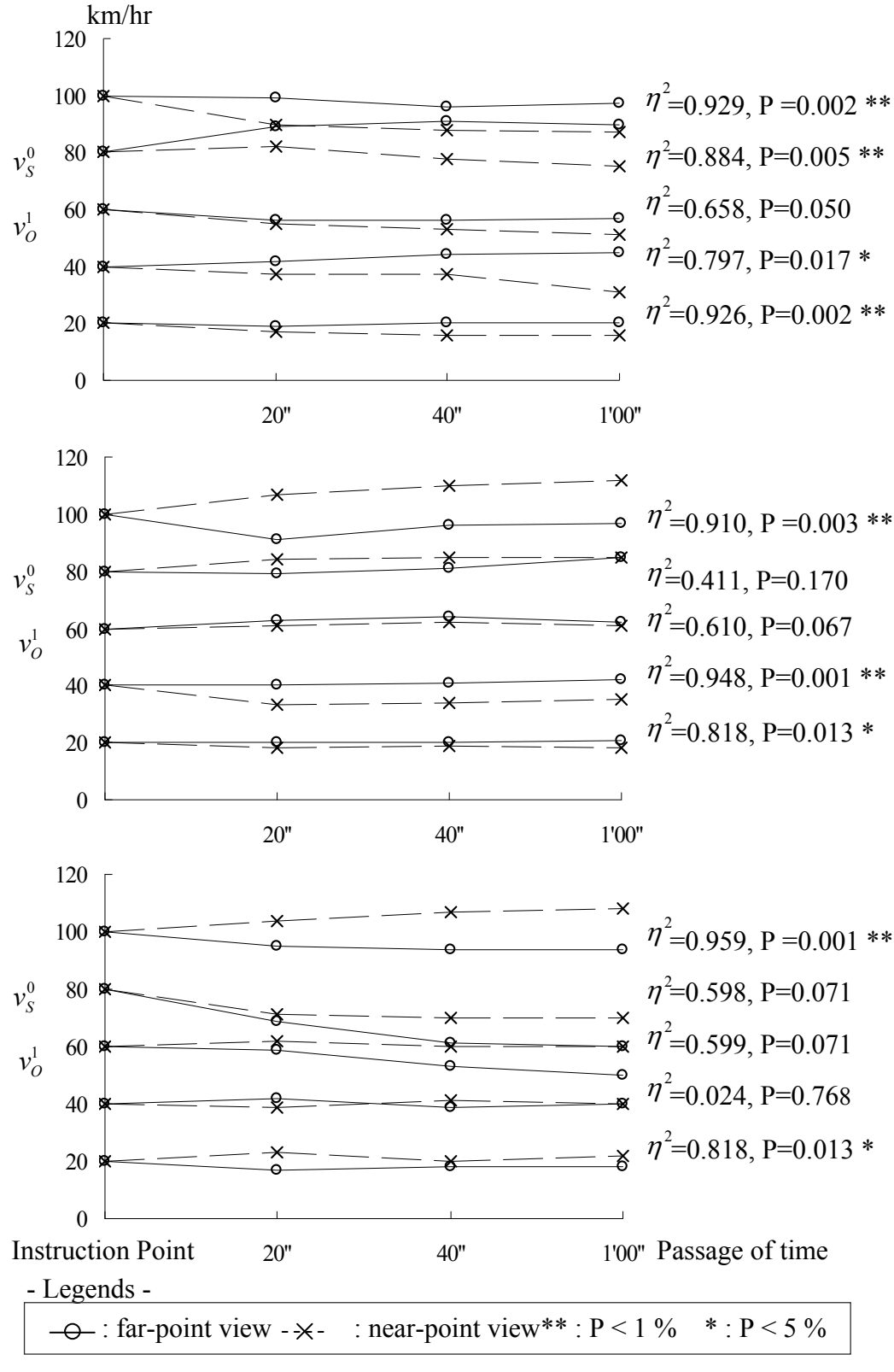


Figure 4. The results of the constant-speed-adaptation experiment.

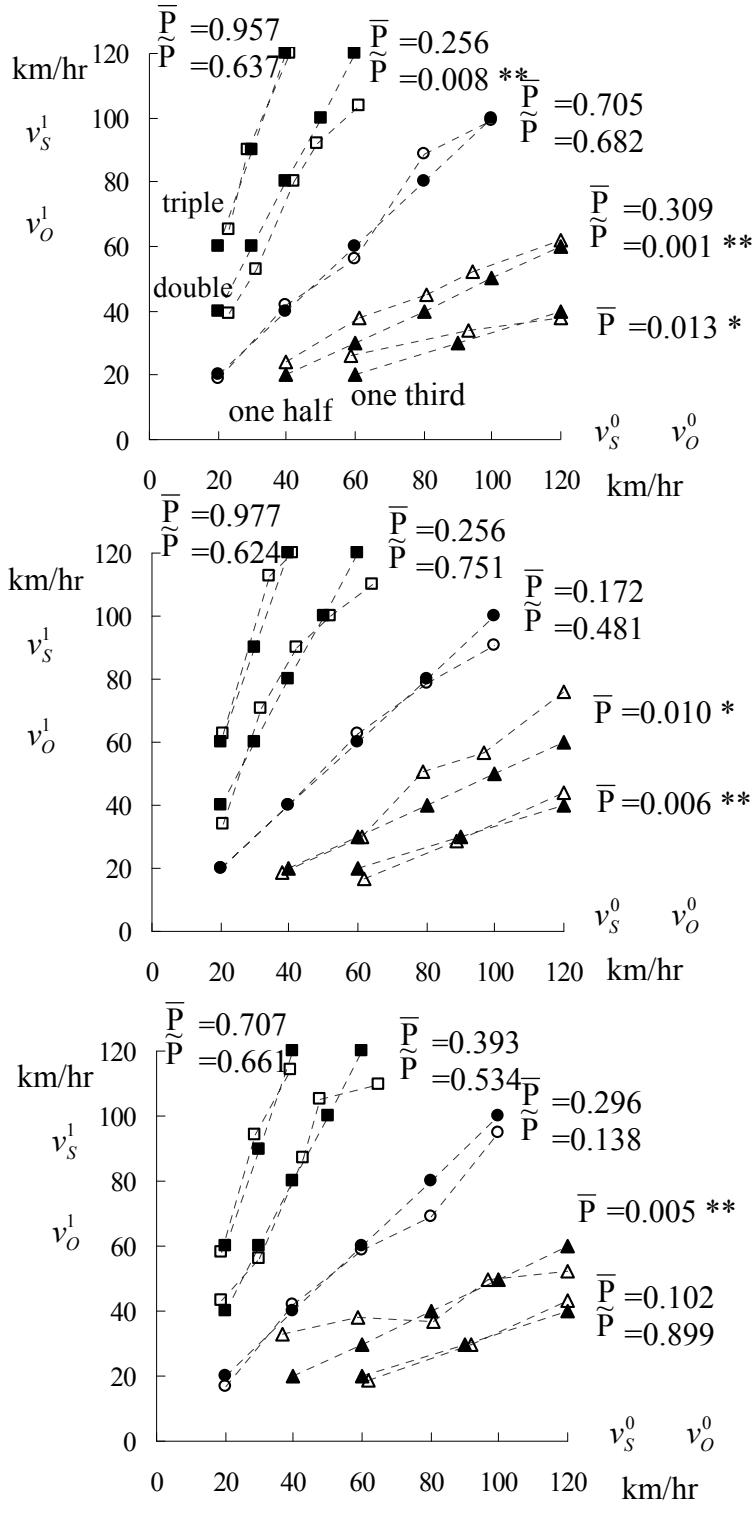


Figure 5. Results of the speed-ratio-adaptation experiment.

If the model of the Eq. (11) is valid, it is likely that the gradient of the regression line constructed by v_S^0 and v_S^1 is statistically different from the one constructed by v_O^0 and v_O^1. Therefore, in order to test this hypothesis, an analysis of covariance (ANCOVA) is needed. The first step of ANCOVA is the parallelism test, to verify whether the two regression lines are parallel. As a first step, the null hypothesis H_0^2 that each regression coefficient of the two groups of data is equal is tested by F-statistic. If H_0^2 is accepted, the second step of ANCOVA is the significant difference test on whether the two regression lines have a difference. The null hypothesis H_0^3 that each mean in the two groups of data has no difference is tested by F-statistic. The result of ANCOVA on the data for a given speed ratio is depicted in Figure 5, where $\overline{P}$ and $\widetilde{P}$ denote each p-value of the parallelism test and the significant difference test, respectively.

Figure 5 shows that the significant difference between the gradients of regression lines is more remarkable for a speed ratio equal to one half or one third, than to the double or triple. This result suggests that the discrepancy between the actual and perceived speeds may be more remarkable in decelerating driving than in accelerating, and the actual speed in the decelerating driving may not be as decelerated as the driver expected. These results may be interpreted as one of the reasons why speeding accidents on the exit-ramp of an expressway are not small in number.

Validation Test of The Speed Perception Structure Model. By using data from both experiments, whether the speed perception structures of all participants can be represented by a power functional form as proposed in the Eq. (11) or not was verified.

First, by implementing the non-linear regression analysis to the following equation, the null hypothesis H_0^4 of β_0=0 is tested:

$$\frac{v_O^1}{v_O^0} = \beta_0 + \beta_1 \left(\frac{v_S^1}{v_S^0} \right)^{\beta_2} \tag{16}$$

where the actual speeds of v_O^1 and v_O^0 used the mean value of all participants' data.

The results of two cases where the baseline speeds are 40 km/hr and 60 km/hr are shown. The t-values of $\hat{\beta}_0$ are t = 1.43, p < 0.39, in the case of 40 km/hr, and t =– 0.88, p<0.54, in the case of 60 km/hr, respectively. The results of the t-test show that H_0^4 is not rejected for both cases at the significant level of 1 %. Accordingly, the functional relationship illustrated in the Eq. (16) is valid.

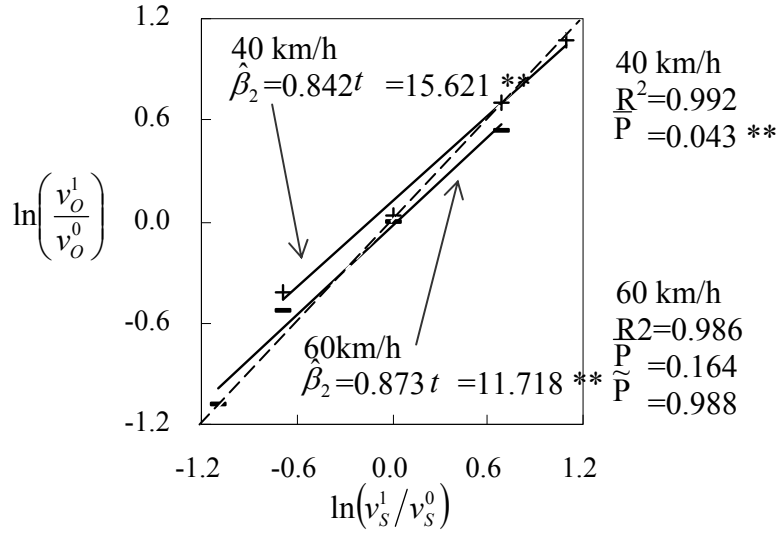


Figure 6. Result of the validation test of the speed perception structure model.

Next, after introducing the regression analysis setting β_0=0 in the equation (16), the parallelism between the gradient of the regression line and the 45-degree line is tested. The result of the test is depicted in Figure 6. The regression line of 40 km/hr crosses 45-degree line significantly, while the regression line of 60 km/hr fits in the 45-degree line statistically. Therefore, the result of this experiment, in particular, was that the speed of 60 km/hr is interpreted to be relatively safe in the sense that a driver can operate a vehicle as expected at this speed. As shown in Figure 6, the results of both of $\hat{\beta}_2$=0.842 in the case of 40 km/hr and $\hat{\beta}_2$=0.873 in the case of 60 km/hr can corroborate that the speed perception structure produces not under-perception but over-perception, because of $\hat{\beta}_2 < 1$.

Finally, let's try to evaluate the sensitivity, α, from the $\hat{\beta}_2$. In the case of setting at ω=0, the estimated value of α is represented by the equation of $\alpha = \delta/(1-\hat{\beta}_2)$. Then, α=6.329δ at 40 km/hr and α=7.874δ at 60 km/hr are obtained, so that $\alpha > 2$ for $1 \le \delta < 2$. This result suggests that the near-miss-to-Weber fraction of distance may be constant.

Conclusions

This paper suggests a theoretical model of speed perception structure, including a visible distance perception structure of a driver. In this model, the visible distance perception structure can be evaluated by employing the sensitivity of the speed perception structure model. This is the great advantage of the theoretical approach over the empirical approach, in modeling driver perception of distance and speed.
As has been shown in this paper, the knowledge obtained by the modeling of the speed perception structure can be organized, as follows:

- Theoretically modeling a speed perception structure based on the optic flow theory results in the empirical rule that driver perception of one's own vehicle

speed can be empirically approximated by a power law, and can also be theoretically reconfirmed.

- In cases of driving at a uniform motion, speed perception structure does not depend on the focal-point-of-view direction, in theory. In practice, in the middle and low levels of the perceived speeds, this point was valid. However, at a high level of perceived speeds, this point was not necessarily valid. Changing the focal-point-of-view directions suddenly while driving at a high speed may produce a greater change of speed than expected. On the other hand driving at a uniform accelerated motion, it is important to look at a far-point that is as far away as possible, from the viewpoint of traffic safety.
- Driver perception of one's own vehicle speed produces not under-perception but over-perception of speed, in theory. This point was corroborated by a model validation test in the experiments.
- In general, speeding accidents result from overlapping among the errors of speed perception, the irrationality of speed choice behavior, and the mistakes of speed control operation. In order to analyze the effects of the engineered traffic safety measures to prevent speeding accidents caused by multiple factors, it is believed that the approach of focusing on a certain level of the driving behavior process and analyzing the effects on the level will be beneficial. Especially, as long as the existence of driver misperception of one's own vehicle speed is premised, a driving-behavior analysis framework incorporating the speed perception structure model suggested in this paper will be beneficial.

References

Gibson, J. J. (1966). "The senses considered as perceptual systems", *Boston: Houghton Mifflin, MA,* 186-223

Kita, H. (2000). "Level-of-service measure of road traffic based on the driver's perception", *Transportation Research Circular E-C018 the Proceedings of the 4th International Symposium on Highway Capacity*, 53-62

Lappe, M., Bremmer, F., van den Berg, A. V. (1999). "Perception of self-motion from visual flow", *Trends in Cognitive Sciences*, 3(9), 329-336

Lee, D. N. (1980). "The optic flow field; The foundation of vision", *Philosophical Transactions of the Royal Society of London, Series B*, 290(1038), 169-178

Marr, D. (1982). "Vision; A computational investigation into the human representation and processing of visual information." *New York: W. H. Freeman & Company,* NY, 8-40.

McGill, W. J., and Goldberg, J. P. (1968). "A study of the near-miss involving Weber' law and pure-tone intensity discrimination." *Perception & Psychophysics*, 4(2), 105-109.

Sivak, M. (1996). "The information that drivers use; Is it indeed 90% visual?", *Perception*, Vol.25, No.9, 1081-1089.

Car Routing and Distribution Model based Kirchhoff's Electric Current Law

Yajing Zheng and Xingchen Zhang

School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China; Email: 06121381@bjtu.edu.cn

Abstract

Railroad network is similar in some conceptions to electric current, such as a fulcrum station to a node, the car routing to the current loop and the car flow to electric current, etc. So it could put some concepts of circuit into car routing, for example, it can make formula about the traffic flow volume at the fulcrum station, which any of fulcrum stations replaces a node in circuit and the traffic flow volume passed replaces the electric current. Based Kirchhoff's electric current law, the paper addresses the car flow routing and distribution model. This model aims to minimize the expected total cost. The model can get car routing and distribution at the same time, avoiding the double-calculation (solved the car flow routing and then worked out the distribution by some traditional ways). Based on the Lingo which is implemented in C#, the paper solved the numerical experiments under the railroad network environment. The results show that this model can satisfy the actual needs and has strong adaptability.

Key Words: car routing; car flow routing plan; distribution; Kirchhoff's electric current law

Introduction

Railway transportation is based on the car routing stated, which spreads the transportation throughout. Choosing a train formation plan, it should be considered the train moving in different directions at fulcrum stations and finding the best solution of the transfer operation. Some measures taken and implemented have relation with the car routing during the daily adjustment plan. The measure chosen has close relationship with the car routing in the daily arrangement, such as

saturation loading, shut-down loading and restriction loading. How to choose a better car routing is become one of the basic problems in the railway transportation system. Generally, the problem of car routing is probably to simplify the organized mode, under the condition of the railroad network environment which states every car routing to get with it and minimize the miles of each car driving.

The main elements of choosing the car routing include:

(1)The distance of transportation;

(2)The time of transportation, and calculating it as the following way:

$$T = \sum \frac{L_{\text{section}}}{v_{travel}} + \sum t_{resorting} + \sum t_{without-resorting}$$

(3)Carrying capacity in district in each direction, sorting capacity at each station and coefficient of utilization of this capacity;

(4)The total fees of transportation, and calculating the traffic volume fees E as the following way:

$$E = N\left(\sum e_{\text{run}} + \sum e_{\text{transfer}}\right) + \sum Me_{\text{sigle}}$$

In this equation:

N ——the traffic flow volume;

e_{run} ——Each vehicle's running cost in the via railway section

e_{transfer} ——Each vehicle's cost …… in technical station;

$\sum Me_{\text{sigle}}$ ——the all running cost of all light locomotives, the M means the needed increase number which accord the car flow N, e_{sigle} means each locomotives' running cost in the via railway section.

It should be used parallel route measure to split flow, when all of the traffic volumes can't be passing the minimum cost way on account of be short of passing capacity and reorganizing capacity in some railway or some district. If the level of usage capacity is much better, after the increasing of freight volume, it may be bring to that the trains are frequency of intersection and overtaking within the district as the increasing of number of trains run. Besides it may be bring to extend the stop time at technical station with the number of increased. And when the usage capacity approaches to saturate, it is possible to lose motility of the transport organization and cause unwarranted loss. Under this condition, part of the car routing is changed to free district, it not only realizes railway technique equipment's potential, but also improves the speed of freight train service. So under the unchanged number of car flow and structure, each car routing defined by speeding the minimize time of the whole railroad network and the total cost is more reasonable to optimize the car

routing management, for example, it shows the method of stating the minimization total cost in car routing in document (SUN Yan, JIANG Lei,2005; SHI Qi-zhou,SHI Yong,1999; SUN Wan-hua，ZHENG Shi-de,1995; Wayne L; WANG Bao-hua, HE Shi-wei, SONG Rui, WANG Bo.,2007; Xie Jinxing, Xue Yi,2005).

Kirchhoff's Electric Current Law

In electronic analysis, Kirchhoff's Electric Current Law means that for each point in the circuit, the sum of current into the point is equal to the sum of current out of the point.
It shows as follows:

To the node A： $I_3 + I_5 - I_4 = 0$

Except for Node A, it can make a formula about the electric current for any of node in the circuit. This is the form of mathematical in Kirchhoff's Electric Current Law.
Railroad network is similar in some conceptions to electric current, such as a fulcrum station to a node, the car routing to the current loop and the car flow to electric current, etc. So it could put some concepts of circuit into car routing, for example, it can make formula about the traffic flow volume at the fulcrum station, which any of fulcrum stations replaces a node in circuit and the traffic flow volume passed replaces the electric current.

The railroad network has some commons with the circuit, but they also have some differences, for example, the traffic flow moves two-way, but in circuit, the electric current only moves from high voltage to low voltage; and in some section, it generally could be considered about each branch volume in traffic flow, but in circuit, it should be thought about each branch volume in electric current between two nodes. So it should be made a concrete analysis of using some ideas of circuit into railroad network according to the special characteristics.

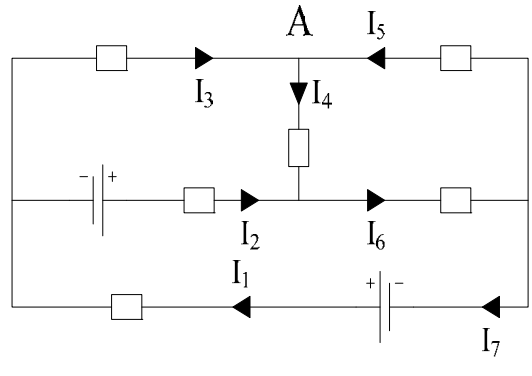


Figure 1. A sketch that indicates the Kirchhoff's electric current law.

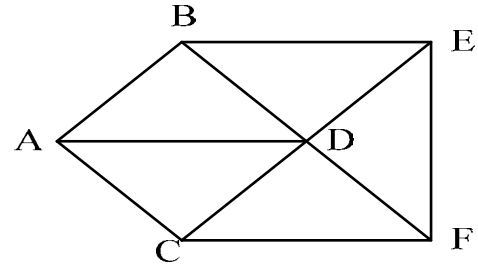


Figure 2. A sketch that indicates a railway network.

The paper states that the car routing in Kirchhoff's Electric Current Law is utilized reasonably and constructed math model suited for railroad network, and also discussed that the model is suitable in different conditions.

From circuit to railroad network

It is a graph indicating the railroad network (Figure 2). In there A、B、C、D、E、F are fulcrum stations in this network:
AB and ended at the F station; N_{AF} means the OD flow from the A station to the F station

Principle of Inseparability of the Car Flow

Railway traffic flow is a strongly controlled flow with unique operating characteristics. If rearranged in certain station, those railway traffic flows that bound for the same arrival station will be marshaled into one flow and be sent to the termination without being marshaled and split again in the progress, which is different from the distribution of road traffic; this is the Inseparability Principle of the Car Flow. Figure 3 is the diagrammatic sketch of the principle:

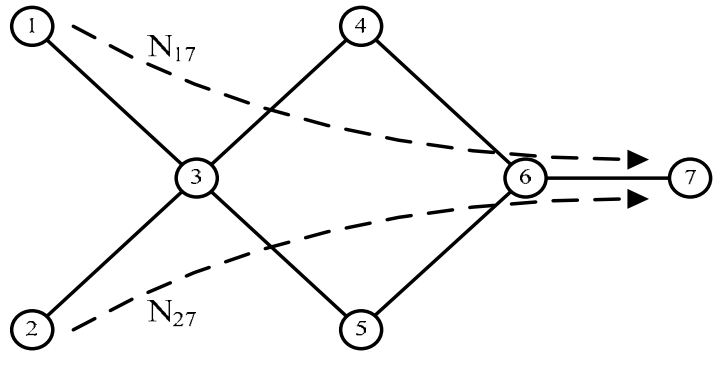


Figure 3. Diagrammatic sketch of inseparability of the car flow.

Figure 3 demonstrates two car flows: the one from fulcrum station 1 to fulcrum

station 7 and the one from fulcrum station 2 to fulcrum station 7. As for car flow N_{17} or N_{27}, it can neither be separated into two car flows nor get through from Route 3-4-6-7 or 3-5-6-7.

The Multi-Objective Plan Model of Car Routing And Distribution Model in Railroad Network

Any fulcrum station in the railway network can be given an equation following the form of fulcrum station A in Chart 2. All the equations can be used as restrictive conditions to make the model of the route map of the car flow and its distribution in railway network.

Related Descriptions and its Parametric Representation

The railway network could describe as a directed graph G(V,M,C,Ⅱ), in which V means the sets of fulcrum stations, M means the sets of connection arcs (adjoining fulcrum stations are connected by two arcs of opposite direction, which respectively indicate the up going direction and the down going direction in the section. in this paper, the order of the subscripting numbers' means an arc's direction), C means the set of expenses incurred when joint arc unitary car flow gets through. Ⅱ indicates the set of joint arc capacitances (the through capacity in the corresponding direction of the section where it is located). In order to simplify the issue, We made an assumption as follows:

(1)All the car flow are put under the administration of fulcrum stations according to the area of car flow absorption in the station. All the trains initiate from fulcrum stations.

(2)Without consideration of the station's capability of receiving and dispatching trains. (in other words, the station's capability of receiving and dispatching trains being infinite.)

(3)The expense incurred from the process of the car flow getting through the fulcrum station is 0.

The definitions of the parameters used in the model are as follows: N is the set of O-D flow (N_{ij} indicates the flowrate from fulcrum station v_i to v_j.); m_{ij} is the directed arc from fulcrum station v_i to v_j; c_{ij} is the expense needed when unitary car flow moves from v_i to v_j, b_{ij} is the through capacity of the directional arc from fulcrum station v_i to v_j, $f(m_{ik}, j)$ is the whole flowrate that gets through the adjoining fulcrum stations v_i and v, and terminates in v_j.

Modeling

Because of the principle of inseparability of the car flow, the model can't writing out directly(we still call the current model CRDNM for the sake of making our explanation easier). In order to solve this question, we can using another model, which we called as CRDM. The CRDM is modified as:
Car Routing and Distribution Model (CRDM) can be constructed as below:

$$\min E\left[\sum_i \sum_k \left(c_{ik} \sum_j f\left(m_{ik}, j\right)\right)\right] \tag{5}$$

s.t.

$$\sum_j f\left(m_{ik}, j\right) \le b_{ik}, \forall m_{ik} \in M \tag{6}$$

$$\sum_k f\left(m_{ik}, j\right) = N_{ij} + \sum_k f\left(m_{ki}, j\right) \tag{7}$$

$$f\left(m_{ix}, j\right) \times f\left(m_{iy}, j\right) = 0, \quad \forall x \neq y \tag{8}$$

$$f\left(m_{ik}, j\right) \ge 0 \tag{9}$$

The explanations of the related variables and their parameters are the same with the former model. The Expression of Restrictive Condition is added to it to ensure the Principle of Inseparability of the Car Flow.

But when O-D flow is a through train originated from one loading point or fully satisfies the Principle of Inseparability of the Car Flow, the modified model is a nonlinear plan model, the solution of which is hard to find with currently on-hand optimization software.

This paper applies the branch and bound method, successively adding in the restrictive conditions, so as to solve the model:
Which means for any fulcrums v_x and v_y（x≠y） adjoining to a certain node, they ramifies when there exist $f\left(m_{ix}, j\right) \neq 0$ and $f\left(m_{iy}, j\right) \neq 0$, after adding $f\left(m_{ix}, j\right) = 0$ and $f\left(m_{iy}, j\right) = 0$ into （CRDM） respectively, the original object is ramified into two sub-object plans, and then find the solutions:
Sub-object plan 1:

$$\min E\left[\sum_i \sum_k \left(c_{ik} \sum_j f\left(m_{ik}, j\right)\right)\right] \tag{10}$$

s.t.

$$\sum_j f\left(m_{ik}, j\right) \le b_{ik}, \forall m_{ik} \in M \tag{11}$$

$$\sum_k f\left(m_{ik}, j\right) = N_{ij} + \sum_k f\left(m_{ki}, j\right) \tag{12}$$

$$f(m_{ix}, j) = 0 \tag{13}$$

$$f(m_{ik}, j) \geq 0 \tag{14}$$

Sub-object plan 2:

$$\min E\left[\sum_i \sum_k \left(c_{ik} \sum_j f(m_{ik}, j)\right)\right] \tag{15}$$

s.t.

$$\sum_j f(m_{ik}, j) \leq b_{ik}, \forall m_{ik} \in M \tag{16}$$

$$\sum_k f(m_{ik}, j) = N_{ij} + \sum_k f(m_{ki}, j) \tag{17}$$

$$f(m_{iy}, j) = 0 \tag{18}$$

$$f(m_{ik}, j) \geq 0 \tag{19}$$

If we want to solve the CRPDNAM, can we add into the railway network the O-D flow successively. We give a solution to the car routing on condition that the flow is inseparable every single time an O-D flow is added in, and based on the final result, we abstract from each network the flowrate distributed to it in this process and then, viewing the network as one with a car flow of 0, add in an O-D flow (in descending order so as to minimize the expense), when all the O-D flowrate are added into the railway network, we can easily get the car routing on the condition that each of the O-D flow are through train originated from loading points):

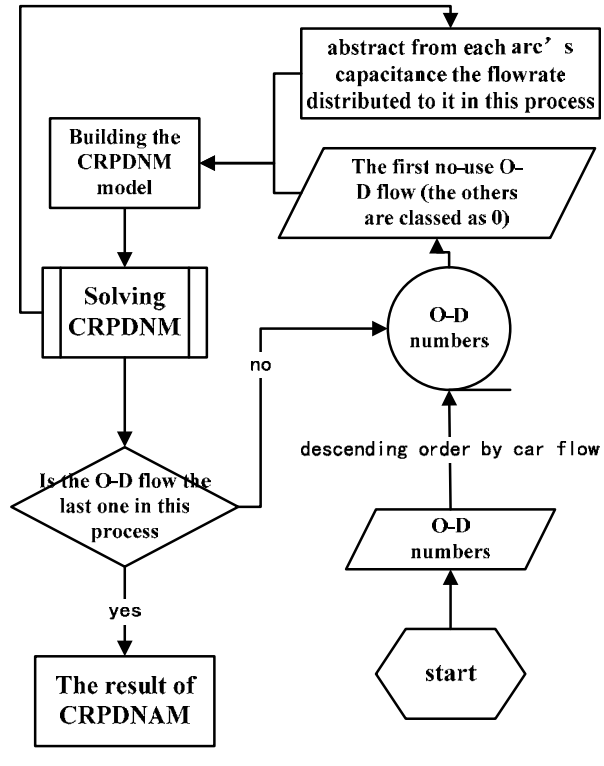


Figure 4.

Specially, when there are differences of priority levels, it is of more significance to shorten the running distance of the prioritized cargo, we need therefore ensure the shortest route being selected by it. While according to the model we advanced here in this paper, this issue can be easily solved by adding in O-D car flow into the model after they are arranged in the order of priority rating. We are not giving any further discussion on this.

Example

Figure 5 shows the railway network of the example, which we take to check the effectiveness of the model and algorithm. In the diagraph, number 1~8 indicates the eight fulcrum stations in the railway network. We use Lingo 10.0 give the solution.

The expenses of unitary car flow getting through are show in Table 1, the car flow (OD flowrate, the differences of the types of trains are omitted) between fulcrum stations are shown in Table 2, capacity of the joint arc（through capacity of the line） are shown in Table 3:

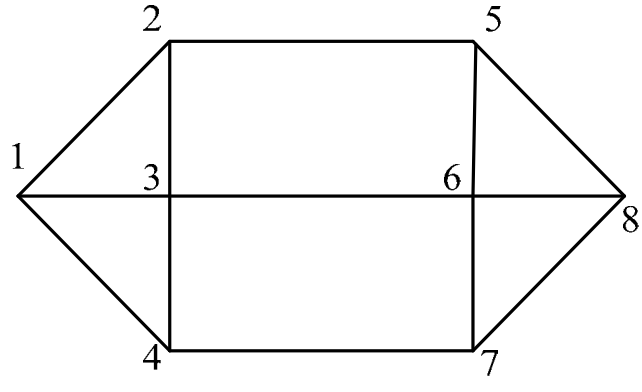


Figure 5.

Table 1. Expenses of car flow getting through sections.

section	Expense/yuan	section	Expense/yuan
1-2（2-1）	3	4-7（7-4）	6
1-3（3-1）	4	5-6（6-5）	9
1-4（4-1）	8	6-7（7-6）	3
2-3（3-2）	5	5-8（8-5）	11
3-4（4-3）	3	6-8（8-6）	8
2-5（5-2）	11	7-8（8-7）	4
3-6（6-3）	8		

Table 2. Car flow between fulcrum stations.

O-D number	Starting	Terminat	Flowrate(unit:train)
1	1	8	250
2	2	8	440
3	5	7	450
4	3	8	360
5	3	7	510
6	5	4	180
7	5	1	120
8	2	7	630

Table 3. Capacity of the joint arc.

Section	Capacity	Section	Capacity
1-2（2-1）	1000	4-7（7-4）	1100
1-3（3-1）	500	5-6（6-5）	550
1-4（4-1）	500	6-7（7-6）	2000
2-3（3-2）	700	5-8（8-5）	2000
3-4（4-3）	1000	6-8（8-6）	360
2-5（5-2）	2000	7-8（8-7）	1500
3-6（6-3）	1200		

The solution can be found through computer programming, with the help of software Lingo (using C# to call the program):

Table 4. Scheme of the distribution of car routing (CRPDNAM).

O-D number	Routing and the distribution of flowrate
1	$1 \xrightarrow{250} 3 \xrightarrow{250} 6 \xrightarrow{250} 7 \xrightarrow{250} 8$
2	$2 \xrightarrow{440} 1 \xrightarrow{440} 4 \xrightarrow{440} 7 \xrightarrow{440} 8$
3	$5 \xrightarrow{450} 6 \xrightarrow{450} 7$
4	$3 \xrightarrow{360} 6 \xrightarrow{360} 7 \xrightarrow{360} 8$
5	$3 \xrightarrow{510} 6 \xrightarrow{510} 7$
6	$5 \xrightarrow{180} 8 \xrightarrow{180} 7 \xrightarrow{180} 4$
7	$5 \xrightarrow{120} 2 \xrightarrow{120} 1$
8	$2 \xrightarrow{630} 3 \xrightarrow{630} 4 \xrightarrow{630} 7$

Current target value min=44680yuan

Conclusion

From the explanation, the model which is based on Kirchhoff's Electric Current Law to choose the car routing and divide the traffic flow can meet the need to choose the

car routing in many limited conditions. And according to the actual demand, the car routing in different conditions is obtained under using the model. A example that is calculated the result in 1min under three different conditions can meet the requirement in practice work, so it proves that the model and example conduces to solve the problem of choosing the car routing in practice condition.

References

He S. W., SONG R, DAI X. L., YANG Y. K., and LEI Z. L.. (2003). "Study on optimal method for evaluating the carrying capacity of railroad network." *Journal of the China Railway Society*, 1-9.

Lin B. L.,et al.(1996) "Optimal weighted car routing in railway network." *Journal of Northern Jiaotong University*, 20(6), 645-650.

Sun Y., JIANG L.(2005) "Route optimizing method of loaded car flow with capacity limit in railway network." *Railway Transport and Economy*, (12), 82-84.

Shi Q. Z., SHI Y. (1999) "Multi-objective linear-programming model and its algorithm for car flow routing with bidirectional heavy and empty cars in railway network." *Journal of the China Railway Society*, 21(1), 1-9.

Sun W. H, ZHENG S. D. (1995) "Study on the optimization algorithm of railway wagon flow path." *Journal of Northern Jiaotong University*, 19(Sup.): 39-44.

Wayne L. Winston Operations research : applications and algorithms eng, Beijing.

Wang B. H, He S. W., SONG R., WANG B. (2007) "Stochastic dependent-chance programming model and hybrid genetic algorithm for car flow routing plan." *Journal of the China Railway Society*, 8, 6-11.

Xie J., Xue Y. (2005) "Optimization modeling and LINDO/LINGO." *Beijing. Tsing-Hua university press*.

Yang H. (2001) "Organization of rail transport." Beijing. *China Railway Pres,* 213-214.

Zha W. X., QIN Z. R.(1997) "Optimal flow allocation algorithm for the wagon path on the railway network." *Journal of Northern Jiaotong University*, 21(3), 259-263.

Three Dilemma Zone Strategies for High-Speed Rural Intersections: Comparison of Field Results

Alixandra Demers[1], George F. List[2] and Isaac K. Isukapati[3]

[1] Research Assistant Professor, [2] Professor and Department Head, [3] Graduate Research Assistant, Department of Civil, Construction, and Environmental Engineering, North Carolina State University, 208 Mann Hall, Campus Box 7908, Raleigh, NC, USA 27695; Email: ademers@ncsu.edu, gflist@ncsu.edu, isaac_assiduous028@yahoo.com

Abstract

High-speed, rural intersections with both passenger car and truck flows present dilemma zone challenges. Three signal-timing strategies were tested in the field at three intersections of varying geometries with the goal of minimizing the occurrence of dilemma zones without sacrificing efficient operation. The strategies focused on hardware changes impacting vehicle sensing as well as mainline green phase extension and termination. The base strategy was volume-density control. The second one (NQ4) added advance detection of high-speed trucks triggering a fixed green extension. The final strategy replaced the second system with a sophisticated advance detection and control system (D-CS) logic based on the work of Bonneson et al. (2002). Results indicate simple volume-density control is surpassed in safety by both other strategies; moreover, the detection-control system dramatically reduced the number of vehicles trapped in dilemma zones at the onset of amber. The second strategy works fairly well in the field. The drawbacks to the NQ4 system are most noticeable at high-volume intersections – it does not actually find times when no vehicles are in dilemma zones, vehicle speeds are not directly used for computing main street hold times (which is accomplished in the D-CS control strategy resulting in improved efficiency), and it is a bit cumbersome and expensive to install. Based on our findings, the addition of the DC-S algorithm into a controller is a worthwhile investment for mixed-traffic, high-speed, rural intersections with dilemma zone issues.

Key Words: dilemma zone; high-speed intersection; field experiment; truck safety

Introduction

This paper reports field results from tests of instrumentation and control strategies for dilemma zone minimization at high-speed rural intersections. Different practices have been created nationwide and worldwide to minimize the number of crashes that occur when main street green terminates (List et al. 2007a & 2007b). The challenge is to find a time when all vehicles can either safely stop or proceed ahead without conflict. Otherwise, upon the onset of amber, if a vehicle is neither able to stop safely before the intersection nor clear the intersection before the conflicting green begins the vehicle can be caught in a "dilemma zone." The decision of the driver to stop at the intersection may cause a rear-end collision; a decision to go through the intersection may result in a right-angle collision.

It is possible to forecast when a vehicle will be in its dilemma zone if one measures its speed and deceleration (acceleration) rate in advance of the intersection (Gates et al. 2007; Kell et al. 1990; Liu et al. 2007). Given these two attributes, the next challenge is to predict the trajectory being followed (List & Eisenman 2006; Liu et al. 2007), and then create a control scheme that ensures no vehicle is in a dilemma zone when the green is terminated (Kell et al. 1990; Bonneson et al. 2002). A sensing strategy is needed along with a control strategy. The former makes it possible to determine the vehicle trajectories; the latter decides when to initiate the amber. For example, if there is just one lane on the main street approach, and all the vehicles are approaching at a known constant speed, then it is quite simple to determine where the sensors need to be placed (Rodgers 1973) and how long the onset of the amber should be delayed. However, in the real world there are significant differences in the speeds of the vehicles; so a more sophisticated strategy is needed.

Consider that Vehicle detection or sensing is currently used for multiple tasks including signal control, ramp metering, and data gathering for traveler information; however, such tasks may require detection in different places along a roadway section. Within signal control there are the oft competing goals of efficiency (Rodgers 1973, Cribbins & Meyer 1975, and Tarnoff & Dufford 1977) and safety (Zegeer & Deen 1978, McCoy & Pesti 2003, and ITE 2003) with recent solutions attempting to meet both needs (Pant & Cheng 2001, Bonneson et al. 2002). Advance warning systems have been examined (McCoy & Cheng 2001, Sunkari et al. 2007) which can reduce dilemma zone occurrence by making drivers aware of the signal's state; However, some municipalities are wary of implementing such systems because of liability concerns if malfunctions happen; hence, this study was undertaken to assess methods that did not require dynamic signage. In this project, three detector/control strategies were tested, as detailed in Table 1.

Data Collection & Extraction

Data were collected at each of three intersections on both main street approaches for every control strategy. Since one of the three intersections already had LVS installed,

only two control strategies were studied at that location. Hence, field studies were conducted for 16 "scenarios" (3+3+2 strategies)*(2 directions).

Table 1. Overview of dilemma zone minimization strategies.

	Strategy 1	Strategy 2	Strategy 3
	Volume-density	Long vehicle system	Detection-control system (Bonneson et al., 2002)
Controller	2070 Eagle	2070 Eagle	2070 Naztec with D-CS Logic
Additional Traffic Control Devices	Detectors placed 121.0 m (400') upstream of stopbar for controlling gap timers.	Two NQ4 LVS devices, one per main street approach.	None
Speed Traps	None	One in each approach lane, 302.4 m (1,000') upstream of stopbar.	One in each approach lane, 302.4 m (1,000') upstream of stopbar.
Dilemma Zone Prevention Action	Relies on standard signal timing principles.	Fixed-time down counter is triggered every time a logical statement of the following form is met: if ((vehicle > 6.7 m (22') long) $\wedge$ (speed > 88 Kph (55 mph))) then reset down counter to 12 seconds. Main street green is extended until max green is reached.	Variable-time down counter is triggered and set based upon each vehicle's length and speed trajectory. In part 1, all vehicles are considered; in part 2, only long vehicles are considered since the main street cycle is nearing the max green time.

Note: Thereby 2 was modified in the field because of equipment constraints. Only the outside through lane had a speed trap.

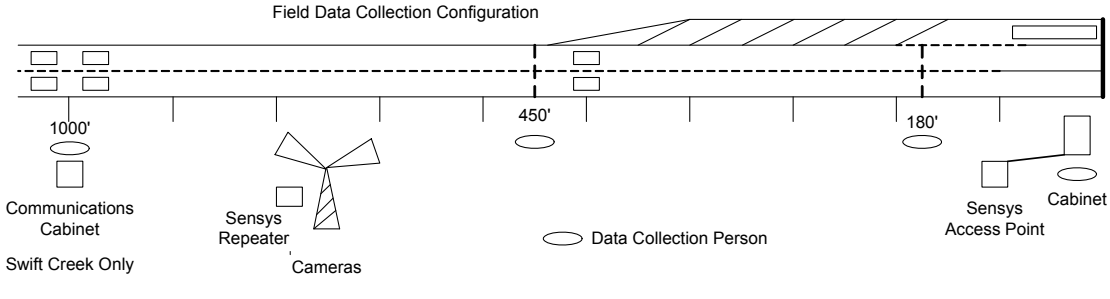


Figure 1. Prototypical instrumentation configuration.

Data were collected as Figure 1 shows. Observers were stationed at each of four locations: at the stopbar as well as 54.4 m (180'), 136.1 m (450'), and 302.4 m (1,000') upstream of the stopbar (Dilemma zones are often thought of in terms of distance to the stop-line, especially since this is easiest to instrument and our team had fixed counters at four locations; however, Zegeer and Deen (1978) suggest that a better measure is time. They assert that the dilemma zone extends from 6 to 2 seconds upstream of the stopbar; this definition has been adopted here for the analysis.) Time stamps were recorded by lane and vehicle type every time a vehicle passed by one of these monitoring points. Signal timing data were also recorded by the observers at 54.4 m and at the stop bar.

A computer program developed by the authors estimated approach speeds and deceleration rates for each vehicle, then ascertained how many vehicles were in dilemma zones at each onset of amber.

Study Sites

More than thirty sites were considered for study. Five were ultimately selected for detailed simulation analyses and three of these were subsequently chosen for field experimentation. Each of the latter intersections is summarized in Table 2.

Equipment Information

The authors do not endorse the use of the vendors and products mentioned, rather, it is the methods that are of utmost importance. However, specific vendors and products are mentioned in this paper since this was a field experiment rather than a simulation exercise and repeatability may only work with the same or similar equipment. The key items used were:

- 2070 Eagle controller running OASIS software (a North Carolina standard)
- NQ4 Long Vehicle System (LVS) devices
- 2070 Naztec controller with D-CS logic incorporated
- Sensys Networks wireless magnetometer point detectors

Results

The ultimate question for the project was this: can improved performance be achieved in the field? Can a better control strategy be found? In essence, the research team wanted to determine if the NQ4 and D-CS control configurations performed better than the baseline volume-density control, and which one was "best", with fewer vehicles in dilemma zones along with no increase in delays.

The strategy for answering these questions was as follows. The time-stamps from 302.4 m, 136.1 m, 54.4 m, and 0 m were studied to see if there were differences between the control strategies in the number of vehicles caught in dilemma zones at the onset of amber. Plus, comparisons of the ability to keep vehicles from being caught in the dilemma zones at the onset of amber were examined. Finally, cycle length trends were evaluated for signal operation efficiency.

Table 2. Field site differences.

	SWIFT CREEK	**US 17**	**NC 280**
Region	Piedmont	Coast	Mountains
Mainline Topography	Sag curve	Flat	Crest curve+
Intersection Type Mainline Geometry Side Street Lanes	4-way intersection 4-lane divided highway Flared, but no left-turn bays	T intersection 4-lane divided highway Left, shared left/right, short flared right-turn bay	4-way intersection 4-lane highway with a TWLTL Flared, but no left-turn bays
Photo N W E S			
Signal Phasing	6-phase volume-density control (1 side street) with NQ4 LVS superimposed; mainline lefts are protected	3-phase volume-density control with mainline turns protected and permitted	2-phase volume-density control with permitted mainline lefts
Min and Max Greens on ML and MA^	ML 14 sec, 120 sec MA 7 sec, 30 sec	ML 14 sec, 90 sec MA 7 sec, 30 sec	ML 14 sec, 60 sec MA 7 sec, 20 sec
Total Volume*	High	High	Low
Truck Volume*	High	High	Low

* To the nearest 1,000 vehicles per day.

+ Complex topography: Both approaches of NC 280 lead uphill to the intersection whereas there is a downslope to the intersection on both side street approaches. This suggests that main street traffic should not commonly incur dilemma zone issues because the uphill approaches to the intersection would aid in deceleration. However, if traffic exits either side street approach quickly, collisions might occur.

ML = Mainline, MA = Minor Approach, TWLTL = two-way left-turn lane

^ For the three sites, through/right movement amber and all red times ranged from 4.7 to 5.4 seconds and 1.2 to 1.9 seconds, respectively.

While it is important to recognize that the dilemma zone herein is defined in terms of time, not location, distance-based locations were required for an efficient data collection strategy. On these high-speed approaches, the vehicles are all traveling at approximately 105 kph (65 mph), so the vehicles are about 54.4 m, upstream of the stop bar when they are 2 seconds away; at 136.1 m, when they are 5 seconds away, and at 163.3 m when they are 6 seconds away. (At 302.4 m, they are about 11 seconds away.)

Ideally, as can be seen in Figure 2, one would like to look backwards in time to see how far away the vehicles were (in time) at t_y, the onset of amber. As was done with the simulation analyses (List et al. 2009), this would involve checking the vehicle trajectories and ascertaining which vehicles were 2-6 seconds away at t_y. From

videotapes, this can be done, but the analysis is labor intensive, and requires extensive field instrumentation (The research team's original plan was to create the dilemma zone counts based on synthesized vehicle trajectories, stitched together from the time stamps at 302.4 m, 136.1 m, 54.4 m , and 0 m, but to-date this has proved to be a significant challenge because of the sensitivity to the synchronization among the time stamps and the not-quite-complete condition of the time stamps collected). Another choice is to take a snapshot at t_y and count the number of vehicles that are 163.3 m to 54.4 m away from the stop bar. This is approximately the same as the number of vehicles that are 2 to 6 seconds away from the intersection.

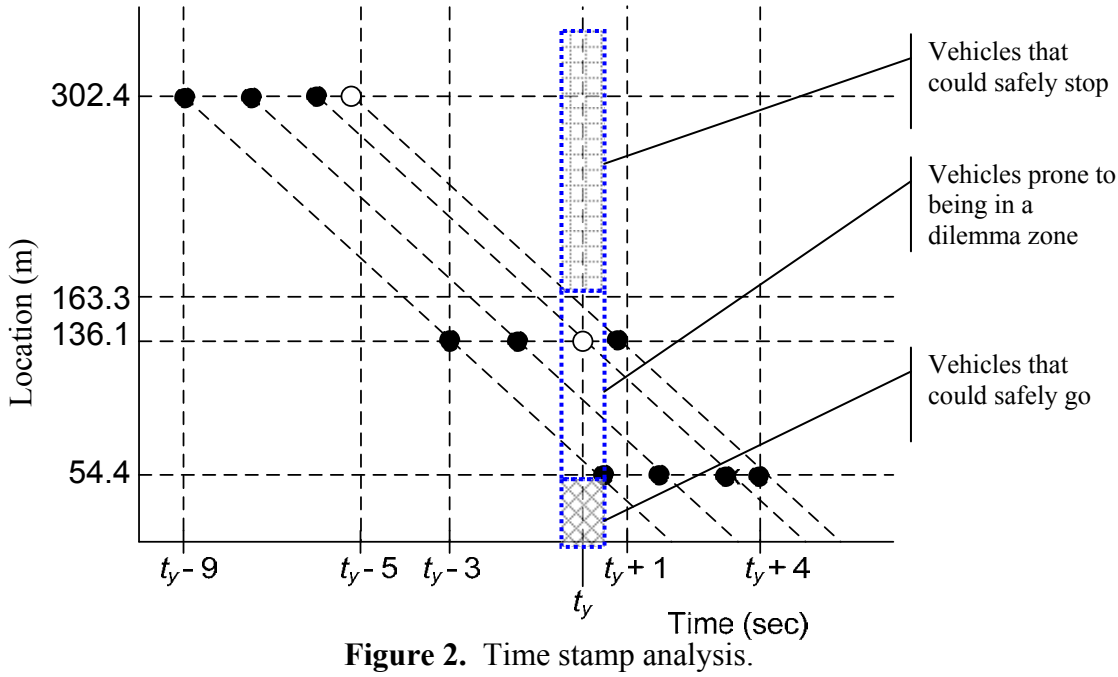


Figure 2. Time stamp analysis.

Recognizing the speeds are not constant, the research team focused directly on the time stamps collected at the individual locations. Consequently, a Visual Basic for Applications program was created by the authors to count the number of vehicles observed at 302.4 m, 136.1 m, and 54.4 m in time windows corresponding to the 2-6 second time window in advance of t_y. For example, as seen in Figure 2, vehicles observed at 54.4 m between t_y and t_y + 4 are about 2-6 seconds upstream of the intersection. When time = t_y, at 136.1 m, vehicles observed between t_y – 3 and t_y +1 are about 2-6 seconds upstream of the stop bar at t_y. Similarly, at 302.4 m vehicles observed between t_y – 9 and t_y - 5 are 2-6 seconds away at t_y. The research team decided to count these occurrences at 54.4 m, 136.1 m, and 302.4 m to be assured that a correct assessment of performance was being obtained, in part because this compensated for time stamps that might have been missed. For example, in Figure 2, the open circle just shy of t_y – 5 is intended to represent a vehicle missed, one for which a time stamp was not entered by the 302.4 m observer. Similarly, at 136.1 m a second open circle is shown, indicating that another time stamp had been omitted. So in this instance, the counts at 302.4 m and 136.1 m would have been 3, while the count at 54.4 m would have been 4. In the aggregate, since the data collection was

quite good, no systematic biases were found among the locations. While individual vehicles might not have been recorded at each location at a specific point in time, in the aggregate, the results from the three locations are consistent and defensible.

Summary Results for Rll Rites. Tables 3 and 4 summarize the results for all six field tests that were performed. The first column in the former shows the percentage of times that no vehicles were 2-6 seconds from the stop bar based on the observations at 54.4 m. For example, in the base case (existing conditions) at US 17 southbound, 64.3% of the time no vehicles were 2-6 seconds from the stop bar. The actual times, as shown in the last column of Table 4, when ambers were invoked have the same characteristics: 62.7% of the time, no vehicles were present. In contrast, the NQ4 control configuration selected times when no vehicles were 2-6 seconds upstream of the amber 77.1% of the time and the D-CS strategy increased that percentage to 91.8%. Moreover, while the base case average cycle length was 62.7 seconds (Table 3 bottom right), the NQ4 control strategy increased the cycle length to 147.9 seconds while the D-CS control strategy kept it constant at 66.6 seconds. Hence, the NQ4 and D-CS control strategies both improved safety, but the D-CS control strategy did it without sacrificing efficiency.

Table 3. Field results for all sites based on observations at 54.4 m.

			AII Vehicles – Cum Prob of being in DZ							
			CumProb(0)		CumProb(1)		CumProb(2)		**AvgCyc (sec)**	**Total Ambers**
Site	**Direction**	**Control**	Amber	Other	Amber	Other	Amber	Other		
Swift Creek	EB	NQ4	69.3	42.5	22.0	30.7	7.1	17.2	78.7	127
		D-CS	84.4	36.4	13.3	33.9	2.2	19.8	62.8	135
	WB	NQ4	83.3	68.3	13.0	17.0	3.7	9.2	95.4	108
		D-CS	87.0	58.8	9.8	21.8	3.3	11.9	77.3	184
US17	NB	Existing	87.2	53.9	10.6	27.2	2.1	12.9	68.0	188
		NQ4	85.3	80.5	12.1	12.6	2.6	5.0	59.7	232
		D-CS	83.8	47.4	15.1	33.1	1.1	14.2	56.9	284
	SB	Existing	62.7	64.3	31.3	27.2	6.0	7.0	62.7	252
		NQ4	77.1	56.5	20.0	28.7	1.9	10.7	147.9	107
		D-CS	91.8	55.7	7.7	28.8	0.4	10.9	66.6	233
NC280	EB	Existing	72.6	63.4	24.2	27.2	3.2	7.5	150.7	95
		NQ4	72.1	59.7	14.4	23.3	10.8	10.3	125.8	111
		D-CS	79.3	66.6	19.5	25.1	1.2	6.6	166.1	82
	WB	Existing	49.4	56.3	41.4	25.7	8.0	11.3	182.8	87
		NQ4	77.8	64.3	20.4	26.4	1.9	7.5	163.4	108
		D-CS	84.5	60.8	12.6	23.2	1.9	10.2	154.2	103

Note: CumProb(#) = the cumulative probability that # vehicles will be in a dilemma zone, shown in percents.

Table 4. Probabilities of no vehicles in dilemma zones at time t_y.

			Probability of No Vehicles		
Intersection	Direction	Control	1000'	450'	180'
Swift Creek	EB	NQ4	59.1	78.7	69.3
		D-CS	76.3	74.8	84.4
	WB	NQ4	83.3	75.9	83.3
		D-CS	81.5	84.2	87.0
US 17	NB	Existing	58.5	51.6	87.2
		NQ4	81.9	66.8	85.3
		D-CS	58.8	64.1	83.8
	SB	Existing	73.0	77.8	62.7
		NQ4	60.0	66.7	77.1
		D-CS	74.7	47.2	91.8
NC 280	EB	Existing	66.3	67.4	72.6
		NQ4	57.7	64.9	72.1
		D-CS	65.9	61.0	79.3
	WB	Existing	65.5	66.7	49.4
		NQ4	67.6	69.4	89.8
		D-CS	62.1	68.9	84.5

Note: Probability of No Vehicles is shown in percents (0 to 100) instead of decimals (0 to 1).

Another interesting observation is that the number of vehicles observed in the dilemma zone counts often decreases as one moves from 302.4 m to 136.1 m and 54.4 m. Notice in Table 4 that in the case of the D-CS control strategy, this is almost always the case. For example, westbound, P(0), the probability of no vehicles being in a dilemma zone at the onset of amber rises in the D-CS case from 81.5% to 84.2% and then 87.0%. It is always true for the D-CS cases if the P(0) estimate at 302.4 m is compared with that from 54.4 m. It is true in three of the nine cases that the 136.1 m value also has to be increasing (i.e. maintaining a trend). The implications are twofold. First, once drivers see the amber, they are taking actions to move themselves out of a dilemma zone as they approach the intersection, mainly by slowing down. Second, the rise in these values is most dramatic for the D-CS control strategies, slightly surpassing the NQ4 results, which means the D-CS strategy is performing best.

Cycle Lengths – All Sites. Finally, while the incentive for exploring more advanced signal control strategies was principally safety, efficiency was also of interest, and an examination of signal cycle lengths provided insight into that issue.

Figure 3 shows the cumulative density functions for cycle lengths for each of the sites and approaches studied in the field. A good outcome is the curve highest and furthest to the left. The cycle lengths for that control configuration are the shortest. Notice that in the case of US 17 southbound, the D-CS (naz) has the distribution that is highest and furthest left, slightly better than that for the NQ4. It is also the best for US 17 northbound, and in this case it is by itself. In the case of NC 280, both

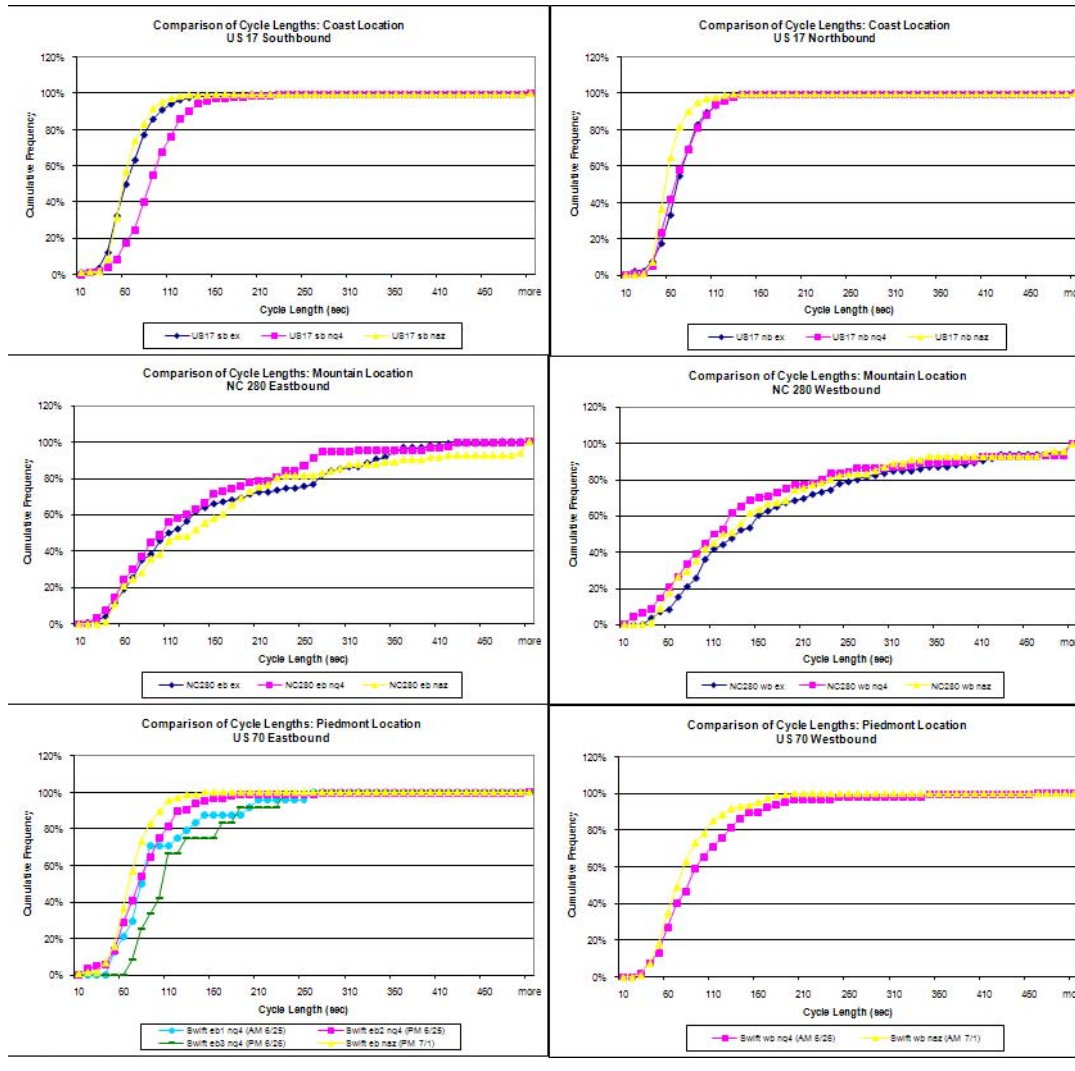


Figure 3. Trends in cycle lengths.

eastbound and westbound, the D-CS result is in the middle, performing on par with the NQ4 strategy. A potential explanation the performance of the D-CS strategy is not distinctively better is the truck (and total) volumes at the NC 280 intersection are low. In the case of Swift Creek, either eastbound or westbound, the dominant performance of the D-CS control configuration is again apparent.

Conclusions

Several conclusions can be drawn from this research. First NCDOT is at or near the forefront of signal control strategies for high-speed rural intersections. During the literature review, a comprehensive assessment of signal control techniques and technologies was conducted. The research team found that few states have a signal design practice that is as well documented as NCDOT's and the procedures in use are at the frontier of best practice.

Second, the NQ4 strategy currently employed at select locations works fairly well in the field. NCDOT can do better, but the practice provides good solutions, certainly better than volume-density control alone, based both on the field simulation experiments. Found were the drawbacks NQ4 system that it does not actually find instances when no vehicles are in dilemma zones, which the D-CS control strategy does, it does not directly take into account for vehicle speed in computing the main street hold time it conveys to the controller, it tends to lengthen the distribution of cycle lengths which means efficiency suffers (a disadvantage the D-CS system avoids) and it is a bit cumbersome and expensive to install. For example, AC power must be brought to a cabinet adjacent to the loops and an AC signal must be brought back to the controller cabinet.

The third conclusion is that the D-CS control strategy developed by Bonneson *et al.* (2002) consistently works very well. In most field tests and all of the traffic simulation model runs, it performed the best. It dramatically reduced the number of vehicles that were trapped in dilemma zones at the onset of amber in the field. It made the intersections quieter; less mainline traffic was brought to an unexpected stop. It shortened the side street queues, shortened the cycle length, and made the signal more responsive to minor movement calls. The addition of the D-CS algorithm into a controller is a worthwhile investment for high-speed, mixed traffic, rural intersections with dilemma zone issues.

References

Bonneson, J. A., Middleton, D., Zimmerman, K., Charara, H., and Abbas, M. (2002). *Intelligent Detection-Control System for Rural Signalized Intersections*, Report Number FHWA/TX-03/4022-2, prepared for Texas DOT, August.

Cribbins, P.D., and Meyer, C.A. (1975). "Choosing detectors – and placing them right", *Traffic Engineering*, 15-18.

Gates, T.J., Noyce, D.A., and Laracuente, L. (2007). "Analysis of dilemma zone driver behavior at signalized intersections". *86th Annual Meeting of the Transportation Research Board CD-ROM*, Washington D.C.

Institute of Transportation Engineers (2003). *Making Intersections Safer: A Toolbox of Engineering Countermeasures to Reduce Red-Light Running* (An Informational Report), ITE, Washington, D.C.

Kell, J. H., Fullerton I. J., and Mills, M. K. (1990) *Traffic Detector Handbook* (FHWA-IP-90-002), McLean.

List, G.F., Demers, A., Isukapati, I.K., Hauser, E., and S. Pulugartha (2009). *Placement of Detection Loops on High Speed Approaches to Traffic Signals,* HWY 2007-13 Final Report, prepared for North Carolina DOT, November.

List, G.F., Demers, A., Isukapati, I.K., Hauser, E., and S. Pulugartha (2007a). *Placement of Detection Loops on High Speed Approaches to Traffic Signals,* HWY 2007-13 Task 1 Report, prepared for North Carolina DOT, June.

List, G.F., Demers, A., Isukapati, I.K., and Hauser, E. (2007b). *Placement of Detection Loops on High Speed Approaches to Traffic Signals,* HWY 2007-13 Task 2 Report, prepared for North Carolina DOT, June.

List, G.F., Eisenman, S. (2006). "Identifying vehicle trajectories and turning movements at roundabouts", *5th International Symposium on Highway Capacity and Quality of Service*, Yokohama, Japan.

Liu, Y., Chang, G., Tao, R., Hicks, T., and Tabacek, E. (2007). "Empirical observations of dynamic dilemma zones at signalized intersections", *86th Annual Meeting of the Transportation Research Board CD-ROM*, Washington D.C.

McCoy, P.T., and Pesti, G. (2003). "Dilemma zone protection on high-speed signalized intersection approaches: advance detection versus advance warning flashers and advance detection". *82nd Annual Meeting of the Transportation Research Board CD-ROM*, Washington D.C.

Pant, P.D. and Cheng, Y. (2001). *Dilemma Zone Protection and Signal Coordination at Closely-spaced High-speed Intersections*, Final Report, prepared for Ohio DOT, November.

Rescot, R.A. (2007). *Automating Traffic Studies at Modern Roundabouts – A Feasibility Study*, Masters Thesis, University of Missouri-Columbia.

Rodgers, L.M. (1973). "Detector placement – a misunderstood subject", *Traffic Engineering*, pp. 42-44.

Sunkari, S.R., Messer, C.J., and Charara, H. (2005). "Performance of advanced warning for end of green system for high-speed signalized intersections". *Transportation Research Record: Journal of the Transportation Research Board, No. 1925, Transportation Research Board of the National Academies*, Washington, D.C., 176–184.

Tarnoff, P.J., and Dufford, D. (1977). "Surveillance system design for computerized traffic control", *Transportation Engineering*, pp. 15-19.

Zegeer, C.V. and Deen, R.C. (1978). "Green-extension systems at high-speed intersections." *ITE Journal*, *Institute of Transportation Engineers*, Washington, D.C..

The Method of Survey and Statistics of Passenger Transport Volume in County Highway Networks

Liqiang Wang[1,2], Xiansheng Li[1], and Yuanyuan Ren[1]

[1] Ph.D. candidate, College of Transportation, Jilin University of China, No. 5988, People Street, Changchun, China 130022, Email: Wangliqiang11111@sohu.com
[2] Senior Rearcher, Heilongjiang Provincial Institute of Highway and Transportation, No.92, Qingbin Road, Harbin, China 150080

Abstract

The county economy in China is the fundamental regional economic unit, and the passenger transport volume is one of the basic indicators to reflect a county economic development. In this paper, by considering the characteristics of the operation of passenger highway transport, the operating passenger cars would be divided into three classes: 1) cars that take operation in the passenger terminal, 2) cars that do not take operation in the passenger terminal but with fixed-lines, and 3) cars that are rentals. Their investigation methods are - statistical reports collection or the concentration survey in passenger terminal, car-survey method and sample survey method in business place. According to general characteristics of the above three types of passenger transport volume and the actual survey conditions of grass-roots management departments, the simplified, easy-to-use statistical model was proposed. It is a two-stage calculating model, which combines multi-class traffic projections and road network transport volume (fluctuations) projections. Finally, the feasibility and effectiveness of the method have been verified by the investigation of a county in Heilongjiang Province.
Key Words: highway transport volume; operation of passenger cars; survey method; statistics method; county

Introduction

At present in China, the provincial operation passenger highway transport volume is counted in accordance with sampling car membership per year (Chen 2005; DOM 2007). Recently with the increase of individual transport operators, the difficulty of sample statistics is gradually increasing. In the “Eleventh Five-Year” period, the road

"Mastery" project, and "Patency" project were gradually accelerated, and the county road network has been improved greatly, which also led to the growth of road passenger transport volume in the county and townships. However, we cannot obtain the road passenger transport volume at the county level according to the current statistical system, in which the sampling frame and the statistics model are designed to obtain the transport volume of only one province (Cen et al. 2000). The government departments had difficulty accurately grasping the basic situation of road transport within the county (city). The result was that they did not reflect the tremendous achievements of the rural road construction and also could not make an objective assessment and scientific plan on the county road transport.

In China, the county economy is the most fundamental regional economy unit, while the highway operation transport volume of passengers is one of the basic indicators, which reflects the level of the countys economic development. Unfortunately, we have not yet researched how to infer the highway operation transport volume of passengers in a county, as well as the existing survey, and statistical methods in transport volume statistics could not be suited for statistics of that transport volume value. So, in this paper, a simple and easy method of investigation and statistical projections for highway passenger transport volume is proposed based on the relevant existing idea of survey and statistical methods at home and abroad.

The Characteristics of Existing Survey and Statistical Methods

The Foreign Existing Survey Methods about Road Passenger Transport Volume. The foreign market economic system made its industry-wide statistical work of transport information carried out earlier. Because of its market economic structure on statistical surveys, the indicators and survey methodology have shown the characteristics of early study and abundant achievements.

The survey of passenger transport volume mainly was the Natural Household Travel Survey (NHTS), which the Bureau of Transportation Statistics of the U.S. Federal Department and Federal Highway Administration had implemented. NHTS collected a sample of 26,000 households nationwide. In addition to the national sample surveys, NHTS collected another sample of 40,000 families in the five regions and four states belonging to highly urbanized areas. The survey contained several features of travel on a specific day (travel day), including travel mode, trip purpose, and trip distance. In addition, transport enterprises were required to provide long-distance passenger information, and the public transport database of some associations such as the United States Public Transportation Association was made as the data supplement. In the 1970s, the transport industry became deregulated, causing access management data and passenger information from the passenger business to become increasingly difficult to gather. In 2001, long-distance passenger data collected by the U.S. National Household Travel was limited more than ever before. The experts considered that the surveyors should take all measures to address the vacancy passenger data, and adopt a flexible solution to the emergence of new data gaps in time (Chen, 2005; DOM, 2007; Cen et al., 2000; USDOT, 2003).

The motor transport survey method used in Japan is mainly a method based on on-the-spot transport investagation and served with a questionnaire approach. In addition to the direct investigation into the operation of road vehicles, the random surveys were also carried in about 1,600 bus enterprise sales and 8,000 trucking sales, and the survey method was a questionnaire directed to users. To investigate properly, one must survey all of the operating vehicles in these places.

In short, the transport survey work in developed countries was mostly based on the transport production sites, such as roads survey sites, sales offices, the bureau of investigation, and so on. The method can effectively avoid the omission of the sample and subsequent errors in the investigation, which can ensure the accuracy of survey data.

However, two above-mentioned methods were not successful because many Chinese people resisted answering the questionnaire and inquiries because of the exposure of their private information (Yang and Qin 2009). Also, these methods of gathering information are costly and require a lot of manpower to implement.

Chinas Current Survey Methods about Road Passenger Transport Volume. The history of the Chinese road transport volume survey and statistics had experienced three major periods, including the statistical reporting period (before the reform and opening up), sampling period (after 1994), and reform period of the survey (2006). The main reason for the changes of road transport volume survey and statistical methods were that economic restructuring led to the changes in social structure, interest adjustments, and change in ideas. As a result, road transport volume survey and statistical methods could not meet the needs of social development at that stage and the data could not reflect the actual situation of the transportation market.

The survey of passenger vehicles adopted a sampling method on the national highway traffic special investigation program. The sampling unit is a single vehicle.

In the sampling design all passenger vehicles of the country belong to the scope of investigation were as a whole and the passenger vehicles of provinces are the sub-population samplings. According to the characteristics of the transport vehicle, province-wide passenger vehicles included in the survey were first divided into inter-provincial bus line classes, classes within the province bus lines, and other commercial buses, rental buses, public transportation vehicles, and other non-commercial passenger were three separate samples. If some vehicles refused to participate in the survey, samples were added on the scene in time until the total number of vehicles accepting the survey met the minimum sample size requirements. If the county road passenger transport volume survey is also carried out in accordance with this approach, it is difficult to implement for the large sample size and to follow sampling vehicle's tracks in a survey cycle.

The Proposal of Investigation and Statistical Methods

Research ideas The survey object studied in the paper is the operation of passenger transport volume on county road network, including the local vehicles and other vehicles from outside.

Combine the actual situation of lack of the statisticians and funds, and passenger transport volume can be classified in accordance with the vehicle operating characteristics, and the simple and easy investigation of statistical methods are put forward. The methods are as follows.

(1) The cars that operate in passenger terminal：Using statistical reports method or the terminal concentration survey method, we can get the line operational data that represent days and months. According to these data, we can calculate the annual transport volume directly.

(2) The cars that do not operate in passenger terminal: By the passenger car-survey method under sample survey method, we can calculate their volumes during the investigative phase, then expand it to an annual transport volume combining the seasonal fluctuation coefficient.

(3) A taxi or charter coaches: Temporary road OD traffic survey (the proportion of total traffic volume, the actual travel path, and transport distance) combines with the village-home survey (self-operated chartered transport volume) and concentration of a sample survey of business establishments (relatively fixed, concentrated off points; the passenger number per day or week, the average number of passengers, travel routes, and transport distance), specifically based on "focus on a sample survey of business establishments" data, follow the survey statistical methods based on car registration database we to conduct traffic surveys, the other two methods of investigation provide the correction factor and the (projected) rate coefficients.

The operation of county road bus passenger transport volume can be obtained by directly adding to the above-mentioned three types of vehicles road passenger transport volumes.

To further simplify the survey and statistical methods, we conduct transport volume surveys of the cars operating in passenger terminals each year. While the other two transport volume types conduct transport volume surveys in the base year, we only calculate the transport volume projection coefficient in ordinary years. At this point, it is so-called "the statistical projections method of sub-model traffic projections combining with road network traffic (fluctuations) projected." A projection cycle is generally five years, including a base year and four ordinary years.

Statistical Calculation Model.

The traffic projections calculation of the bus entering passenger station.

(1) Projection ideas

For this kind of bus transport volume, we can choose representative business months or weeks of the year to collect business data, and the unit is the pick-up rate.

When calculating, first of all figure out the attendance of each departure. The average transport volume of each route multiplied by the average attendance is the average volume of passenger transport. The volume multiplied by the departure times of each route is the completive passenger transport volume of the corresponding route. The completive passenger transport volume multiplied by the transport distance is the turnover of passenger traffic. This value multiplied by 365 is the annual passenger transport volume. Finally, the completive annual passenger transport volume of all the operation lines at county will be added to obtain the total passenger transport volume.

(2) Projection model

The equation of the average attendance b :

$$b = \frac{\sum_{i=1}^{n} a_i}{n}$$

Where,a_i: attendance of each trip, and n: times of trip of each route.

The formula of route passenger volume q_l :

$$q_l = m \times b \times n$$

Where,m: route average capacity,b: route average attendance, and n: times of trip of each route.

The formula of route revenue passenger turnover Q_l :

$$Q_l = q_l \times d$$

Where,d: carrying distance

The equation of total passenger volume Q_{11} :

$$Q_{11} = \sum_{i=1}^{t} 365 \times q_l$$

Where,t: Number of operation routes in the county.

The equation of annual route passenger turnover Q_{12} :

$$Q_{12} = \sum_{i=1}^{t} 365 \times Q_l$$

Non-stop operation of fixed-line transport volume projections (including non-stop operation of fixed-line bus and rental business passenger car) . The investigation method is the same as 1994 and 2006 ministerial pilot programs, and the data are selected by the local vehicle household registration database. When ensuring the 95% confidence level, the limits of the relative error of the main objective of the investigation (including the county's passenger traffic volume and passenger transportation volume) is 5% or less. The calculation of the total number of regional sample vehicles, the distribution of each sub-district, and the distribution

of each sub-layer, etc. will not be repeated. Their specific steps and calculation method are the same as the ministerial programs of 1994.

Road operation passenger transport volume projections in ordinary year.
(1) Calculation ideas

First of all, calculate the passenger transport volume of the fixed-line (first type) in the base year and ordinary years. For the other two types of models, use their own methods to calculate the annual "observing" transport volume of the base year (the missing data are mainly from the illegal operators and some nonlocal vehicles in the survey area). Then, (the road network traffic is known) calculate (mainly for the latter two models) the road network transport volume of the base year with travel proportional coefficient of operational bus, the average occupancy rate and so on. Finally (for the latter two models) according to the change of the traffic volume, the number of buses, and the operational coefficient of carrying passengers, adjust the data based on the value of base year.

(2) Calculation model

The formula of road-net network passenger volume at the survey period Q_{31} :

$$Q_{31} = T \times K_i \times \delta_i$$

Where,*T:* traffic volume matrix at the survey period,K_i: Average tons (seat) matrix, and δ_i :Average actual load rate matrix.

The equation of the road-net revenue passenger turnover volume at the survey period Q_{32} :

$$Q_{32} = Q_{31} \times L_i$$

Where,L_i:The average transport distance matrix

Case Analysis

The case study is based on the operating data of the county road passenger station in Heilongjiang province in July 2008.Actual traffic volume of one day in July was investigated to validate the computational method of road passenger transport volume of the county operation vehicles.

The transport volume of the line bus getting into the station The paper made one county in Heilongjiang the experimental area and conducted the survey of the passenger station. We had collected the operating data from January 25, 2008, to May 26, 2008. The data included the number of passenger routes, departure frequency of each route, the time, origin, and destination of each route, the capacity for each frequency and each route, the actual carrying capacity, and other information, some of the original information is shown in the Table 1.

According to the model in this paper, the result based on the monthly attendance is shown in Table 2.

Table 1. Calculated passenger volume and turnover volume based on the route attendance.

Date	volume(people)	RPK(people*kilometre)
5.10	1251.3	39044.1
5.11	1143.8	34056.0
5.12	1185.7	37104.1
5.13	1174.6	34382.6
5.14	1167.5	38655.8
5.15	1060.7	32011.9
5.16	1269.3	39254.8
5.17	1282.4	37186.7
5.18	959.7	28465.6
5.19	1179.8	36346.7
5.20	1129.1	35401.2
5.21	1017.7	31721.9
5.22	978.8	32559.1
5.23	1162.6	35272.8
5.24	1244.2	38700.8
5.25	1143.6	33542.8

Table 2. Transportation volume of the bus getting into the station.

Monthly volume (people)	43184	Monthly RPK(people*kilometre)	1285418
Annual volume (people)	518209	Annual RPK(people*kilometre)	15425025

The transport volume of the operation car getting out of the station The paper made one county in Heilongjiang the experimental area and conducted the survey of the passenger station. The investigation was conducted in July 2008. For the survey, we selected 13 passenger routes in which bus lines don't enter into the terminal, while the surveyors chose one or two bus lines to follow and survey. For taxis, we selected several places where taxis go such as the railway station and trade market. According to the statistic inference model in this paper, the results of the transport volume of the operation car is shown in Table 3.

To compare with the survey from the local transportation department in 2008 (Communications Dept. 2008), in that the sampling survey method was adopted, and the mean error is 6.7%. And also, there were lower costs in the survey of the paper, which means the method in the paper is valid and effective for statistics of highway passenger transport volume in the county level.

Table 3. Transportation volume of the bus getting out of the station.

Monthly volume(people)	452	Monthly RPK(people*kilometre)	7232
Annual volme(people)	164980	Annual RPK(people*kilometre)	2639680

Table 4. passenger transport volume of the county road.

Annual volme(people)	683189	Annual RPK(people*kilometre)	18064705

Passenger transportation volume of the county road network The data from Table 1 plus the data from Table 2 was obtained from the passenger transport volume of the county road network. The result isshown in table 4.

Conclusions

A survey was conducted on different operating style of buses to classify the passenger transport volume. The purpose of this survey aimed at the county-level passenger transport volume statistics with enough convenience and high speed, and the calculated result was similar to the actual statistics.

However, in the county, the travel times of passengers are concentrative and their trips are mostly with the purpose of hurrying, shopping, or going to work, so the passenger transport volume has the characteristics of instability, prominent seasonality, change with the time interval, single–direction, and so on. In order to improve statistical accuracy, we can carry out the survey in the summer, winter or conduct more surveys to amend the key projection parameters, such as the actual number of passengers on board.

References

Chen, Y. (2005). "Study of freeway transportation volume [J]." *China Journal of Highway and Transport*, 18(2):94-98.

Comprehensive planning department of ministry of communications. Passenger entry and exit survey analysis report in May 2007 [R]. *Beijing: Comprehensive planning department of ministry of communications*, 2007.

Cen, Y., Zou, L., and Yang, Z. (2000). "The application of sampling survey technology in transportation statistics." *Journal of Highway and Transportation Research and Development*, Vol. 117(12):67-71.

Bureau of Transportation Statistics. (2003). "U. S. department of transportation. transportation statistics annual report [R]." Washington D.C.: *Bureau of Transportation Statistics*.

Yang, M., and Qin, H. (2009). "A statistic comparison between highway transportation amount in China and that in America [J]." *Transportation Enterprise Management*, 2009, 2:69-70.

Heilongjiang province highway OD survey report [M]. (2008). *Conmmunications Department of Heilongjiang province*.

Estimating Dynamic Transport Population for Official Statistics Based on GPS/GSM

Yinyi Ma[1], Jan van Dalen[1], Chris de Blois[2] and Jo van Nuen[1]

[1] Erasmus University Rotterdam, the Netherlands; Email: YMa@rsm.nl
[2] Statistics Netherlands

Abstract

Transport and traffic data collection methods have developed along two main directions. One is the off-line method using pen-and-paper surveys or face-to-face interviews, and the other is the on-line method using internet surveys or XML messages from companies' administrations. The transport and traffic statistics based on these data are rather static representing an average, macro-view of transportation. In order to further enhance transport and traffic statistics, linking traffic to social economic fields, real-time collected data based on GPS/GSM should be available. Marked advantages of GPS and GSM data collection are that transport and traffic data are captured automatically at highly frequent rates, which provide dynamic information and offers opportunities to reduce administrative burdens. And also with the help of GPS information, the individual vehicles can be identified by connecting with other data source. However, though GPS/GSM data collection sounds a promising technology, its adoption is seriously hampered by the fact that few vehicles are equipped with GPS transponders and that not all drivers use a GSM. GPS/GSM-collected data thus constitute only a limited part of the whole transport and traffic data for a delineated area and specific time slots. In this paper, we aim to validate the usage of GPS information for traffic statistics. Traffic density is theoretically applied to integrate traffic engineering, social economy and statistics methods. Then with the limited number of GPS/GSM vehicles, we take their advantages of doing the timing and identification issues for carrying out density estimation. Further, statistic method of Horvitz-Thompson is used to up-scale the dynamic density to the whole network and time period for official statistics.
Key Words: dynamic transport population estimation; GPS/GSM sample data; loop detector data; Horvitz-Thompson estimator

Introduction

Transport data collection methods for travel survey have developed along two main directions. One is the off-line method using pen-and-paper surveys or face-to-face interviews, and the other is the on-line method using internet surveys or XML messages from companies' administrations. Both survey methods are essentially applied to gather information about demographic, socioeconomic, and trip-making characteristics of individuals and households, and about the movement of motor vehicles and cargos. Normally, these data are collected during a single day or week per year for each household or vehicle in the sample. For the traffic data collection, detectors or fixed sensors along the motorways generate loop data about traffic flows. Once the data is obtained, aggregation is carried out to produce traffic statistics. Consequently, the transport and traffic statistics which are based on these data are static and represent an average, macro-view of transportation. With these kinds of data, the detailed information in the traffic and transport processes can not be represented, and the heterogeneity of travel behaviour can not be test.

Current data collection practice is subject to serious challenges, such as the administrative burden, data accuracy, and in the case of loop detectors, the sensitivity for errors and malfunctioning. At the same time, the last decade has seen a massive increase of traffic-stated signals generated by GPS (Global Positioning System) and GSM (Global System for Mobile communications). The question arises whether and how this real-time captured electronic data could serve as alternative, probably complementary, data to construct traffic statistics. This paper explores the combined use of GPS data from transport companies and GSM data from telecommunication companies for developing transport and traffic statistics. A marked practical advantage of GPS and GSM data collection is that transport and traffic data are captured automatically at highly frequent rates. This provides dynamic and real-time information and offers opportunities to reduce administrative burden. And also with the help of GPS information, the individual vehicles can be identified by connecting with other data source. Since GPS devices are installed inside vehicles, they can capture instantaneous, real-time and accurate information, including position, direction, instant speed and time. These information can represent real-time traffic states in temporal-spatial dimensions, give relatively precise transport information and exact travel behaviour.

Though GPS/GSM-based statistics sounds promising, its adoption is seriously hampered by the fact that relatively few vehicles are equipped with GPS transponders and that not all drivers carry a GSM. Recent estimates of the penetration rate of usable GPS and GSM devices are around 2% of the traffic population. This low penetration rate is not sufficient to provide an exhaustive coverage of the transportation network. GPS/GSM-collected data constitute only a limited part of the whole transport and traffic data for a delineated area and specific time slots. A substantial continuously varying number of vehicles thus go unobserved. Thus, combing loop detector data and GPS data could be a good approach currently.

The aim of this paper is to validate the usage of GPS information to improve traffic statistics. Following this, the research questions can be formulated as: (1) What is the proper approach to integrate loop detector data, GPS data and statistical knowledge? (2) What is the model to represent dynamic traffic characteristics for official statistics with GPS information?

The paper proceeds as follows. In section 2, characteristics of traffic statistics are described. The comparison of the mechanisms, advantages and disadvantages of loop detector data and GPS data is carried out in section 3. Following this, beginning with traffic flow theory, the connection among loop detector data, GPS data and statistical knowledge is represented. Then in section 5, modelling dynamic traffic population is carried out intergraded both traffic and statistical knowledge. Finally, the conclusion and discussion in section 6 finalize the paper.

Traffic Statistics Characteristics

Statistics is the scientific discipline that aims to describe population characteristics on available data. Normally, national statistics office is responsible to produce official statistics. Traffic statistics, as one of the official statistics, consists of highly aggregated data to represent the traffic performance of a road network in a certain geographical area or country. It includes traffic counts (vehicles/year), total travel distance, transport modes and so on. Most of the statistics offices get data from authorities responsible for the road network. In the case of the Netherlands, for instance, central bureau of statistics (CBS) obtains data from the ministry of transport and water work.

However, the drawbacks of this traffic statistics are showing up. First, current traffic statistics are monthly/yearly-based, which means that they are either static or homogenous. They only show the state-value, instead of process-value. Thus it is impossible to represent the dynamic traffic states. Second, traffic never exists alone and totally independently. It always connects with many other activities, which are derivated during the travel periods, such as road maintenance, safety, emission, trade and investment. Unfortunately, although this relation is quite important from the social economic perspective, there is no connection between these topics and the current traffic statistics, since traffic statistics is aggregated in a high level.

Therefore, in order to validate the relation, reconstructing traffic statistics in an innovative approach need to be carried out, which should be suitable to link with the social economic aspects. Precisely, the updated traffic statistics could represent the detailed information, for instance, the road usage, travel behaviour heterogeneity and travel routes. Meeting all these requirements needs a proper data collection method, in order to either reduce the administrative burden in statistics office, or update the traffic statistics to connect the social economic perspectives.

Regarding these requirements, the following section has the further insight into traffic data and data collection method, and finds the opportunity for improving traffic statistics.

Comparison between Loop Detector Data and GPS Data

Traffic is taken to refer to vehicles that drive on the roads of a certain network. In order to understand the traffic performance, and to support traffic policy measures, traffic date is collected to identify the traffic states through certain equipments. Apart from traditional survey data, more direct, real time measures of traffic states are obtained through loop detectors and GPS/GSM. This section compares the latter two observation methods in order to assess their relative use for making traffic statistics. Currently, loop detectors are widely. And with the development of telecommunication, GPS/GSM becomes promising to obtain traffic data.

Loop Detector Data. The most popular approach at present to capture traffic data on the roads is through inductive loop detectors embedded in the pavement. The data generated by loop detectors provide information about traffic flow of vehicles counts per hour and time-mean speed of vehicles passing the loop detectors during a time interval.

Working mechanism. Technically, the loop detector system includes one or more turns of insulated loop wire wound in a shallow slot sawed in the pavement of a road or highway. The loop system forms a tuned electrical circuit in which the loop wire is the inductive element (Ki and Baik, 2006). When a vehicle passes over the loop or stops within the loop, the inductance of the loop is decreased, which is observed by a sensor.

The main detector types are the single-loop detector and the double loop detector. Both types have the same mechanism to count the traffic volumes per time period, but have different ways of calculating the speed of passing vehicles. For the single-loop detector, the speed $v_{i,t}$ of a vehicle i passing during time interval t is calculated as the effective length, vehicle length(l_v) plus detector length (l_d), divided by the time elapsed between the switching on (t_1) and off (t_2) of the detector (Ki and Baik, 2006):

$$v = \frac{l_{v,t} + l_d}{t_2 - t_1} \tag{1}$$

If n_t vehicles pass the loop detector during time interval t, then time-mean speed (v_t) generated by loop detector can be formulated as:

$$v_t = \frac{1}{n_t} \sum_{i=1}^{n_t} v_{1,t} \tag{2}$$

Since the loop detector can not measure the length of vehicles, a mean vehicle length ($\bar{l}_v$) is assumed, which may lead to inaccurate measurements of speed. For instance,

longer trucks will produce underestimated speed, and shorter passenger cars will produce overestimated speeds (Sun and Ritchie, 1999).

Moreover, if the traffic composition systematically various between the hours of the day, the days of the week, or road segments, then these inaccuracies may give rise to biased estimation of traffic speed in the population.

For the double loop system, speed is measured as the distance (d) between detectors divided by the difference between the time when the first detector turns on (t_1) and the second detector turns on (t_3):

$$v = \frac{d}{t_3 - t_1} \tag{3}$$

This approach avoids the requirement of knowing the length of vehicles, and consequently is more accurate.

To further improve the accuracy of the system, the Gordon et al. (1996) recommend the use of turn-off time in additions to turn-on times in the following equation:

$$v = \frac{1}{2}\left(\frac{d}{t_3 - t_1} + \frac{d}{t_4 - t_2}\right) \tag{4}$$

where t_2 is the time when the first detector turns off and t4 is the time when the second detector turns off.

Advantages and disadvantages of loop detector data. In order to measure traffic speed in a road network, it is necessary to have a substantial number of loop detectors in place. In the Netherlands, for instance, loop detectors are set in the pavement every 500 meters. They are generally considered reliable with a well defined zone of detection (ITS Decision, 2007). Obviously there are several benefits of using loop detectors. First, they are relatively easy and convenient means to obtain data with the simple calculation methods. Second, when properly installed and maintained, loop detectors perform well in both high and low traffic volumes, and in different weather conditions. Third, since they are in the pavement, the possibility of getting damaged is low, and thus they have a long life cycle. Fourth, they meet even the most stringent vehicle flow error specifications required by Intelligent Transportation System application. Lastly, even with crosstalk problems and a high proportion of lane changes, loop detectors have overcounts of only 0.8% and 0.4% (ITS Decision, 2007).

However, there are also disadvantages of using detectors to measure speed. First, traffic speed monitoring and vehicle counting using induction loops involve relatively high installation and maintenance costs. Secondly, loop data are prone to malfunctioning detectors and transient errors (Jain and Coifman, 2005), even if a double loop is more accurate than a single loop for measuring speed. Errors may occur because of the scanning time of the detector card, variations in loop spacing, irregular driving vehicles, and presence of multiple vehicles in the detection zone (Ki

and Baik, 2006), which result in noisy measures. Several solutions have been proposed to cope with these shortcomings. Jain and Coifman (2005) advance a data cleansing tool to identify and exclude malfunctioning detectors and transient errors. Coifman and Krishnamurthy (2007) present the extended Kalman-filtering method to match vehicle measurements between detector stations in order to provide information about the conditions over the link between the detectors. From the perspective of developing traffic statistics, the main drawback of loop detector data is its inability to identify vehicles. The availability of identified traffic data would allow them to collect information about the roads traveled, the propose of road usage, or emission. Global Positioning System (GPS) and Global System for Mobile communications (GSM) should be good approaches to compensate the identification issue.

GPS/GSM Data. With the development of telecommunication and wide usage, the Global Positioning System (GPS) and Global System for Mobile communications (GSM) are promising to be used in the traffic field, tracking and tracing the vehicles. Compared with the fixed loop detectors, GPS and GSM as the in-vehicle devices can capture the dynamic real-time information and even be identified.

Working mechanism. GPS is a spaced-based global navigation satellite system, which consists of three segments: space composed of 24 to 32 satellites, control and user. Currently, the USA controls the military GPS system, while Europe is developing a civil system named Galileo. GPS provides reliable positioning, navigation, and timing services to worldwide users on a continuous basis in all weather, day and night. For traffic, GPS can offer the real time data about location (latitude, longitude and elevation), time in minimum time intervals of 0.3 seconds, speed and number of satellites to identify the source.

Speed measurement with GPS is based on a series of "trackpoints" of position estimates at regular time intervals. The difference between the known satellites carrier frequency and the frequency determined at the receiver is known as a Doppler shift, which is directly proportional to velocity of the receiver along the direction to the satellite, regardless of the distance to this satellite. The vehicle speed thus obtained is very accurate. However, the measurement accuracy is not constant (Chalko, 2002), but depends on the number of tracked satellites and on their geometrical distribution above the horizon. If the number of measured satellites is 3 or 4, then the data is regarded as reliable.

Advantages and disadvantages of GPS data. Compared with loop detectors, GPS/GSM offers the identified information of each vehicle. This information can be used to have the vehicles along the road network, but also to examine variation in usage during certain time periods. By linking GPS/GSM data with vehicle characteristics estimates of emission can be obtained.

The longer the measuring period, the more frequent are the speed samples and the better the accuracy of speed estimates. Additionally, since the time interval between GPS signals is small, the detailed performance can be obtained.

However, at present not all the vehicles have GPS/GSM devices to guide the routes. In Greece, only 3%-5% (Wang and Papageorgiou, 2005) of vehicles have GPS-devices installed. Hence, the real-time data resource is limited, but still gets growing. In addition, the cost of bilateral communication is so expensive that drivers are unwilling to pay for the service. Some companies are trying to solve this cost issue by realizing bilateral communication through the internet rather than the GPS signal channel. But it is still in the research and test stage.

Although the number of GPS-installed vehicles is limited, the potential of the technology is sufficiently promising to test its use for improving traffic statistics.

Theoretical Perspective of Dynamic Traffic Estimation

The development of traffic statistics based on captured data draws from the fields of engineering and the economic and social sciences. In this section the theory from traffic aspect is discussed to build a path for traffic statistics.

The interaction between vehicles, drivers and infrastructures is the subject of traffic flow theory, which aims at understanding and developing an optimal road network with efficient movement of traffic and minimal traffic congestion. Relevant concepts in traffic flow theory are distance travelled, traffic volume, traffic density, time-mean and space-mean speeds. Understanding traffic flow characteristics is essential for planning, design and operation of transport system, and further to make traffic statistics.

Moreover, the interests of social scientists and economist are concerned with issues like the intensity of cargo transport and business vehicles in the road networks. Other topics of interest are congestion, environmental issues, safety and substantiality.

Relation amongst Flow, Density and Speed.Classical traffic flow theory defines the fundamental relation amongst flow, density and average speed: flow is density * speed (Hoogendoorn, 2007). The higher the density, the more intense the flow, assuming the speed constant. Likewise, a higher average speed implies a more intense flow. Density is defined as the average number of vehicles in a road segment during a time interval; speed as the average speed (in km/h) of these vehicles during the interval; and flow as the number of vehicles per road segment and time interval. This formulation carries the implicit assumption that traffic is homogeneous and stationary within the road segment and the time interval.

For official statistics, it is important to link the technical aspects of traffic measurement with the requirements from social and economic objectives. Then availability of identified vehicle data seems necessary to realize the connection.

Especially, the measurement of traffic density perhaps a major role in traffic flow theory and economic theory and policy. We therefore take density to demonstrate the use of GPS data.

Density Estimation. Traditionally, traffic density is defined as the number of vehicles presents in a certain part of a roadway lane during a certain period of time, and is expressed in vehicles/km.

Using loop detector data, density measurement models have been proposed by models Coifman (2003), Gazis and Knapp (1971), and Nahi and Trivedi (1973). In addition, Daganzo (1994) introduces the concept of Cell Transmission Model (CTM) to detect changes in density on the motorway with an open-loop density estimator. The CTM is extended to the load network level in Daganzo (1995). For the purpose of density estimation, Munoz et al. (2005) apply the CTM derived switching-mode model as a freeway traffic density estimator to determine the controllability and observability properties of highway.

With the development of vehicle telematics, information about individual vehicle could be a potentially valuable source to compensate for the deficiencies of point sensors. In Qiu et al. (2010), 5% of GPS data are applied to estimate density, which provide a moving window over distance for speed and position, combined with loop data giving traffic speed and flow. Di (2010) implements a hybrid extended Kalman filter to integrate the approximated Markov Compartment with detector and GPS measurements.

The mentioned studies are important sources of traffic density estimation, but none connects density to economic statistics. Following, we discuss a method of estimating dynamic traffic population, connecting traffic theory and statistical methods, and making a path to social science applications.

Classifed Estimating Dynamic Traffic Population

In order to represent traffic statistics in a better way that it can connect with the social economic issues, the information from both road network and the characteristics of vehicles is necessary. In the road-network perspective, road usage is paid attention on the number of vehicles in the specific length of road. It may link to the life cycle of road for the road maintenance issue. Then in the vehicles aspect, knowing the vehicle composition and characteristics of vehicles can have a linkage to the emission and other social economic issues. At the same time, both perspectives need identify the vehicles and get the detailed information.

Road usage is demonstrated by occupancy rate (β). A vehicle passing over a loop temporarily occupies it, approximately from the moment the front of the car is at the beginning of the loop until its rear is at the end of loop (Hoogendoorn, 2007). Note this individual occupancy period as b_i. For a period T, in which n vehicles pass, the occupancy rate (β) is defined as:

$$\beta = \frac{1}{T}\sum_{i=1}^{n} b_i \tag{5}$$

For each vehicle, the vehicle length is L_i and speed is v_i. Then $b_i = (L_i + L_{loop})/v_i$. Suppose vehicles are approximately of the same length, then b_i is proportional to $1/v_i$, and b can be written as:

$$\beta = \frac{1}{T}\sum_{i=1}^{n}\frac{L_i + L_{loop}}{v_i} = \frac{L_i + L_{loop}}{T}\sum_{i=1}^{n}\frac{1}{v_i} = (L_i + L_{loop})\frac{n}{T}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{v_i} = (L_i + L_{loop})\frac{q}{u_{mean}} = (L_i + L_{loop})k \tag{6}$$

It appears that occupancy rate (β) is proportional to the (calculated local) density (k) if vehicle lengths (l_v) are equal (Hoogendoorn, 2007). However, if a mix of passenger cars and trucks is present, then the meaning of b is less obvious. But combined with GPS information, the vehicle length can be obtained through identifying the vehicles. The calculation from this aspect will be improved. Then density needs to be updated as well to represent the dynamic changing.

In this section, the methodology of estimating dynamic traffic population is given, which combines both traffic theory and statistics method. Through traffic theory, the classified traffic densities (according to different types of vehicles) in one segment of motorway can be obtained. Then up-scaling in the temporal-spatial dimensions based on statistics method is carried out to estimate dynamic traffic population.

Traffic Density Estimation. In classical flow theory, traffic density (k) is loosely defined as the number of vehicles in a road segment divided by the length of the segment. More precisely, let us consider a road segment that begins and ends with two detectors at locations x_i and x_{i+1}, between which there are an on-ramp and an off-ramp also with loop detectors $x_{on(i)}$ and $x_{off(i)}$. At each time interval t, the system records the cumulative count of vehicles at each detector, $C(x_i,t,t_0)$, since initiation at time interval t_0. The cumulative count of vehicles in the road segment between x_i and x_{i+1} defined as the difference between the cumulative counts at the detectors.

$$N(x_i, x_{i+1}, t, t_0) = C(x_{i+1}, t, t_0) - C(x_i, t, t_0) \tag{7}$$

If we denote the distance between the two detectors as $L(x_i, x_{i+1})$ and the number of lanes involved as $LN(x_i, x_{i+1})$, then the traffic density is defined as:

$$k(x_i, x_{i+1}, t, t_0) = \frac{N(x_i, x_{i+1}, t, t_0)}{L(x_i, x_{i+1}) \times LN(x_i, x_{i+1})} \tag{8}$$

Here, GPS vehicles are applied to realize the timing issue. When GPS vehicle arrives at loop detector x_{i+1} at a certain point time, and the point time is involved in the time interval $t(x_{i+1})$ of loop detector, then traffic counts changing in the certain time period when GPS vehicle pass both continuous loop detectors can be calculated. In this small part of highway section, there are three parts: main lane, on-ramp and off-ramp. For each part, the formulations about vehicle counts with GPS timing are given.

- Main lane vehicle counts.The main lane vehicle counts changing can consider the traffic counts from loop detector x_i only. When GPS arrives at loop detector x_i, at that time the time interval is $t_{GPS}(x_i)$ When GPS arrives at loop detector x_{i+1}, at that time the time interval is $t_{GPS}(x_{i+1})$. Thus from time interval t0 to time interval $t_{PGS}(x_i)$, there is cumulative traffic counts from loop detector x_i, denoted as $C[x_i, t_{GPS}(x_i), t_0]$. Similarly, the loop detector xi+1 generates cumulative traffic counts, denoted as $C[x_i, t_{GPS}(x_{i+1}), t_0$. Hence, the vehicle counts changing in the main lane is formulated as:

$$N_{mainlane} = C[x_i, t_{GPS}(x_{i+1}), t_0] - C[x_i, t_{GPS}(x_i), t_0] \tag{9}$$

- On-ramp vehicle counts.On-ramp vehicle counts are generated by on-ramp loop detector, with cumulative counts ($C_{on\text{-}ramp}$). GPS vehicles also play a role to do the time issue between loop detector x_i and x_{i+1}. Thus the traffic counts changing on ramp can be formulated as:

$$N_{on-ramp} = C_{on-ramp}[t_{GPS}(x_{i+1}), t_0] - C_{on-ramp}[t_{GPS}(x_i), t_0] \tag{10}$$

- Off-ramp vehicle counts.The algorithm for off-ramp vehicle counts changing is quite similar to on-ramp one.

$$N_{off-ramp} = C_{off-ramp}[t_{GPS}(x_i), t_0] - C_{on-ramp}[t_{GPS}(x_{i+1}), t_0] \tag{11}$$

Once the separated traffic counts in the three parts of segment are formulated, the total density between the both loop detectors in the highway segment is traffic counts divided by the segment length (L) times number of lanes (LN). So the total density is:

$$k(x_i, x_{i+1}, t_{GPS}, t_0) = \frac{N(x, x_{i+1}, t_{GPS}, t_0)}{L(x_i, x_{i+1}) \times LN(x_i, x_{i+1})} = \frac{N_{mainlane} + N_{on\text{-}ramp} + N_{off\text{-}ramp}}{L(x_i, x_{i+1}) \times LN(x_i, x_{i+1})} \tag{12}$$

Furthermore, GPS vehicle is not only used to do the timing, but vehicle characteristics can be identified, such as type of vehicles and length of vehicles. Then the density can be denoted by $k_{truck}(x_i, x_{i+1}, t_0, t)$ if the GPS vehicle is truck, $k_{pc}(x_i, x_{i+1}, t_0, t)$ if it is private car, $k_{bc}(x_i, x_{i+1}, t_0, t)$ if it is business car, and so on. If there are more GPS vehicles in the lane ln of segment i, such as NT trucks with GPS, NPC private cars with GPS, considering the average calculated density (k) from different kinds of vehicles, the final estimated value is formulated as:

$$k_{i,\ln}(x_i, x_{i+1}, t_0, t) = \frac{\sum_{nt=1}^{NT} K_{truck(nt)}(x_i, x_{i+1}, t_0, t) + \sum_{npc=1}^{NPC} K_{pc(npc)}(x_i, x_{i+1}, t_0, t) + \ldots}{NT + NPC + \ldots} \tag{13}$$

Once the dynamic density estimation and the length of vehicle are available, occupancy rate can be obtained in a more accurate way. It can better represent road usage with life-cycle and maintenance of pavement. But the fleet composition is not easy to get, which really depends on the increasing number of GPS vehicles. The more GPS vehicles, the clearer the types of vehicles.

Dynamic Traffic Population Estimation Based on Statistical Method. Having defined the density in a lane of a motorway segment with the help of GPS-installed

vehicles, the next step is to up-scale the sample results to estimate the desired population parameters. The Horvitz-Thompson estimator is a general estimator for a population total, which can be used for any probability sampling plan (Borchers et al., 1998). The formulation of the Horvitz-Thompson estimator is formulated as:

$$\tau_{\pi} = \sum_{i=1}^{v} \frac{y_i}{\pi_i} \tag{14}$$

where, τ : Horvitz-Thompson estimator;
π : the probability;
y: sample;
v: number of sample.

In order to estimate the network density, the up-scaling issue can be based on the detected flow which can represent the traffic states in the network. The proposed density ($k_{i,l,t}$) in one lane (l) of one segment (i) can be regarded as a referenced value, and then through the differences between the flow ($q_{i,t}/q_{1,t}$) as a weight, the up-scaling can be carried out. If the network has S segments and each segment has L lanes, the network density in the whole time period T is formulated as:

$$k_{network} = \sum_{t}^{T} \sum_{i=1}^{S} \sum_{l=1}^{L} \frac{q_{i,t}}{q_{1,t}} k_{i,l,t}(x_i, x_{i+1}, t_0, t) \tag{15}$$

Summing Up.This proposed method of estimating dynamic traffic population is combined with traffic theory and statistical method. According to density definition, the density can be estimated with the help of specific GPS vehicles. Since GPS information can identify vehicles, densities can be classified according to different types of vehicles. Then Horvitz-Thompson estimator is applied to scale up the traffic counts to the population level in the temporal-spatial dimensions.

Conclusions and Discussion

This paper represents the statistic method to estimate dynamic traffic characteristics for updating official statistics, linking with social economic perspective. From data collection aspect, current loop data source and GPS/GSM data are compared to demonstrate the advantages and disadvantages, finding the opportunity for GPS/GSM data to improve traffic statistics. Then from the theoretical perspective, traffic density is applied to be the integration of traffic engineering, social economy and statistics. Furthermore, dynamic density estimation is modeled based on loop detectors and GPS/GSM information. GPS vehicles are applied, in the data-collection aspect, to do the timing issue and identify vehicles. Finally, based on Horvitz-Thompson Estimator, dynamic traffic population in temporal and special dimensions is estimated. With this method, official traffic statistics can be more accuracy than the current one. So it is valid to apply GPS information to improve traffic statistics. More importantly, for the further research, the model can be extended to link with social economic perspectives, taking emission for instance, since identification of vehicles is possible based on GPS/GSM information.

Reference

Borchers, D., Buckland, S., Goedhart, P., Clarke, E. and Hedley, S. (1998), "Horvitz-Thompson estimators for double-platform line transect surveys", *Biometrics*,pp. 1221–1237.

Chalko, D. (2002). "High accuracy speed measurement using GPS (Global Positioning System) ", *Scientific Engineering Research P/L* .

Coifman, B. (2003), 'Estimating density and lane inflow on a freeway segment', *Transportation Research* Part A 37(8), 689–701.

Coifman, B. and Krishnamurthy, S. (2007), "Vehicle reidentification and travel time measurement across freeway junctions using the existing detector infrastructure", *Transportation Research Part C*,15(3), 135–153.

Daganzo, C. (1994), "The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory", *Transportation Research* Part B: Methodological 28(4), 269–287.

Daganzo, C. (1995), "The cell transmission model, part II: Network traffic", *Transportation Research* Part B 29(2), 79–93.

Di, Xuan. Liu, H. X. (2010), "A Hybrid Extended Kalman Filtering Approach for Traffic Density Estimation along Signalized Arterials Using GPS data", in *Transportation Research Board 2010 Annual Meeting CD-ROM*.

Gazis, D. and Knapp, C. (1971), "On-line estimation of traffic densities from time-series of flow and speed data", *Transportation Science* 5(3), 283.

Gordon, R., Reiss, R., Haenel, H., Case, E., French, R., Mohaddes, A. and Wolcott, R. (1996), "*Traffic control systems handbook*", Publication Number FHWA-SA-95-032 .

Hoogendoorn, S. (2007), "Traffic Flow Theory and Simulation". ITS Decision (2007). http://www.calccit.org/itsdecision/serv and tech/Traffic Surveillance/road-based/in-road/loop summary.html.

Jain, M. and Coifman, B. (2005), "Improved speed estimates from freeway traffic detectors", *Journal of Transportation Engineering* 131, 483.

Ki, Y. and Baik, D. (2006), "Model for accurate speed measurement using double-loop detectors", *IEEE Transactions on Vehicular Technology* 55(4), 1094–1101.

Munoz, L., Sun, X., Horowitz, R. and Alvarez, L. (2005), "Traffic density estimation with the cell transmission model", in 'Proceedings of the *American Control Conference*', Vol. 5, pp. 3750–3755.

Nahi, N. and Trivedi, A. (1973), 'Recursive estimation of traffic variables: Section density and average speed', *Transportation Science* 7(3), 269.

Qiu, T. Z., Lu, X.-Y. and Chow, A. H. (2010), "Freeway Traffic Density Estimation with Loop Detector and Probe Vehicle Data", in '*Transportation Research Board 2010 Annual Meeting CD-ROM*'.

Sun, C. and Ritchie, S. (1999), "Individual vehicle speed estimation using single loop inductive waveforms", UC Berkeley: California Partners for Advanced Transit and Highways (PATH). Retrieved from: http://www. escholarship. org/uc/item/2zz4b12p .

Wang, Y. and Papageorgiou, M. (2005), "Real-time freeway traffic state estimation based on extended Kalman filter: a general approach", *Transportation Research Part B*, 39(2), 141–167.

Time Contours: Using IC Toll Information to Describe Public Transport Networks Accessibility

Lianxiong Gao[1,2], Jianping Wu[1] and Rui Liu[3]

[1] School of Computer, Beijing University of Posts & Telecommunications, Beijing 100876, China, E-mail: gaolianxiong@gmail.com
[2] School of Physics & Electronics, Yunnan Nationalities University, Kunming 650031, China
[3] School of Resources-Environment and Geoscience, Yunnan University, Kunming 650091, China

Abstract

Accessibility is a fundamental indicator of location advantages. This paper presents one algorithm that measures time accessibility of public transit network using the bus line flow and transfer flow information. It first obtains flow and transfer information from the IC Card Toll System and estimates the transition probability matrix. Then the service area of bus stops is bisected by a Voronoi diagram. The time cost walking to and from bus stops is calculated. Finally, the average travel time cost of each location is calculated. One real-world public transit network in Kunming is calculated using this methodology. The area covered by the public transit network is first partitioned by 10m*10m grids. And then the travel time cost of every grid is calculated, which forms time contours.
Key Words: accessibility; public transit network; Voronoi diagram; IC toll information pricing strategies

Introduction

Significant attention is being placed on accessibility analysis of public transportation networks (PTN) in order to make them more desirable and to increase their utilization. Accessibility can be generally defined as the advantage of one place, at which some particular activities can be reached via a particular transit system (Weibull, 1976). Researchers from different areas defined it from different perspectives in the literature. These perspectives are:

(1) spatial separation measures (Kirby， 1976) that measures costs such as distance, travel time and travel fee (Moseley, 1979; Recker et al, 2001);
(2) potential measures (Bureau of Transportation Statistics, 1999; Hansen,1959; Vickerman, 1974) that use potential to assess the spatial advantages with respect to their attractiveness and spatial interactions, and,
(3) probability measures (Geertman and Ritsema, 1995) that assesses the expected value of trip time of one place to be reached from other places according to the time costs and frequencies.

While each perspective has its points, each neglects other worthwhile aspects of accessibility. A spatial separation measure neglects differences among attractiveness and the facts that activities declined with high travel costs. Conversely, potential measures generally do not consider temporal constraints, and the potential usually has no direct interpretations. Probability measures are mathematically rigorous but hard to calculate because related information is often unavailable.

In this paper, we propose one approach to use the expected travel time of one place to index its accessibility. The bus line flow and transfer flow information is first collected from the IC Card Toll System. Then the transition probability matrix can be estimated, which reveals the probability one trip begins at one bus stop and ends at another. Then the service area of bus stops is bisected by a Voronoi diagram. The time costs traveling to and from bus stops are calculated. Finally, the expected commute time of one place is calculated. One real-world PTN in Kunming is calculated using this methodology. The area covered by the public transit network is first divided into 10 m*10 m grids, and then the travel time cost of every grid is calculated, which forms time contours.

PTN Travel Time Cost

PTN Accessibility. In PTNs, there are two aspects of accessibility. The first is physical access to PTN services. The more people that live and/or work nearby PTN services, the greater the possibility the service will be utilized. The second aspect of accessibility is the accessibility with the PTNs. That is the space reachable from a particular location over a period of time. The shorter the distances along a route, the faster the travel speeds and the wider area one can cover within a given travel time budget.

Measuring PTN Accessibility Using Travel Time Cost. The travel time cost, $c(x,y)$, of location (x, y) can be formulated as:

$$c(x,y) = \sum_i p(x,y,i)(w(x,y,i) + c(i)) \quad (1)$$

where $w(x, y, i)$ is the time cost walking from location (x,y) to bus stop i.

$$w(x,y,i) = \frac{1}{v_w} d(x,y,i) \quad (2)$$

And v_w is the walking speed.

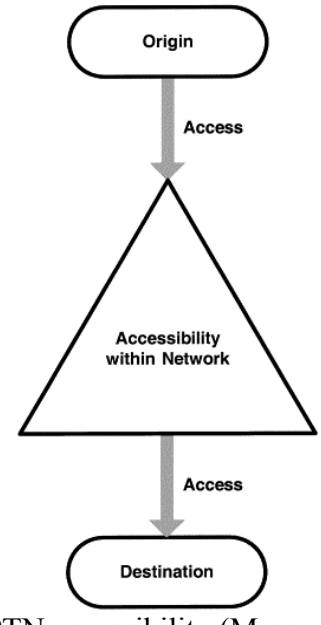


Figure 1. PTN accessibility (Murray et al. 1998).

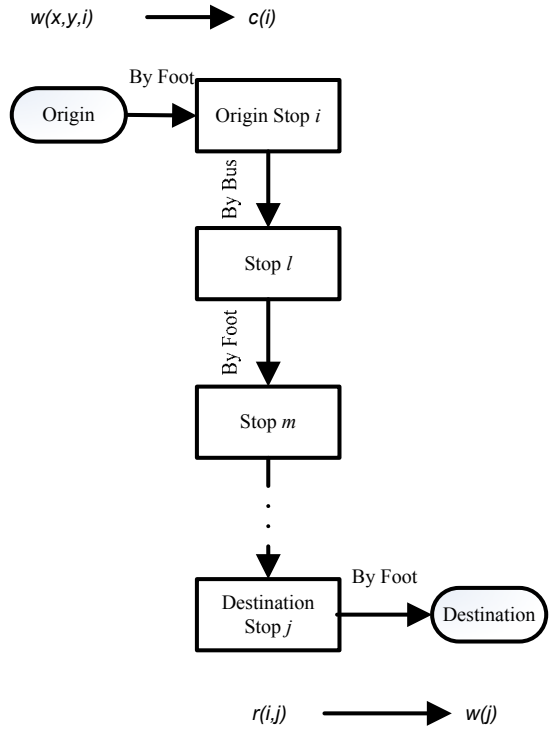


Figure 2. Public transport trip chain.

One typical PTN trip chain is regularly composed of access to PTN, waiting, transfer, or riding in PTN and departure from PTN.

Stop choice probability, $p\ (x,y,i)$, is the probability of one passenger starting the trip at location (x,y) choosing bus stop i as his or her first bus stop within the network,

which depends on *od(i)*. The count of passengers originates the trip within the network at bus stop *i*,and the impedance from (*x*, *y*) *to i*, the impedance function, *g*(*r*), of distance *r*=(*x, y, i*) is often specified as negative exponent function of *r*: *g*(*r*)=*exp*(-*βr*) (Liu and Zhu, 2004).

Assume that *p (x,y,i)* is proportional to *od*(*i*) and *p (x, y, i)* is proportional to*g*(*r*), then *p* (*x*, *y*, *i*) can be specified as:

$$p(x,y,i)=\frac{od(i)*e^{-\beta d(x,y,i)}}{\sum_{i} od(i)*e^{-\beta d(x,y,i)}} \tag{3}$$

where β is the decay parameter relating the decay pattern to the effective range of correlations that are meaningful. The travel time cost from bus stop *i* is *c(i)*.

$$c(i)=\sum_{j}(r(i,j)+w(j))p(i,j) \tag{4}$$

where,
r(i,j) is the in-vehicle time cost from bus stop *i* to bus stop *j*, including in-vehicle time, waiting time, and transfer time;
w(j) is the mean walking time cost departing from *j*;
p(i,j) is the possibility of one passenger originating the trip within the network at bus stop *i* destined for bus stop *j*.

And define $\bullet$ Operation of Matrix as:

$$\begin{aligned}&\mathbf{C}_{n\times 1}=\mathbf{A}_{n\times n}\bullet\mathbf{B}_{n\times n}\\&with\\&c_{i1}=\sum_{k}a_{ik}b_{ik}\end{aligned} \tag{5}$$

Eq. 4 can be written in matrix form:

$$\mathbf{C}=(\mathbf{R}+\mathbf{1}\cdot\mathbf{w}^{T})\bullet\mathbf{P} \tag{6}$$

where

$$\begin{aligned}&\mathbf{P}=[p(i,j)]_{N\times N}\\&with\\&\mathbf{R}=[r(i,j)]_{N\times N}\\&\mathbf{C}=[c(i)]_{N}^{T}\\&\mathbf{1}=[1,1\cdots,1]_{N}^{T}\\&\mathbf{w}=[w(i)]_{N}^{T}\end{aligned} \tag{7}$$

Detailed Procedure

***Transition Probabilities Matrix* P.** If the OD matrix of the PTN is known, the transition probability matrix **P** can be deduced by normalizing the OD matrix. Otherwise, it can be estimated from bus line flows and transfer flows.

When one passenger makes a decision about which route to ride, he or she usually takes into consideration walking time, waiting time, transferring time, and in-vehicle time to minimize the total travel cost. Yang made an investigation about the affects of the passenger's route-choice strategies in Nanjing in 1999. The investigation focused on transfer times, distance, and total time cost. It reveals that among these factors, 41.6% of passengers' first concern is about transfer count, 30.93% about total time cost, and 18.60% about distance (Wang et al. 2000). If the PTN trips are partitioned by transfer count, and neglected the trips of more than two transfers, then **P** can be reformulated as:

$$\begin{aligned} \mathbf{P} &= \sum_{k=0}^{\infty} \mathbf{P}_k \\ &\approx \mathbf{P}_0 + \mathbf{P}_1 + \mathbf{P}_2 \end{aligned} \tag{6}$$

where,

$$\mathbf{P}_k = [p_k(i,j)]_{N\times N}$$

and $p_k(i,j)$ is the probability of passengers originating the trip within the network at bus stop i destined for bus stop j through k transfers.

None-Transfer Transition Probabilities Matrix $\mathbf{P_0}$. $p_0(i, j)$ is related to the flow $f_0(l)$ of bus lines including bus stop i and bus stop j and the proportion of the bus line flow, $q(i,j)$, just boarding at bus stop i and alighting at bus stop j. It can be written as:

$$p_0(i,j) = \frac{\sum_l f_0(l) q(i,j,l)}{\sum_l \sum_j f_0(l) q(i,j,l)} \tag{7}$$

Estimation of q(i, j, l). Number of passengers' in-vehicle stops is Poisson distributed (Dou et al., 2007). So, the flow intensity from i to j on bus line l can be approximated by:

$$\begin{aligned} &q(i,j,l) = \frac{\lambda^k e^{-\lambda}}{k!} \\ &\lambda: \text{ average number of in - vehicle bus stops} \\ &k: \text{ number of bus stops of line } l \end{aligned} \tag{8}$$

One-Transfer Transition Probabilities Matrix $\mathbf{P_1}$. If two bus lines include common bus stops, passengers can transfer between these bus lines at these bus stops. Otherwise, if some bus stops on two bus lines are close enough, passengers still transfer by walking. And so, there regularly exist a great number of routes from one stop to another by transferring. But most of these routes are too circuitous and few passengers will choose them. Define the utility of route r from i to j, $u(i,j,r)$, *as:*

$$u(i,j,r) = \frac{c(i,j,r)}{c(i,j)} \tag{9}$$

where $c(i,j,r)$ is the cost of route r, and $c(i,j)$ is the minimum cost from i to j.

Based on Logit route choice model (Wang et al., 2002), the proportion of the route flow from i to j on route r q (i, j) through one transfer can be calculated by:

$$q_1(i,j,r) = \frac{e^{-\theta u(i,j,r)}}{\sum_r e^{-\theta u(i,j,r)}} \tag{10}$$

where parameter θ indicates the level passengers are familiar to the PTN. If it is zero, then all possible routes will be equally chosen by passengers.

The computation of $\mathbf{P_2}$ is similar to $\mathbf{P_1}$.

In-Vehicle Time Cost Matrix **R.** The average travel time cost from i to j can be computed by:

$$r(i,j) = \sum_r c(i,j,r) * p(i,j,r) \tag{11}$$

where the travel time cost of route r is dependent on waiting time, in-vehicle time, and transfer time.

Departure Time Cost. Service areas of bus stops can be represented by Voronoi diagrams (Atsuyuki and Atsuo, 1997). Given a set of two or more but a finite number of distinct points in the Euclidian space, we associate all locations in that space with the closest member(s) of the point set with respect to the Euclidean distance. The result is a tessellation off the plane into a set of regions associated with members off the point set (Okabe and Sugihara, 1992). The mathematical definition is the following:

If S is a set of points, $S = \{ p_1 , p_2 , \ldots, p_n \}$, then the Voronoi polygon of p_i is defined as:

$$VOR_i = \{ x : d(x, p_i) < d(x, p_j) \mid p_i , p_j \in S , p_i \neq p_j \} \tag{12}$$

where d is the Euclidian distance.

An example of an ordinary Voronoi diagram is presented in the figure below.

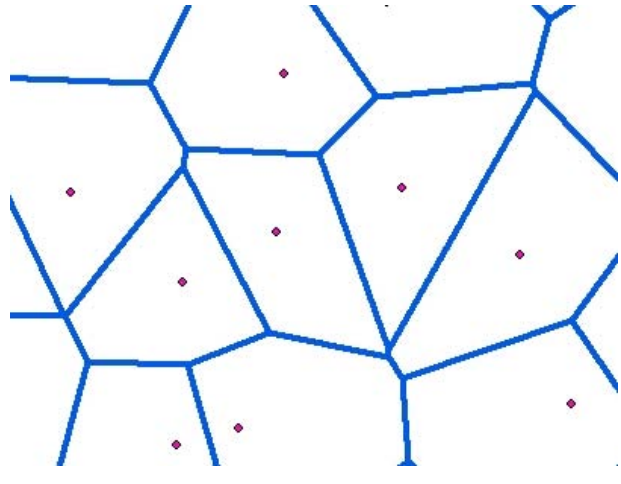


Figure 3. Voronoi diagram of bus stops.

Time Cost Departing from Bus Stops w(j). To simplify the computation of time cost departing from bus stops, the service area of one bus stop is approximated to one circular region with the same area. And assume that the probability density of the walking distance r departing the bus stop is $k*g(r)$, then $w(j)$ can be written as:

$$w(j) = k\int_0^R \frac{r}{v_w} g(r)dr$$

$$s.t. \quad (13)$$

$$k\int_0^R g(r)dr = 1$$

It can be deduced that:

$$k=\beta/(1-exp(-\beta R)) \quad (14)$$

$$w(j)=k(1- exp(-\beta R)- \beta R*exp(-\beta R))/(\beta^2 * v_{ww}) \quad (15)$$

Case Study

As an example, the public transport network of Kunming in China is studied using the method proposed. The network is composed of 994 bus stops and 148 bus lines and its service covers an area of 20 km*30 km.

The information of non-transfer bus line flows and transfer route flows was collected by intelligent card toll system in January 2009. The spatial information of bus stops and bus lines were gained with GPS, and GIS for the network was then established. The service area was partitioned by grids of 10 m*10 m. The travel time cost of each grid was then calculated using Eq. (1). Each grid was treated as one pixel, and its travel time cost was considered as the gray level. Then one raster image of 2000*3000 pixels was drawn. Connecting pixels of the same travel time cost, time contours of this PTN are drawn, as shown in Figure 4. The travel time cost of the upper-left point in the picture is about 120 minutes and the travel time cost of the two highlighted lines intersection point is about 20 minutes. The difference of two neighboring time contours is 5 minutes.

In the calculating process, parameters are set as follows: v_b=15 km/h, v_w=5 km/h, β=1/500 m^{-1}, θ=3, and λ=6.

From the results shown in Figure 4, it can be seen that the area along the two highlighted lines is of good accessibility, which conforms to the reality that they are arterial roads and both have bus-only lanes.

But the travel time costs of some peripheral area of the PTN were undervalued (e.g., the six highlighted points in Figure 4). The cause is that Eq. 12 takes into account only the bus stop count but neglects the fact that few passengers travel uptown. This should be amended in future research.

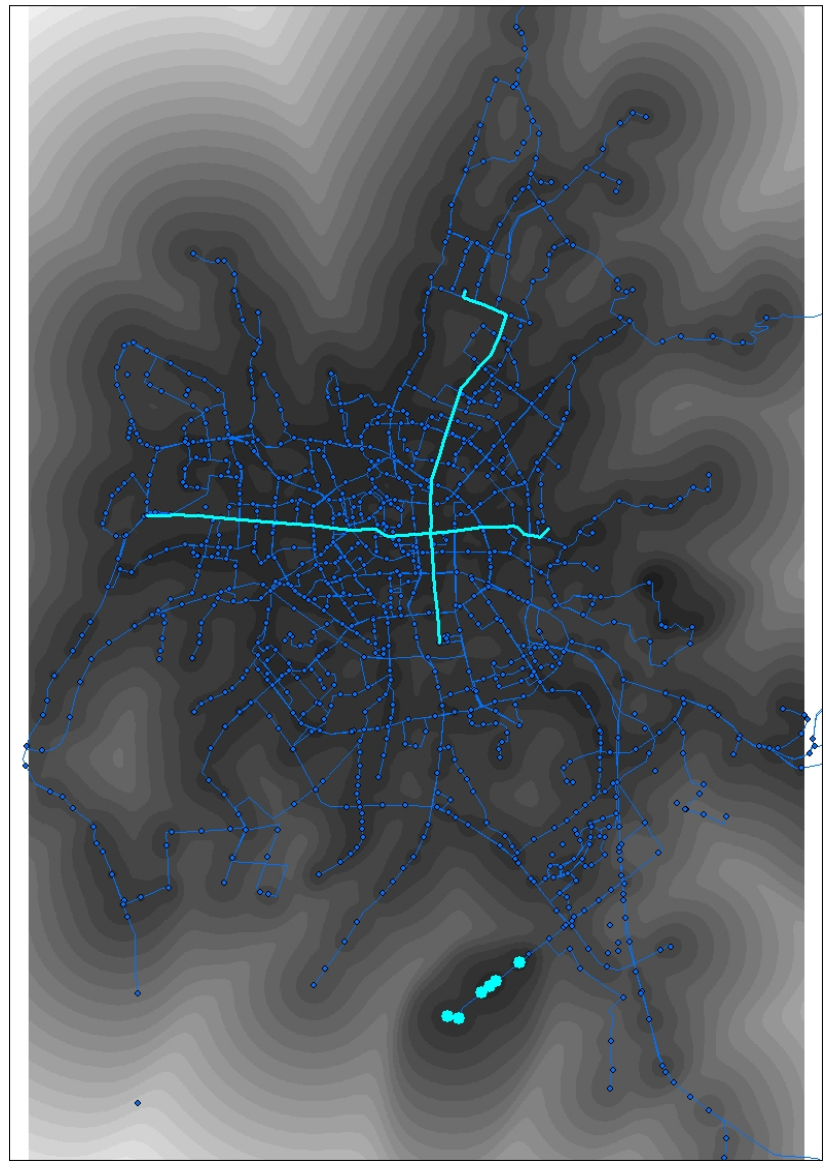


Figure 4. Time contours of Kunming PTN.

Conclusions

This paper proposed one method evaluating PTN accessibility based on measuring travel time cost using IC card toll information. The methods turned out to be operable and practical through a case study in Kunming city. It can be employed to appraise PTN accessibility, and offer a visual tool for PTN planning and management administration to strengthen the weak area of PTN to optimize the service. In practice, the spatial distribution of passengers has to be estimated indirectly for lacking of related information. And some factors such as bus speed and waiting time are space-dependent and time-dependent but they are simplified to constant parameters. These lead to some deviations from the facts, which should be corrected by conducting a field survey.

References

Atsuyuki, O., and Atsuo, S. (1997). "Locational optimization problems solved through Voronoi diagrams." *European Journal of Operational Research*, 98(3), 445-456.

Bureau of Transportation Statistics. (1999). *Transportation Statistics Annual Report (TSAR)*, U.S. Department of Transportation.

Dou, H., Liu, H., and Yang, X. (2007). "OD Matrix Estimation Method of Public Transportation Flow Based on Passenger Boarding and Alighting." *Computer and Communications*, 25(2), 79-82.

Geertman, S. C. M., and Ritsema, V. E. (1995). "GIS and models of accessibility potential: an application in planning." *International Journal of Geographical Information System*, 9(1), 67-80.

Hansen, W. G. (1959). "How accessibility shapes land-use." *Journal of the American Institute of Planners*, 25, 73-76.

Kirby, H. R. (1976). "Accessibility indices for abstract road network." *Regional Studies*, 10, 479-482.

Liu, S., and Zhu, X. (2004). "Accessibility Analyst: an integrated GIS tool for accessibility analysis in urban transportation planning." *Environment and Planning B*: *Planning and Design*, 31,105-124.

Moseley, M. J. (1979). *Accessibility: the Rural Challenge*, London: Methuen.

Murray, A. T., Davis, R., and Stimson, R. J. (1998). "Public transportation access." *Transportation Research Part D*, 3(5), 319-328.

Okabe, A., Boots, B., and Sugihara, K. (1992). *Spatial Tessellations Concepts and Applications of Voronoi Diagrams*, John Wiley & Sons, Chichester. New York – Brisbane – Toronto – Singapore.

Recker, W. W., Chen, C., and McNally, M. G. (2001). "Measuring the impact of efficient household travel decisions on potential travel time savings and accessibility gains." *Transportation Research A*, 35,339-369.

Vickerman, R. W. (1974). "Accessibility attraction and potential: a review of concepts and their use in determining mobility." *Environment and Planning A*, 6, 675-691.

Wang, W., Yang, X., and Chen, X. W. (2002). *Urban Public Transport System Planning and Management*, Science Press*, Beijing*, 112-113.

Wang, W., Yang, X., and Ma, W. T. (2000). "GIS-based public transit passenger route choice model." *Journal of Southeast University*, 30(6), 87-91.

Weibull, J. W. (1976). "An axiomatic approach to the measurement of accessibility." *Regional Science and Urban Economics*, 6, 357-379.

Research and Development of Public Transit Route Inquiry System Based on GIS

Mingming Chen[1], Xi Chen[2]and Huimin Niu[3]

[1] Assistant, [2]Graduate Student, [3] Professor, School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou, Gansu, P.R. China, 730070; Email: Lzcmm2000@mail.lzjtu.cn

Abstract

Public transit route choice is the key technology of the public transit passenger information system. In this paper, a public transit route inquiry theory, which is based on the difference in travel behavior of public transit passengers, is presented. The critical issue in the theory is to analyze the optimization inquiry objectives, which consist of transfer times, travel time and travel costs, and to design the different algorithms. Combined with the characteristics of geographic information systems (GIS), a public transit route inquiry system is developed. It is not only able to provide accurate, reliable and comprehensive information service for passengers, but also to provide intelligent decision support for the bus company in bus dispatching management. Application in the public transit network of Lanzhou city showed that the proposed algorithm was feasible and effective. Furthermore, the system was intuitive and had a friendly interface that was easy and simple for users to operate.

Key Words: public transit route; inquiry system; transfer times; travel time; travel cost

Introduction

With the continuous development of cities, urban roads have been continually reconstructed, and the number of the bus lines is also increasing. It is convenient for passengers due to the multiple public transit route choices. On the other hand, it will cause some difficulties on how to choose the optimal transit route. Even the residents who are familiar with the urban route information can easily find a travel route using

a traffic map. It is not necessarily the most excellent, just a reference of options. Therefore, the establishment of a public transit route inquiry system is the key in solving the current inefficient transit route inquiry problem.

Research and development of the urban public transit route inquiry system is aimed at showing the optimal travel time, travel routes and transfer stop for passengers, providing accurate, reliable comprehensive travel service information and initiative and reasonable guidance and planning, reducing the unnecessary traffic and thereby attracting more passengers to choose a public transit mode. Furthermore, it also can provide intelligent decision support for the bus company in bus dispatching management.

Urban public transit route inquiry is to find the optimal public transit route between any origin and destination. The corresponding inquiry algorithm is the key technology of the public transit passenger information system, but it is also a hot topic within the public transit field in recent decades. Recently, it has received much attention by many experts and scholars.

Koncz and Greenfeld (1996) formulated a strategy for solving static multiple-optimal-path transit network problems. Nguyen (1998) used implicit enumeration of hyper paths in a logit model for transit networks. Guido (2004) proposed a route choice algorithm with on-line information at stops. Mues and Pickl (2005) developed two approaches on transhipment and time windows within vehicle routing problems. Yetis and Nurcan (2008) presented a route choice model of transportation network in Denizli using the fuzzy logic model, logistic regression model and survey-based data. Yang et al. (2000) presented a shortest path algorithm, which used the least number of transfer times as the primary object and shortest path distance as the second object. Zhao et al. (2004) built a dynamic algorithm of minimal transfer times for a viable route. Su and Shi (2005) presented an optimal route choice of public traffic network based on shortest path searching. Peng and Hu (2009) designed a public transit optimum algorithm based on the least transfer times.

Up to now, most existing main objectives of algorithms are based on transfer times, while the travel time and travel cost as the object has taken less attention.

In the aspect of system development and applications, it is relatively undeveloped in China. It still has many difficulties; such as the user interface is not simple enough, and passenger access to public transit information is not comprehensive.

This paper is organized as follow. The next section presents the public transit route inquiry theory. The corresponding algorithm is designed in Section 3. In Section 4, the public transit route inquiry system of Lanzhou city is developed to illustrate the application of the former theory. The last section highlights the conclusions and suggests future research directions.

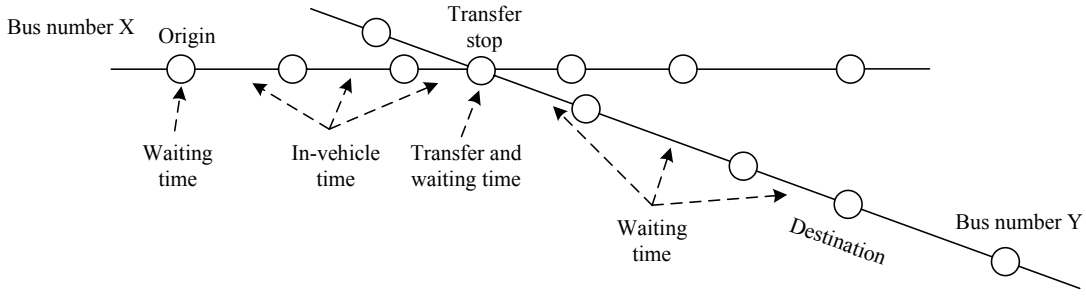


Figure 1. Travel time sketch map.

Public Transit Route Inquiry Theory

Public transit route choice is the key technology of the public transit passenger information system. The basic requirement is fast, convenient and economical. "Fast" refers to the passengers reaching their destination within the shortest time; "Convenience" means that the route choice of passengers needs to have lower transfer times; "Economical" indicates that the cost for passengers is at a minimum from origin to destination. The transit route choice depends on the purpose of passengers. As the trip purposes are different, passengers will weigh all the possible options in order to maximize the psychological satisfaction as a basis for decision-making.

The problem of the public transit route usually focuses on solving the shortest path between any two stops in the transit network. As long as there are edges between two vertices, we can search the path without considering the transfer price, and therefore the shortest path may cause passengers to transfer to many buses to reach their destination, which it is not the best option for bus passengers. So the transfer time is a very sensitive factor for passengers. Although the distance of the path is longer than the shortest path, while the needed transfer times are lower, the majority of passengers will choose the option with a longer distance.

The objective of minimum travel time of transit route can be simply defined as the time it takes for passengers from an origin to reach their destination. In public transit, travel time can be determined by many factors, such as transfer times, bus running time, bus frequency, the distance between the origin and the destination, etc.

Figure 1 shows an example of a passenger's total time required to complete a trip, including the waiting time for bus number X and Y, in-vehicle running time and the transfer time.

When choosing the public transit route, the first consideration of passengers is the lower transfer times, then the travel cost. Passengers prefer to reach their destination with the lowest cost possible. If the number of transfer times is increased, the unpredictable factors will also be greatly increased. There are two types in the ticket system, flat fare and sectional fare. If the bus line uses the flat fare system, the

optimal bus route with the objective of minimal travel cost can be generated in the route which has fewer transfer times. If the bus line uses the sectional fare system, it will be more economical for the passengers that have a shorter distance between their origin and destination, whereas the passengers with a longer distance choose the flat fare bus line because it will be relatively economical for them.

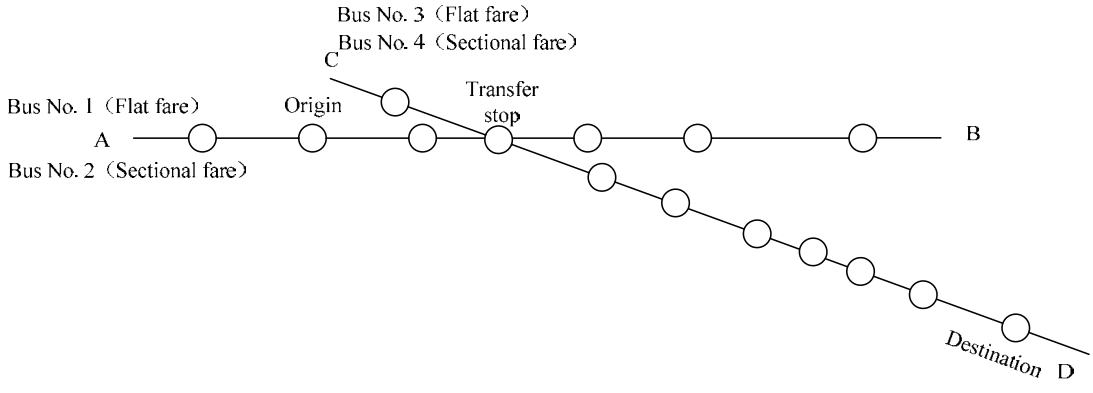


Figure 2. Travel cost sketch map.

Figure 2 shows an example of travel cost with one transfer time. Suppose the flat fare ticket is 1 Yuan, and the sectional fare ticket is 0.5 Yuan within three stops (each stop after three is charged 0.5 Yuan). So the minimal cost of the route is 1.5 Yuan, and the travel route of the passengers is to take bus number 2 from origin to the transfer station, and take bus number 3 to the destination.

Public Transit Route Inquiry Algorithm

Transfer Times. Transit network can be transformed into a directed graph, which consists of a series of stops and arcs, noted as $G(V, E)$, where V is the stop set of G and E, the arc set of G. The algorithm for public transit route inquiry with the objective of transfer times can be designed with a transfer matrix in the public transport network. The transfer matrix $H = (h(i, j))$ can be described as the passengers' transfer times from origin to destination, where $i \in V$, $j \in V$, $h(i, j)$ denotes the transfer times from stop i to stop j. Define the initial transfer matrix $H_0 = h_0(i, j)$.

Then the algorithm for transfer times can be taken as follows:

Step 1: Generate the initial transfer matrix H_0: If a passenger can reach destination j from the origin i without transfer, $h_0(i, j) = 1$; otherwise, $h_0(i, j) = 0$.

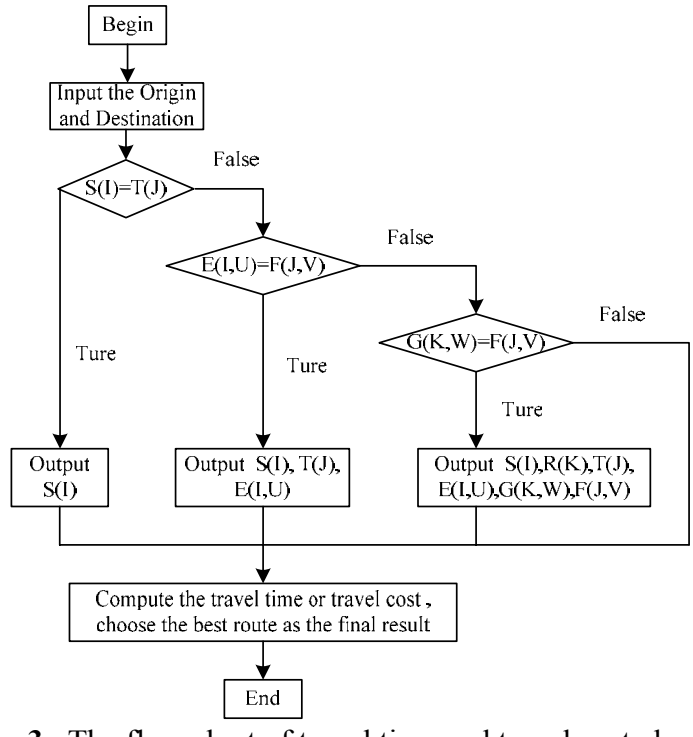


Figure 3. The flow chart of travel time and travel cost algorithm.

Step 2: Generate the first transfer matrix H_1 : If $h_0(i,j)=1$, then $h_1(i,j)=h_0(i,j)$. If $h_0(i,j)=0$ and $\exists k$, $h_0(i,k)=1$, $h_0(k,j)=1$, then $h_1(i,j)=2$; otherwise, $h_1(i,j)=0$.

Step 3: Generate the second transfer matrix H_2 : If $h_1(i,j)=1$ or $h_1(i,j)=2$, then $h_2(i,j)=h_1(i,j)$. If $h_1(i,j)=0$ and $\exists k$, $h_1(i,k)>0$, $h_1(k,j)>0$, then $h_2(i,j)=3$; otherwise, $h_2(i,j)=0$.

Step 4: Generate the n th transfer matrix H_n with the method of former step.

Step 5: Stop until there is no zero element in the matrix H_n .

***Travel Time and Travel Cost*.** Let A be the origin stop, and let B be the destination stop. Define $S(I)$ as the bus line running through the origin stop A , and $T(J)$ as the bus line running through the destination stop B . Define $E(I,U)$ as the bus stop on the bus line $S(I)$, and $F(J,V)$ as the bus stop on the bus line $T(J)$. Define $G(K,W)$ as the bus stop on the bus line which is running through the bus stop $E(I,U)$. The algorithm for travel time and travel cost can be shown in Figure 3.

The Development of the Public Transit Route Inquiry System

The system which uses the Windows system for the operation platform, VB6.0 as the tools of the system development, MapInfo and MapX for GIS platform and controls, develops the public transit route inquiry system of Lanzhou City. The system mainly

has four modules: map operation, map output, information inquiry and route inquiry, shown in Figure 4.

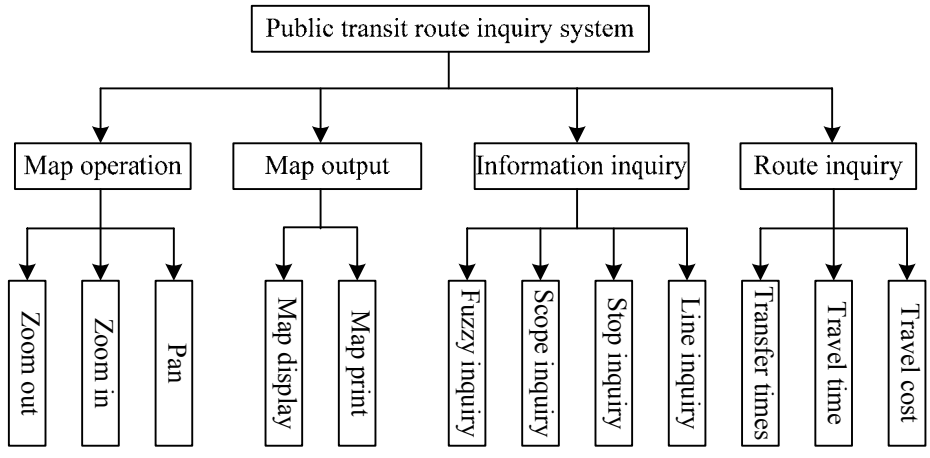


Figure 4. The function framework of system.

University" and "Railway bureau" (long-distance OD pair), respectively, using the different travel goals scheme to inquiry, the three inquiry results with corresponding goals have the obvious difference, shown in Figures 5, 6 and 7.

Assuming the origin and destination of a passenger are "Lanzhou Jiaotong University" and "West station" (short-distance OD pair), respectively, comparative analysis with different inquiry goals can be found that the transfer time of three goals are all zero, namely the passenger can go directly to the destination. But travel time is slightly different. Since bus 3 and bus 103 are both express buses, travel time is relatively smaller. The optimal travel cost is 1 Yuan, just the same.

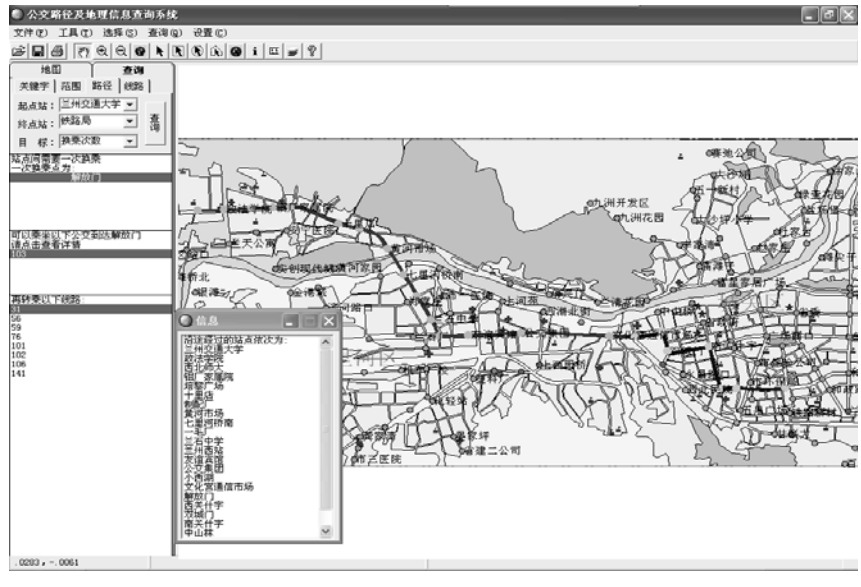


Figure 5. Long-distance transfer numbers inquiry result.

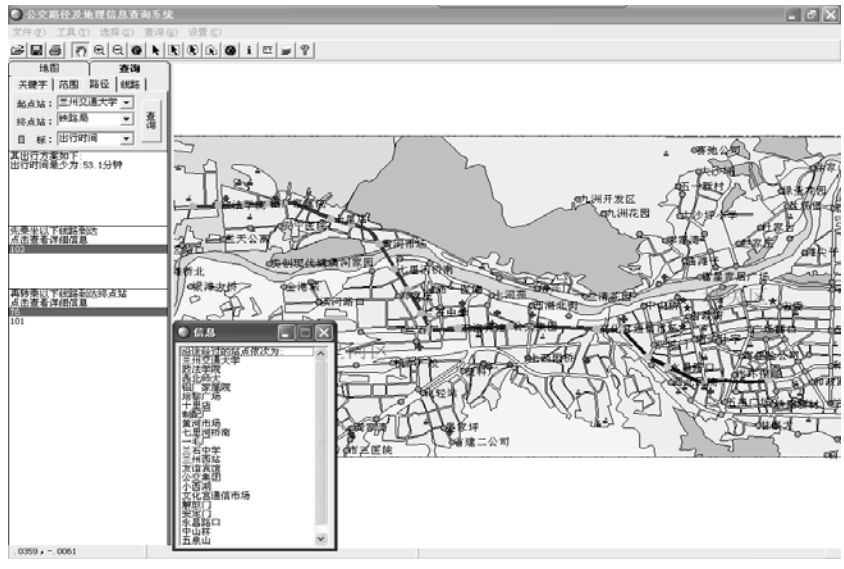


Figure 6. Long-distance travel time inquiry result.

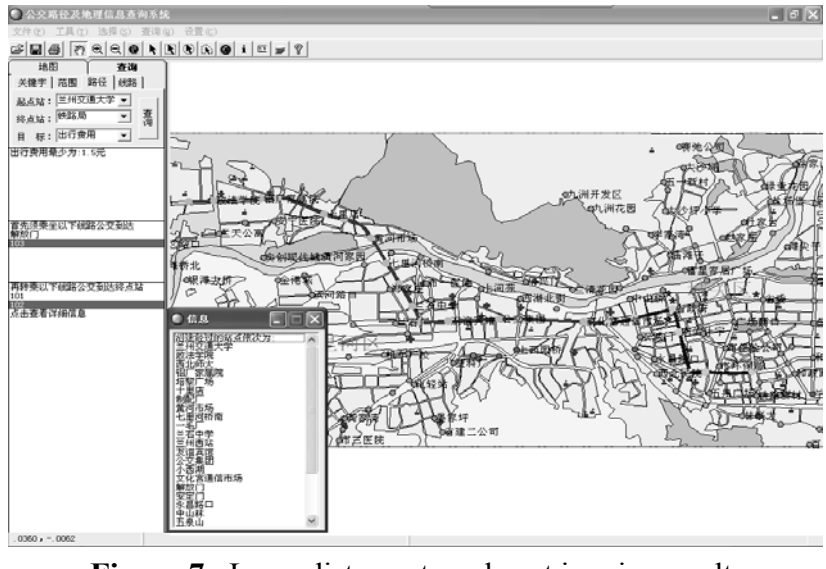


Figure 7. Long-distance travel cost inquiry result.

Therefore, if the OD trip distance was not far, the system-generated target routes with different purpose were almost identical, but in the long distance of OD trip, it would vary greatly with each other.

Conclusions

The urban public transit route inquiry system, which uses the database and tools of GIS, has realized the stop inquiry, line inquiry, scope inquiry and route inquiry of public transit and the auxiliary display with the electronic map. The system is intuitive with a user-friendly interface. It can not only show the inquiry results with the way of text, but also can highlight the graphics and position of the selected transit route.

This paper uses Lanzhou city as an example. Without loss of generality, the inquiry system can be used with other cities. The user can input the corresponding geographic data and basic bus information, and the system becomes completely suitable for any urban transit route inquiry. The system not only speeds up the process of the urban information construction and the information service, but also can be widely applied in the government, transportation enterprises and other public places.

Furthermore, combining the public transit inquiry system and the touch screen technology is an important topic for further research. The touch screen can be installed in bus terminals, hotels, squares and other public places to provide public transit inquiry service directly and can also combine the network technology to release on the Internet, making the system more convenient to the user.

Acknowledgments
The work described in the paper was supported by National Nature ScienceFoundation of China (No. 50968009) and the Research Fund for the Doctoral Program of Higher Education (No. 20096204110003).

References

Gentile, G. (2004). "Route choice on transit networks with on-line information at stops." *University of Rome*, 1-15.

Koncz, N., and Greenfeld, J. (1996). "A strategy for solving static multiple-optimal-path transit network problems." *Journal of Transportation Engineering*, 218-225.

Mues, C., and Pickl, S. (2005). "Transshipment and time windows in vehicle routing." *Proc. 8th International Symposium on Parallel Architectures, Algorithms and Networks*, 113-119.

Murat, Y. S., and Uludag, N. (2008). "Route choice modeling in urban transportation networks using fuzzy logic and logistic regression methods." *Journal of Scientific & Industrial Research*, 19-27.

Peng, J. K., and Hu, P. (2009). "Application research of gis based on map objects in public transportation query system." *Computer Engineering and Design*, 3469-3471.

Sang, N. (1998). "Implicit enumeration of hyperpaths in a logit model for transit networks." *Transportation Science*, 54-64.

Su, A. H., and Shi, F. Z. (2005). "Optimal route choice of public traffic network based on shortest path searching." *Journal of Engineering Graphics*, 55-59.

Yang, X. M., Wang, W., and Ma, W. T. (2000). "GIS-Based public transit passenger route choice model." *Journal of Southeast University*, 87-91.

Zhao, Q. X., Ma, Z. Q., and Zhang, F. (2004). "Algorithms for public transit trip with minimal transfer times and stops." *Computer Applications*, 136-138.

Conceptual Framework and Mathematical Model for Optimizing Train Blocking and Scheduling Synchronization

Haiying Li[1], Lingyun Meng[2], Jianrui Miao[3], Xi Jiang[4] and Zhenhuan He[5]

[1] Associate Professor, [2] Lecturer, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 10044, China, Email: lihaiying@jtys.bjtu.edu.cn
[3] Lecturer, [4] Associate Professor, [5] Lecturer, State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

Abstract

Aiming at synchronization of train blocking and scheduling problems both in space and time, this research first constructs a conceptual framework for synchronizing train blocking plan and train timetable, based on a feedback mechanism between them via "quasi train path graph". On the basis of the conceptual framework, a bi-level mathematical programming model and improved genetic algorithm are proposed. Finally, comprehensive numerical experiments including both fixed fee and variable fee are conducted to demonstrate applicability of the above method. Results show that solutions of this model could identify train operating time window and train series for each goods OD. In addition, the solutions have considered connection time for train blocks in marshalling stations so as to make the train blocks controllable.
Key Words: railway transportation organization; train blocking plan; train timetable; synchronization

Introduction

With the rapid economy development, transportation quality should be improved significantly to meet the increasing requirements for transportation convenience, efficiency and equilibrium. To achieve such goals, train blocking process and train scheduling process should be combined together to significantly improve the transport reliability and some other important indices. That is, in the planning process, "goods flow" and "train path" must be connected via a feedback mechanism. This requires good-quality synchronization between train blocking plan and train timetable.

Although there are a wide range of studies devoted to train blocking plan optimization or train scheduling problems (e.g. Brannlund (1998), Cai(1992, 1998), Cao (1993(a), 1993(b)), Caprara(2002), Carey (1994(a), 1994(b), 1995)), there is few study that is focused on synchronizing train blocking plan and train timetable. Carl van Dyke(2006) presented a multi model applied systems to solve such problems in the context of North America railroad. However there are many differences on the freight transportation mode between North America railroad and Chinese railway, thus those systems are not quite suitable for solving the train blocking and scheduling problem in China. Jorg (2001) introduced MAS method to build a train timetabling system. In this research, they viewed each train block as an agent and the agents negotiate to finally get the optimal solutions. Haiying Li et al. (2007) presented a MAS-based approach to combine the train blocking and scheduling problem and corresponding algorithms, by considering the characteristics of Chinese railway. In their study, they discussed the definitions of various types of agents, interacting mechanism between agents, and so on.

In this research, we first proposed a conceptual framework for synchronizing train blocking plan and train timetable. By means of bi-level programming techniques, we further put forward mathematical formulations to formalize the conceptual framework in detail. Corresponding to the framework and model, an improved genetic algorithm was designed and implemented to get suboptimal solutions.

Frameworkfor Synchronizing Train Blocking Plan and Train Timetable

Problem Statement. Before further detailing problem statement, we first define the synchronization problem as constructing a combined train blocking plan and train timetable to get the optimal shippers' satisfaction and profits of railway transport company, based on transport needs (e.g. goods OD) and layout of transport resources.

Typically speaking, train blocking plan is used to identify organizing pattern of train blocks in space from macro-planning perspective, while train timetable is used to identify occupancy of trains on resources both in space and time from micro-planning perspective. Thus, to synchronize train blocking plan and train timetable, we have to build a connection bridge between them. So we proposed a "Quasi train path graph" to solve this problem. "Quasi train path graph" is a plan that further identifies train operating time windows on the basis of train blocking plan. It has some features of train timetable, but it differs from train timetable. That is, it doesn't identify the detailed arrival and departure times of trains at each station. So, it has more detailed information than train blocking plan but less detailed information than train timetable, which means that it could be used to connect train blocking plan and train timetable.

Along with this idea, we divided the synchronization optimization problem of train blocking plan and train timetable into two sub problems, one is synchronization problem between train blocking plan and "quasi train path graph", and the other one is synchronization problem between "quasi train path graph" and train timetable.

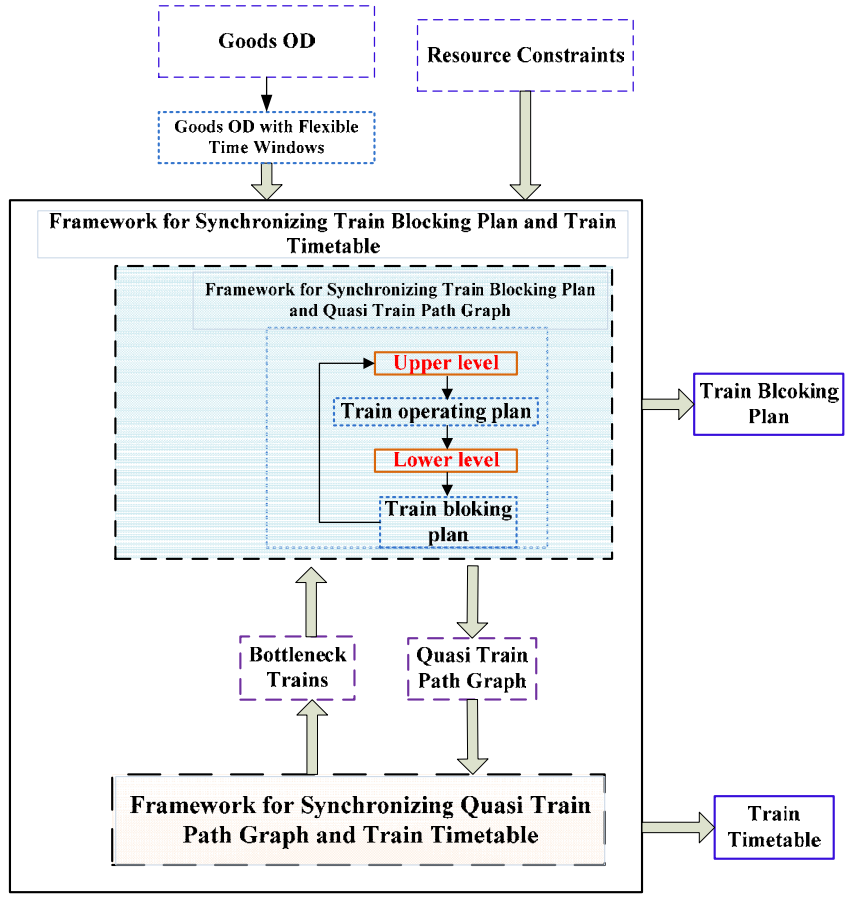


Figure 1. Conceptual framework for synchronizing train blocking plan and train timetable.

Conceptual Framework Model for Train Blocking Plan and Train Timetable Synchronization. Based on the above analysis, we could propose a conceptual framework as illustrated in Figure 1.

As illustrated in Figure 1, there are 4 steps to achieve synchronization of train blocking plan and train timetable in detail:
Step 1: On the basis of investigation and prediction of goods flow, and subdivision of transport demands, introduce freights OD with flexible time windows to represent the requirements of train blocks on arrival and departure times.
Step2: Identify the train block combination scheme with the constraints of arrival and departure time windows, then solve and put out the train blocking plan and "quasi train path graph".
Step 3: Schedule trains by the train scheduling model with the inputs of quasi train paths.
Step 4: If it could not allocate train paths in certain time windows during train scheduling process, then send the bottleneck trains back to upper model and adjust corresponding train blocking scheme or train operation time windows, back to Step 3. Repeat the above four steps to get satisfactory/suboptimal solutions.

Table 1. Subscripts and parameters used in mathematical formulations.

Symbol	Description
N_{mn}^{k}	= wagon block, including Q , $t_S^{'}$, $t_S^{''}$, $t_E^{'}$, $t_E^{''}$, P_{mn}.
Q	= number of wagons in N_{mn}^{k}
$t_S^{'}$	= the expected earliest departure time for N_{mn}^{k}
$t_S^{''}$	= the expected latest departure time for N_{mn}^{k}
$t_E^{'}$	= the expected earliest arrival time for N_{mn}^{k}
$t_E^{''}$	= the expected latest arrival time for N_{mn}^{k}
P_{mn}	= serial of arcs in the path from v_m to v_n
$\mathrm{C_{mn}}$	= capacity of the section a_{mn} maximum number of trains that can be served in the section over some period time(i.e. one day)
$\mathrm{M_{v_m}}$	= capacity of the yard v_m, maximum number of wagons that can be marshalled in the yard over some period time (i.e. one day)
$\mathrm{m_f}$	= maximum number of wagons that can be grouped into one single train
T	= set of trains
d_i^{ll}	= the earliest departure time for train i at section l
d_i^{lu}	= the latest departure time for train i at section l
a_i^{ll}	= the earliest arrival time for train i at section l
a_i^{lu}	= the latest arrival time for train i at section l
d_i^{l}	= the earliest departure time for train i at the origin yard
d_i^{u}	= the latest departure time for train i at the origin yard
a_i^{l}	= the earliest arrival time for train i at the destination yard
a_i^{u}	= the latest arrival time for train i at the destination yard
t_{mn}^{S}	= start time of capacity loss time window for maintenance reason
t_{mn}^{E}	= end time of capacity loss time window for maintenance reason
v_i^{B}	= origin yard of train i
v_i^{E}	= destination yard of train i
Y_{mn}^{k}	= serial of trains that wagon blocks N_{mn}^{k} belongs to
B_{v_m}	= realized marshalling count of yard v_m
$M_{mn}^{k,i}$	= binary variable, 1 if N_{mn}^{k} is marshalled in yard v_i, 0 otherwise
$P_i^{a_{mn}}$	= binary variable, 1 if train i passes through arc a_{mn}, 0 otherwise
l_{mn}	= length of all sections that trains in Y_{mn}^{k} pass through
S_{mn}^{k}	= revenue to railway company from N_{mn}^{k}
$Z_{mn}^{t,i}$	= cost to railway company for operating train i from v_m to v_n at time section t
L_{mn}^{k}	= convenience to shippers from N_{mn}^{k}

Table 1. Subscripts and parameters used in mathematical formulations (continued).

C_{mn}^{k}	= cost to shippers for N_{mn}^{k}
$Z_{mn}^{m_c,i}$	= punishment cost for the less realized grouped wagons in train i with count m_c
$Z_{v_m}^{c}$	= average required marshalling cost for yard v_m
$Z_{v_m}^{t}$	= average required marshalling time for yard v_m
ω_1	= weight of railway company revenue
ω_2	= weight of shippers convenience
O_I	= the departure train units on yard I
D_I	= the arrival train units on yard I

Table 2. Decision variables used in mathematical formulations.

Symbol	Description
$x_{mn}^{t,i}$	= binary variable, 1 if allocate one train (i) from v_m to v_m at time section t, 0 otherwise
$y_{mn}^{k,i}$	= binary variable, 1 if train i contains N_{mn}^{k},0 otherwise

Mathematical Formulations

This section focuses on a bi-level programming model. The model consists of two levels, i.e. upper level and lower level. A solution of the upper level planning stage determines the train operating plan and the aim of the upper stage is to maximize total weighted sum of profits of railway company and the convenience of shippers. More specifically, the profit of railway company mainly depends on the revenue by transporting goods and the cost caused by operating trains on railway lines and marshalling wagons in yards. Based on the solution in the upper stage, the lower level planning stage is designed to determine how wagons are grouped into trains in order to minimize the cost of shippers, in result, solution in the lower stage reflects profit of upper stage.

Symbols and Parameters. We use a graph $G=(V,A)$ to represent a railway network. V is the set of vertices, $v_m \in V$,Each vertex represents one yard on the network. A is the set of arcs, a_{mn} represents the arc from $v_m$to v_n, $a_{mn} \in A$, $w_{mn}$is the weight of a_{mn}. Table 1 lists subscripts and mathematical parameters to formalize mathematical formulations.

Mathematical Formulations.

Upper-level planning formulations. The upper-level planning could be formulated as below.

$$\text{Maximize } [\omega_1 (\sum_{m=1}^{P}\sum_{n=1}^{P}\sum_{k=1}^{K} S_{mn}^{k} - \sum_{i\in T} Z_{mn}^{t,i} - \sum_{i\in T} Z_{mn}^{m_c,i} - \sum_{v_m\in V} B_{v_m} Z_{v_m}^{c}) + \omega_2 \sum_{m=1}^{p}\sum_{n=1}^{p}\sum_{k=1}^{K} N_{mn}^{k}(Q) L_{mn}^{k}] \quad (1)$$

Subject to:
(1) Section capacity constraints

$$\sum_{i \in T} x_{mn}^{t,i} P_i^{a_{mn}} \le \mathrm{C}_{mn},\ a_{mn} \in A \tag{2}$$

(2) Capacity loss constraints for maintenance reason

$$\begin{cases} t_{mn}^{E} \le d_i^{ll} \le 24, t_{mn}^{E} \le a_i^{lu} \le 24 \\ t_{mn}^{E} \le d_i^{ll} \le 24, 0 \le a_i^{lu} \le t_{mn}^{S} \\ 0 \le d_i^{ll} \le t_{mn}^{S}, 0 \le a_i^{lu} \le t_{mn}^{S} \end{cases} \tag{3}$$

Lower-level planning formulations. The lower-level planning can be formulated as:

$$\text{Minimize} \sum_{m=1}^{P} \sum_{n=1}^{P} \sum_{k=1}^{K} C_{mn}^{k} \tag{4}$$

Subject to,
(1) Yard capacity constraints

$$\sum_{m=1}^{p} \sum_{n=1}^{p} \sum_{k=1}^{K} M_{mn}^{k,i} N_{mn}^{k}(Q) \le \mathrm{M}_{v_m},\ v_m \in V \tag{5}$$

(2) Physical limits to the number of wagons that can be grouped into a single train

$$\sum_{m=1}^{p} \sum_{n=1}^{p} \sum_{k=1}^{K} y_{mn}^{k,i} N_{mn}^{k,i}(Q) \le \mathrm{m}_f,\ i \in T \tag{6}$$

$N_{mn}^{k,i}(Q)$ represents the count of wagons marshalled into train *i of* N_{mn}^{k}.

(3) OD transfer constraints, For each Y_{mn}^{k}, we have

$$\begin{aligned} & p \ge 1 \\ & v_i^B = v_m, i = 1 \\ & v_{i-1}^E = v_i^B, p \ge i \ge 2 \\ & v_i^E = v_n, i = p \end{aligned} \tag{7}$$

These constraints are considered to ensure each wagon block is finally marshaled into trains.

(4)Train transfer headway constraints, for any Y_{mn}^{k}

$$a_i^u + Z_{v_i^E}^t \le d_{i+1}^l,\ i=1 \ldots p\text{-}1 \tag{8}$$

(5)Flow balance constraints

$$O_I + \sum_{m=1}^{p} \sum_{n=1}^{p} \sum_{k=1}^{K} \sum_{i \in T, v_i^E = I} y_{mn}^{k,i} N_{mn}^{k}(Q) = \sum_{m=1}^{p} \sum_{n=1}^{p} \sum_{k=1}^{K} \sum_{i \in T, v_i^B = I} y_{mn}^{k,i} N_{mn}^{k}(Q) + D_I \tag{9}$$

Numerical Experiments

The above model is a nonlinear integer and multi-objective programming problem. It is a combination optimization problem. By considering the complexity of the above model, normal algorithms are too slow or not capable enough to come up with good enough solutions. Thus, heuristic algorithms are needed to solve instance problems of real-world size. Although there are quite a few meta-heuristic methods that could be used to solve NP problems, genetic algorithm has its own advantages. Genetic algorithm is a search technique used in computing to find exact or approximate solutions to optimization and search problems. The major advantage of genetic algorithms is their flexibility and robustness as a global search method. In addition, genetic algorithms are parallel because they have multiple offspring thus making it ideal for large problems where evaluation of all possible solutions in serial would be too time taking. Furthermore, they perform well in problems where the fitness landscape is complex, where the fitness function is discontinuous, noisy, changes over time or has many local optima. In this study, we design and implement a specific genetic algorithm based on analyzing the characteristics of the problem to solve the above model.

Figure 2 depicts a rail network for experiments which consists of 11 yards and 16 weighted sections. We designed two types of experiments. Experiment A is that the transport company will afford the punishment lonely caused by less-weight train operation, which means that the transport price is fixed, it will not change with the transport quality. The other one Experiment B is the contrast, that is, the transport price is variable and it will change with the transport quality.

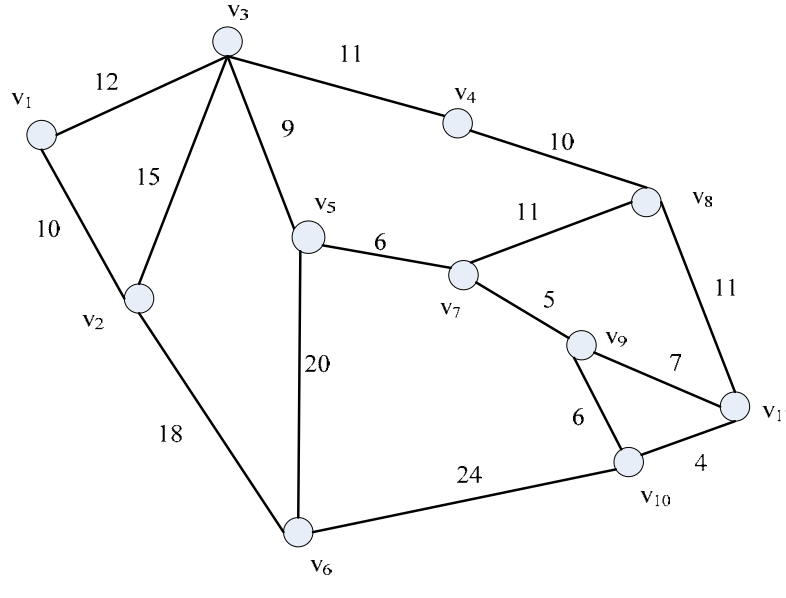


Figure 2. Illustration of rail network and weight of arcs in numerical experiments.

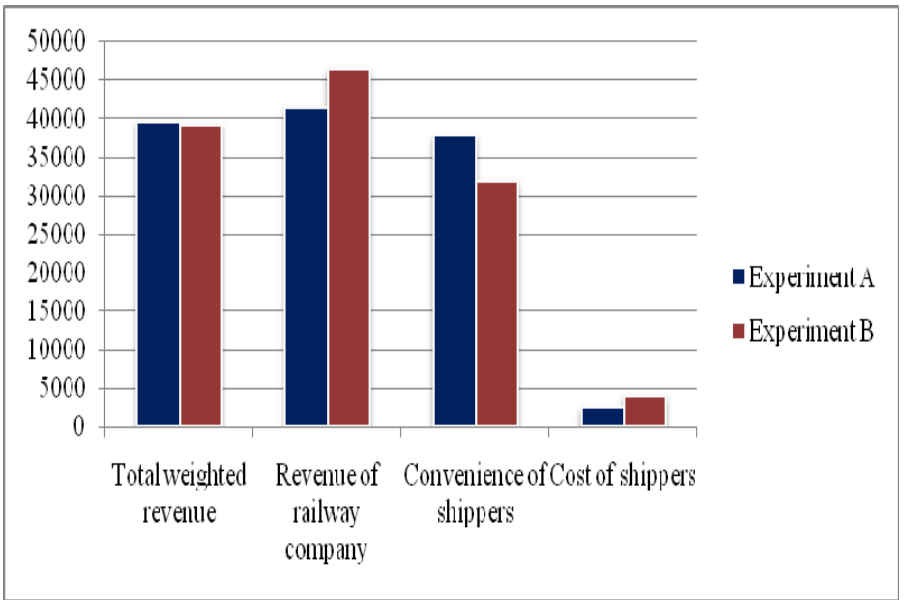


Figure 3. Comparison analysis of revenue and cost.

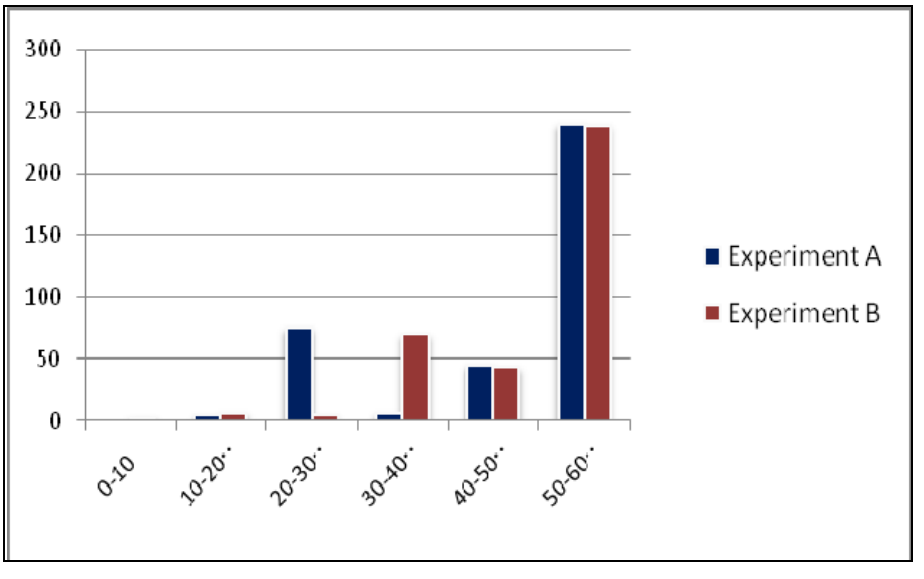


Figure 4. Comparison anaylisis of the number of wagons that trains contain.

Based on solving process, we could present the below comparative analysis according to the results from experiment A and B. Figure 3 shows that in experiment B, the revenue of railway company and the cost of shippers rise significantly with the reduction of the convenience of shippers. Figure 4 shows the number of wagons that trains contain in experiment A and B.

Based on comparison analysis, we can conclude that, in the condition of variable transport price, the revenue of transport company will improve significantly, but the total fee of shipper will increase. Meanwhile, the probability of combining train blocks will increase, in result, this will reduce the convenience of shippers.

Conclusions and Future Researches

This paper aims to synchronize train blocking plan and train timetable. First, the research constructs a conceptual framework model for synchronizing train block plan and train timetable, based on the feedback mechanism between train blocking plan and train timetable via "quasi train path graph". Then, on the basis of this framework, a bi-level programming model and corresponding improved genetic algorithm are proposed. Finally, comprehensive numerical experiments are conducted to demonstrate the applicability of the method.

In the future, further studies should be made on identifying the weights of shippers and transport company in the objective function, calculation method of transport revenue and cost. With the increase scale of network, the solution algorithm efficiency should be improved for practical uses. In addition, how to build robust train timetable by means of "quasi train path graph" highly requires further studies.

Acknowledgement

The work described in this paper was supported by National Natural Science Foundation of China (No.60736047). The authors deeply appreciate the support.

References

Brannlund, U., Lindberg, P.O., Nou, A., and Nilsson, J.E., (1998). "Railway timetabling using Lagrangian relaxation." *Transportation Science* , 32(4), 358–369.

Cai, X., Goh, C.J., and Mees, A.I., (1998). "Greedy heuristics for rapid scheduling of trains on a single track." *IIE Transactions,* 30 (5), 481–493.

Cao Jiaming, and Zhu Songnian (1992). "Optimizing train blocking plan, 0-1 programming." *Journal of the China Railway Society*, 14(4), 49-57.

Cao Jiaming, and Zhu Songnian (1993)(a). "The 0-1 quadratic programming method for the single group train formation plan 0n railway network." *Journal of the China Railway Society*,15(2), 62-70.

Cao Jiaming, and Zhu Songnian (1993)(b). "Synsetic optimization of train formation plan on railway network." *Journal of Southwest Jiaotong University*, 5.

Caprara, A., Fischetti, M., and Toth, P. (2002). "Modeling and solving the train timetabling problem." *Operations Research,* 50 (5), 851–861.

Carey, M. (1994a). "A model and strategy for train pathing with choice of lines, platforms and routes." *Transportation Research Part B,* 28 (5), 333–353.

Carey, M. (1994b). "Extending a train pathing model from one-way to two-way track." *Transportation Research Part B,* 28 (5), 395–400.

Carey, M., and Lockwood, D. (1995). "A model, algorithms and strategy for train pathing." *Journal of the Operational Research Society,* 46 (8), 988–1005.

Carl van Dyke (2006). "Strategic Policies of Multi Modal Applied Systems." *MultiTrain.*

Jorg Bocker, Jurgen Lind, and Bernd Zirkler(2001). "Using a multi-agent approach to optimize the train coupling and sharing system." *European Journal of Operation Research*,134, 242-252.

Haiying Li, and Hao Yang (2007). "Study on The Theory and Methodologies of Synchronizing Train Blocking and Scheduling." Beijing Jiaotong University.

An Energy-Efficient Control Algorithm and Simulation of Urban Rail Transit Systems

Haidong Liu, Yun Bai, Jing Yang, Fangming Zhou, Baoshan Wang, Peng Du and Baohua Mao

School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China; Email: hdliu@bjtu.edu.cn

Abstract

Energy efficiency of trains is an important and beneficial task. According to the research conducted on energy-efficient control both by domestic and international professionals, it is studied that the energy-efficient algorithms of the train of Urban Rail Transit with fixed running time. The algorithms adjust the maximal speed of whole lines and minimal speed of sections, which are used as the limited speeds, according to a given operational time. At the same time, using a forward calculating algorithm makes the best use of coasting to reduce times and time of traction and braking, which helps to decrease energy consumption. A computer simulation system has been developed, and some cases have been designed and analyzed for train movement under different times and different loco-tractions. It provides a method for optimizing real control model of the train of Urban Rail Transit.

Key Words: energy-efficient; fixed-time movement; control algorithm; traction calculation; computer simulation

Introduction

With the growing domestic urban rail transit development, it has become more important for the urban rail transit train to be energy efficient. The train's energy consumption involves many factors, and improving the train's control method to achieve a more energy-efficient operation, under the conditions of certain traction and braking performance, traction weights, lines, and existing the operation plan, train marshalling state of operation and management plans, is a cost-effective and directly viable energy-saving way.

Compared with the inter-city trains, urban rail transit train has obvious differences, which are embodied in the train braking characteristic of the traction, train's traction weights, stop spacing, line condition and so on. For the urban railway, the train uses stepless traction, where traction weights change very little, stand a relatively short distance, line slope commonly used energy-efficient form. Therefore, the question of energy savings has its own particularity. If there is no need to consider handle position conversion, the short distance between stations and the setting of energy-saving slope are conducive to achieving traction-coast operating mode, the smaller changes of train weight make the timing computing of the train more efficient, etc., which brings about favorable conditions to the study of urban rail transit train energy-savings.

Researchers have examined possible strategies concerning train energy-savings at home and abroad. However, because of the complexity of the train operating environment, the energy-savings algorithm model and its solution technique is a very difficult task. Since the 1980s, Australia, Britain, Germany, Japan, the United States, and many other countries have researched the experiments about the train energy-efficient control, summing up energy-efficient train control methods using the microcomputer technology to develop the computer optimization manipulation of the train guidance system, computer simulation system and so on (Weidong and Zili 1997; Zili 1994; Howlett and Pudney 1995; Chang and Sim 1997; Wong and Ho 2004).

Based on the general train movement calculation system and the train energy savings and time accuracy research, this article has conducted the preliminary research to the train energy-saving movement system of urban rail transit (Mao 2000; Liu 2005; Ho et al. 2002; Ding et al. 2004). This article has considered energy-efficient and regular train control methods to design the corresponding algorithm for train operation, and take advantage of computer simulation, combined with some practical examples, to achieve the simulation results.

The Existing Energy-Efficient Running Process

According to the study at home and abroad, we can divide the trains' operation on the station (Weidong and Zili 1997; Zili 1994; Howlett and Pudney 1995; Chang and Sim 1997; Wong and Ho 2004). Therefore, the whole section of the energy-saving operation optimization problem can be transformed for the sake of saving interval between each station optimization problem. The model is as follows:

$$\min E = \sum_{i=1}^{N} E_i$$

$$E_i = f_i(T_i, V_i, C_i, L_i, TR_i, X_i), \qquad i = 1, \ldots N$$

$$\sum_{i=1}^{N} T_i = T$$

$$V_i = f_i(V_{\max}, V_{\min}, V_{站})$$

$$C_i = f_i(F, W, B)$$

where E_i is the energy optimization result of i section; T_i is for each station interval of time consumption; V_i is the train's limited running speed; C_i is for the manipulation of operating conditions; L_i is for the line conditions; TR_i is the conditios of the train; X_i is the other factors impacting energy consumption; N is the total number of intervals between stations; T is the total running time of intervals; $V_{\max}$ is the maximum allowable speed of trains, when no time constraint for the line limited speed, when there are time constraints then for adjusted limited speed; $V_{\min}$ is the minimum speed of the trains; V_s is the lowest-stop speed of the trains; F is traction work conditions for the train; W presents train coast control; B is braking conditions for the train.

As the factors involved in the process of energy-saving operation is very complex, the mathematical models to solve the above-mentioned energy-saving operation is a very difficult task. Therefore, based on the above-mentioned international and national studies, when we design the train energy-saving operation by computer algorithms, we can follow these principles: (1) using the largest traction in the acceleration process; (2) compute the parking brake system in the greatest power; (3) avoid using the air brake except for the parking brake; and (4) in the constrained co-ordination with both speed and coasting mode.

Based on the above principles, and according to literature study at the same time (Howlett and Pudney 1995), take the combination of the upslope, flat slope, and downhill as an example, the process of energy-efficient train operation is shown in Figure 1.

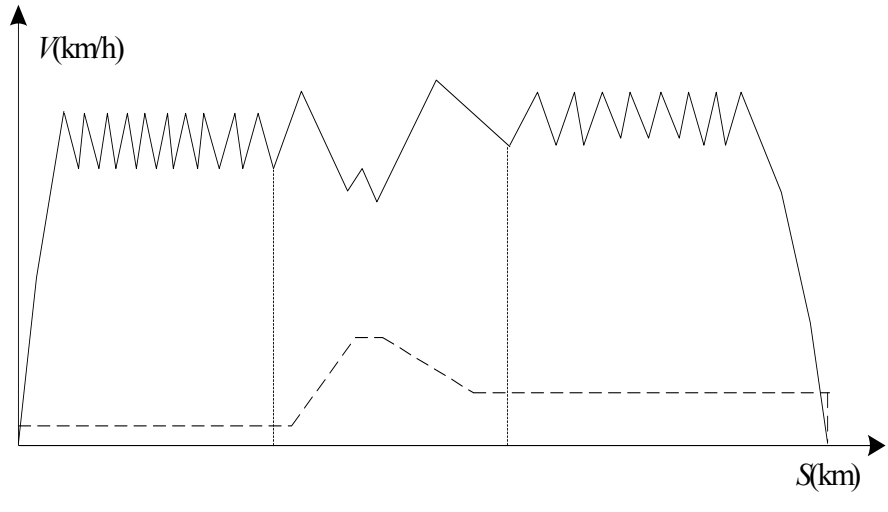


Figure 1. Train in the section on the ramp to run process.

As shown in Figure 1, the operation of the train is as follows: when the train runs on a straight ramp, the way to energy-savings is running in the target speed by traction-coasting pair; when the train runs close to a ramp, it will tract to accelerate the speed in advance, with sufficient kinetic energy to rush the slope. If the train slowed down when in the uphill traction, then take the advantages of traction to rush out of slope; if the train's speed continued to rise, when the speed has exceeded the target speed, coasting control should be used, when the rate dropped to a certain figure, use the traction. After the train drives out of uphill, use the coasting in advance to reduce the train's running velocity, in order to make full use of the potential energy of the ramp to coasting . If coasting speed does not surpass the limited speed on the downhill grade, then the train will coast to get down from the slope, otherwise running under braking. After the train leaves the slope, restore the traction coasting to the movement.

According to the above study, a train on single-uphill, single-downhill, or on a simple combination of uphill and downhill, the operation processes are relatively simple. However, the actual condition of the ramps is very complex, it is difficult to determine their control methods according to the above process. Therefore, combined with the above-mentioned operation process, the following energy-saving control method can be used: take the average speed of the trains as the target speed, and move with the target speed in pulling - lazy line in order to achieve the purpose of energy-savings. The specific process is as follows: in the train start-up phase, make full use of the maximum traction to tow. When running in the interval, take advantage of the coasting conditions as far as possible, and in order to save energy, avoid using the braking condition, when timing in the interval. If the train's target speed is less than the current limited speed, when the train's speed close to the target speed, the train must run under the traction - lazy line mode; if the target speed is faster than the current limited speed, then operate in the lazy line when the train is running near the limited speed, if the speed continues to increase, then run under braking. The train will brake under the largest force if the following point of speed-control curve is valued zero. The train's operating mode is shown in Figure 2.

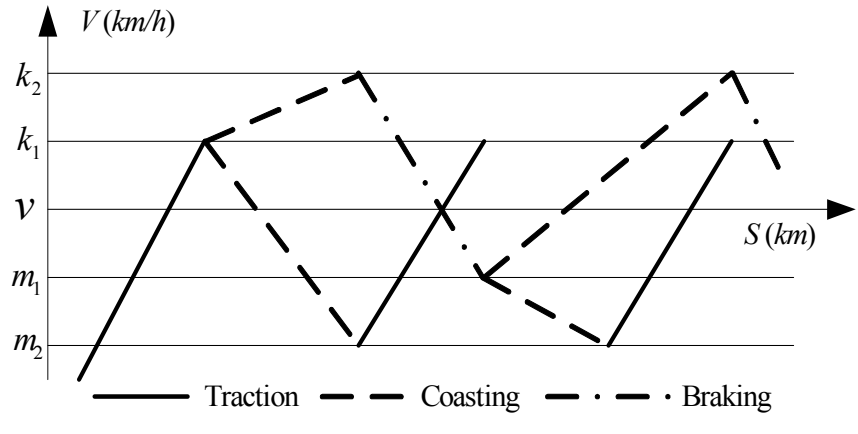


Figure 2. The train's energy-saving operating mode.

In Figure 2, v for the target speed, k_1、k_2 is velocity higher than v, m_1、and m_2 is velocity lower than v. The specific manipulation of the process is as follows: when the train's speeds have reached or exceed the speed of k_1, the train will use the coasting mode, if the train is on the downhill, the speed continues to increase, when its speed is greater than the value of k_2, then use a braking mode. In the downhill road, in order to prevent the frequent switch between coasting and braking, the critical speed from brake to coast could be valued under the object speed. Thus, when the train operating to speed is m_1 under braking, then change to the coasting, if the train's speed increases again, there are $m_2 \sim k_2$ of the coasting space; if the lazy-line speed of the train continued to decrease, when down to m_2 , converted to traction and continue to run.

Improved Energy-Efficient Algorithm

The typical mode of traction-coasting for running is used in the described algorithm where we can't obtain the reasonable breaking point from traction to coast. The algorithm also brings about excessive traction-coasting pair when the train is running, increases the checking brake and quadratic traction on a long steed downhill section to cause the increase of energy consumption. To avoid the shortcomings of the above algorithm, reduce the traction time. The improved algorithm adds the forecast of handle control, which uses the forward prediction to get the coasting point. The forward prediction can make the times of traction and traction time lower, get the least traction-coasting pair, and make the coasting speed be close to the maximum allowable speed.

An heuristic approach is used for the improved energy-efficient algorithm. The model uses the following assumption:

(1) The maximal speed of line is V_{max}, the minimum running speed of the train is V_{min}, and the minimum stop of the train at station is V_s.

(2) The maximum acceleration of the train when starting is a_{max}, when it's running in the block, the maximum acceleration of the train is a, and the maximum deceleration is b_{max}.

(3) The train should start with the maximum traction, and its speed should be smaller than a_{max}.When the running speed is up to V_{min} , we should use the forward prediction . The details are as follows:

Firstly, use the traction handle to push forward the train in one time step ; then use the coasting handle, if the forward speed of coasting is no more than the limited speed. Choose the traction handle in the next time step , or continue to use the coasting handle .

(4) If its speed is below V_{min} when the train is coasting , the train turns to traction.

(5) To make sure that the coasting speed doesn't exceed the limited speed, use braking when the two speeds are close if the coasting speed of the train is on the increase.

(6) In the process of forward coasting, if the speed of the train when coming into the station is below V_s, the train turn into traction in advance to make sure the speed exceeds V_s.
(7) Apply the stop algorithm when the train is near the station. If the train meets the stop error, use braking to make the train stop, and make sure the braking deceleration is below b_{max}.

The train will travel as close to the limited speed as possible in the block, so the runtime for energy efficiency in the block is the minimum runtime under given conditions.

Energy-Efficient Algorithm under Timing

The energy-efficient algorithm under timing, which is more applicable in the real operation for the train, should satisfy the scheduled time strictly and is described as follows:
(1)The limited speed of line is V_{max}, the minimum running speed in the block is V_{min}, and the minimum stop of the train at station is V_s.
(2) The maximum acceleration of the train when starting is a_{max}, when it's running in the block , the maximum acceleration of the train is a_q and the maximum deceleration is b_{max}. Make sure that the acceleration don't exceed a_q and the deceleration don't exceed b_{max} while train is running.
(3) Time constraint, the time train spends in a block is a fixed value T;
(4) Compute the energy efficient simulation using the maximum traction and braking and making the speed fall between (V_{min},V_{max}) and obtain the runtime in the block T_{run}.
(5) Time error is supposed to be T_e. When the difference between T_{run} and T meets T_e, turn to step (7), or adjust the value of V_{max} and V_{min} based on the difference.
(6) If $T > T_{run}$, let V_{max} be smaller, and make the adjusted V_{max} be the maximum speed, or let V_{min} be larger, and make the adjusted V_{min} be the minimum speed. Turn to step (4);
(7) Output the results, the simulation ends.

There are many factors that have an effect on the energy-efficient algorithm under timing .To meet the time error and get the desired result, it is preferred that the system runs many times. So it's very important to improve the efficiency.

The Process of Train Running and Algorithm

Use the above improved energy-efficient algorithm. The running train for energy-efficiency is shown in Figure 3.

Figure 3 shows us that the train is running with the maximum traction when starting. Judge the position where the running speed of train is close to V_{min}, and make traction turn into coasting . If the coasting speed exceeds V_{max}, make coasting turn into braking, or turn into traction if the speed is below V_{min}. Use braking to make the train stop if the train meets to the condition of braking.

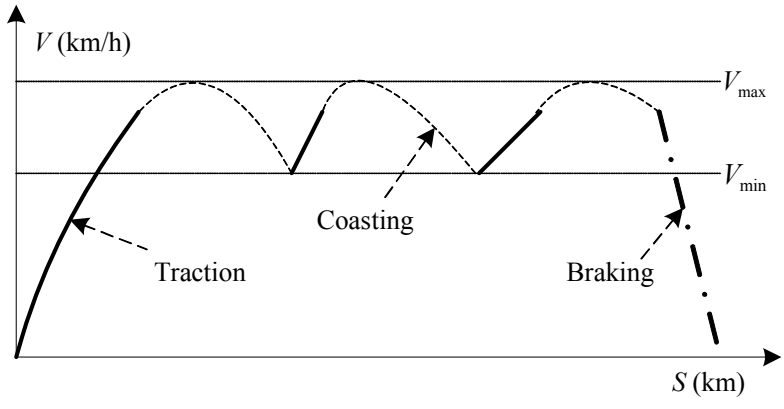


Figure 3. Running train for energy-efficiency.

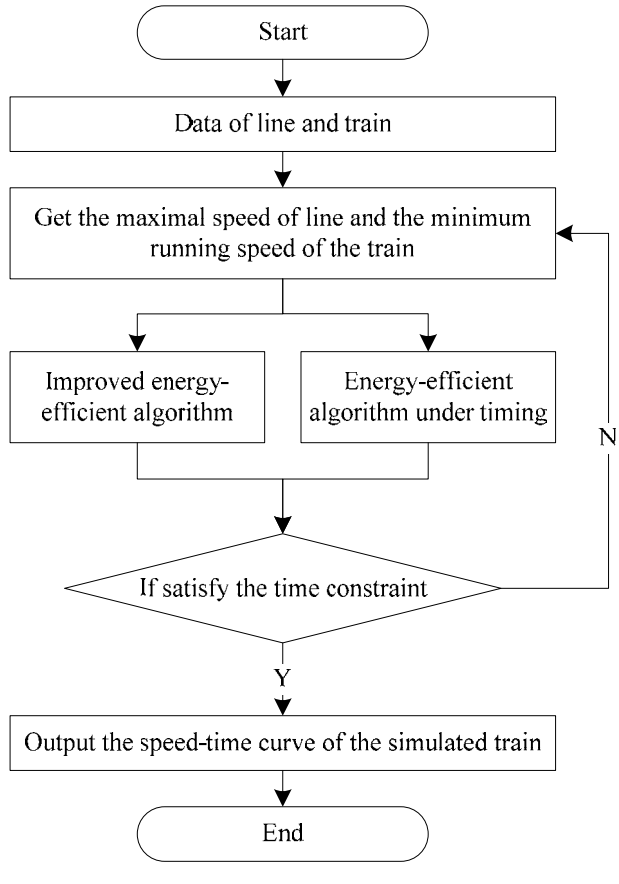


Figure 4. The flow about the energy-efficient of the running train.

Figure 4 shows us the flow about the energy-efficient of the running train. The process is as follows:

Firstly, the data of line and train should be acquired for getting the stress state of the train when running. Then we should get to know the maximal speed of line V_{max} and the minimum running speed of the train V_{min}.

Secondly, calculate the train operation curve using the improved energy-efficient algorithm and the Energy-efficient algorithm under timing. Output the speed-time curve of the simulated train if the running time satisfy the time constraint. Or else adjust the value of V_{max} and V_{min} according to the given time, then recalculate until the running time meets the time constraint.

Simulations and Results Analysis

Energy-efficient running strategies have been designed based on the algorithms discussed above, a simulation system has been established, and field data has been used to simulate. The data used to simulate is as follows:

Table 1. Line data.

Line name	Station count	Line length(m)	Line limit speed(km/h)
line1	20	22465	80
line 2	27	30763	80

Table 2. Train data.

Train name	Train weight(t)	Maximal speed(km/h)
D1	150	80
D3	280	80

Table 3. Simulation data.

Speed (km/h)	Maximal accelerate (m/s^2)	Maximal decelerate (m/s^2)
60～80	0.83	1

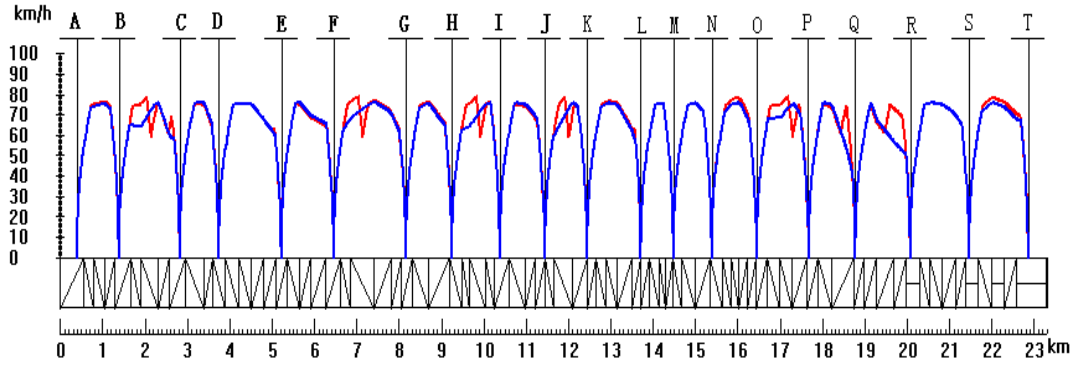


Figure 5. Speed-distance comparing curve of energy-efficient running on line 1.

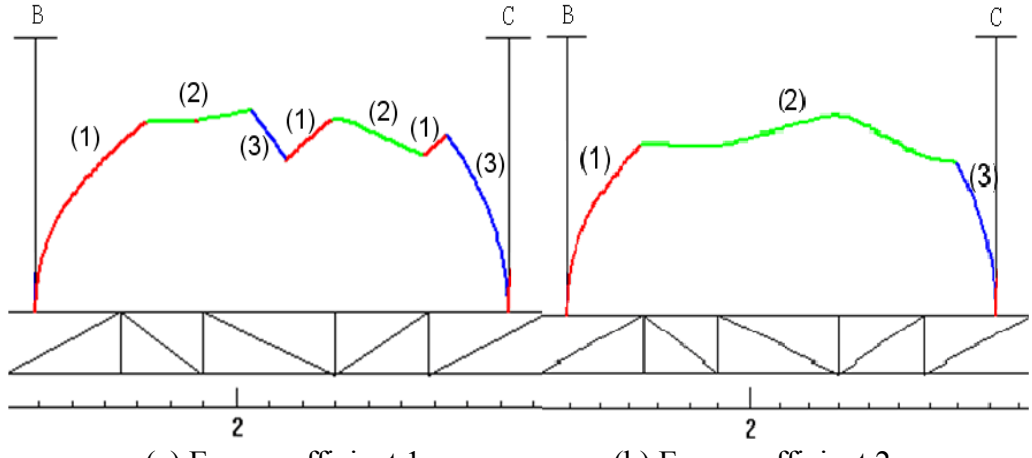
Firstly, train D1 is selected to simulate using two energy-efficient schemes based on data of line 1. Figure 5 showed the speed-distance curve of train running.

As shown in Figure 5, three situations can be obtained from the tracks of two energy-efficient curves between sections.

(1) The two energy-efficient curves in Figure 5 are basically in coincidence. In most of the sections, like A-B, C-D sections, the two energy-efficient curves are basically in coincidence, and manipulate modes, that using a traction-coast handle to accomplish the traction between sections, are basically the same. Because the distance between these sections is short, and gradient is slow, mostly under 10‰. Trains can accomplish the running using one traction-coast handle, moreover, under the limit speed.

(2) The two energy-efficient curves in Figure 5 are different, and there is a braking in curve 1. The traction-coast handle is still used in curve 2, like B-C, F-G sections. Figure 6 shows the speed-distance curve in the B-C section.

In Figure 6, (1) shows the traction curve segment, (2) shows the coast curve segment, and (3) shows the brake curve segment.



(a) Energy-efficient 1 (b) Energy-efficient 2

Figure 6. Energy-efficient running comparing in B-C section

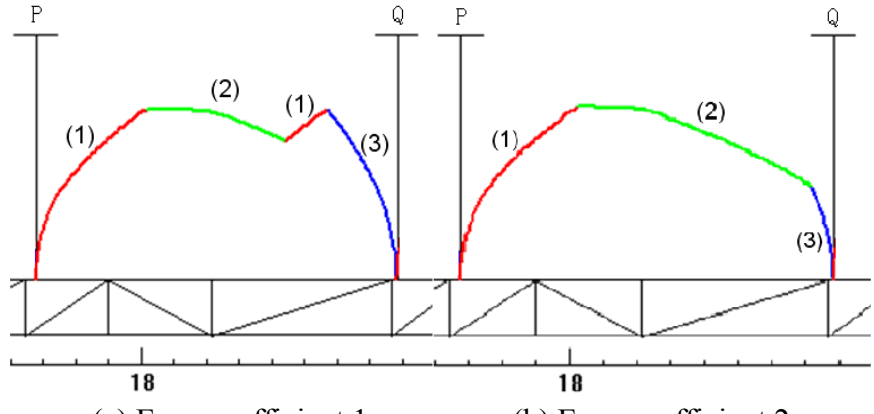
Table 4 shows the grade data in B-C section.

Table 4. Grade data in B-C section.

No.	Start point (m)	Length(m)	Grade(‰)
1	1290	360	3
2	1650	250	-3
3	1900	400	-23.4
4	2300	280	30
5	2580	370	3

As shown in Figure 6(a) and Table 4, the maximum steep downhill track in B-C section is grade No. 3 with a grade of 23.4%, and the brake handle is used to avoid exceeding the limit speed after a traction-coast handle. After that, the traction-coast handle is used again when the speed has been reduced. The brake handle consumes a lot of energy. Compared to Figure 6(a), Figure 6(b) has been using forward deduce at the beginning of traction based on the scope and train situation, and choosing suitable turning point from traction to coast, so the traction times of train between sections are reduced. To avoid exceeding the limit speed, the speed of turning point from traction to coast is 65.02 km/h in Figure 6(b) and is slower than the corresponding speed in Figure 6(a). In Figure 6(b), forward deceleration is used to control the speed under the limit, so the train speed will be under the limit in the following coasting, and only one traction-coast handle is used in the whole section. Comparing running time and energy consumption use in two energy-efficient schemes, the running time and energy is 95.21s and 17.93kWh separately in scheme 1, and 97.81s and 10.46 kWh in scheme 2. Scheme 2 saves 41.7% more energy than scheme 1.

(3) Two energy-efficient curves are different, and there are less traction-coast couples in curve 2. Some sections, like O-P，P-Q section, have this situation. Figure 7 shows the speed-distance curve in the P-Q section.
In Figure 7, (1) shows the traction curve segment, (2) shows the coast curve segment, and (3) shows the brake curve segment.



(a) Energy-efficient 1 (b) Energy-efficient 2

Figure 7. Energy-efficient running comparing to the P-Q section.

Table 5 shows the grade data in P-Q section.

Table 5. Grade data in P-Q section.

No.	Start point (m)	Length (m)	Ggrade (‰)
1	17750	250	3
2	18000	320	-2
3	18320	550	28.73
4	18870	230	3

As shown in Figure 7(a) and Table 5, the maximum steep downhill track in the P-Q section is grade No. 3 with grade of 28.73‰. A traction handle is used after a traction-coast because train speed is under the shield. There are two traction handles in this situation, which cost more energy. Compared to Figure 7(a), Figure 7(b) has been using forward deduce at the beginning of traction based on the scope and train situation, and choosing suitable turning point from traction to coast, so the traction times of trains between sections are reduced. To avoid slow speed in long and steep ramps, the speed of turning point from traction to coast is 75.82 km/h in Figure 6(b), which is higher than the corresponding speed in Figure 7(a). In Figure 7(b), advance acceleration is used so the train speed will be higher than the minimum speed in the following coasting, and only one traction-coast handle is used in the whole section. Comparing running time and energy use in two energy-efficient schemes, the running time and energy-consumption are 78.39 s and 17.05 kWh in scheme 1, and 81.28 s and 12.95 kWh in scheme 2. Scheme 2 saves 24.0% more energy than scheme 1.

Then, train D3 is selected to simulate using two energy-efficient schemes. Table 6 shows the whole-line running results of train D1 and train D3.

Table 6. Results of energy-efficient running.

Section	Distance (m)	D1				D3			
		Energy-efficient 1		Energy-efficient 2		Energy-efficient 1		Energy-efficient 2	
		Time(s)	Energy (kWh)	Time(s)	Energy (kWh)	Time(s)	Energy (kWh)	Time(s)	Energy (kWh)
A-B	993	71.83	12.06	72.13	11.82	72.24	21.43	72.38	21.13
B-C	1432	95.21	17.93	97.81	10.46	94.78	36.20	97.8	19.07
C-D	924	68.66	12.39	68.35	12.65	69.62	22.63	69.35	23.24
D-E	1472	96.50	12.77	96.57	12.14	97.55	22.84	96.69	21.86
E-F	1239	86.03	12.35	85.44	12.60	85.87	22.60	85.56	22.90
F-G	1705	107.43	14.85	108.19	9.44	107.82	25.53	108.86	15.94
G-H	1091	76.45	12.20	77.35	11.97	77.17	21.83	77.17	21.83
H-I	1132	79.04	15.19	80.16	9.71	79.39	26.34	80.87	16.39
I-J	1061	75.62	12.61	75.2	12.87	75.95	23.60	75.07	24.20
J-K	1000	72.34	15.05	73.44	9.18	72.56	25.82	72.97	15.68
K-L	1268	86.06	12.93	86.68	12.08	86.39	23.64	86.3	22.09
L-M	768	60.14	12.58	60.65	12.58	61.12	23.62	61.59	23.92
M-N	908	67.00	12.20	67.00	12.20	67.66	21.77	67.55	22.08
N-O	1058	73.75	12.05	74.63	11.41	74.12	21.08	75.26	20.15
O-P	1228	83.22	16.88	84.72	11.30	84.17	29.93	84.19	21.53
P-Q	1106	78.39	17.05	81.28	12.95	79.30	30.71	81.32	24.65
Q-R	1298	89.45	17.89	94.56	14.72	89.76	34.25	89.69	35.16
R-S	1383	90.76	12.28	91.27	12.28	92.01	22.13	91.69	22.43
S-T	1399	90.42	12.16	91.92	11.51	91.74	21.79	93.08	20.55
Total:	22465	1548.32	263.42	1567.34	223.86	1559.22	477.73	1567.41	414.81

As shown in Table 6, the running time of train D1 is 1567.34 s in scheme 2, 19.02 s more than 1548.32 s in scheme 1; the energy-consumption of train D1 is 223.86 kWh in scheme 2, 39.56 kWh more than 263.42 kWh in scheme 1, and 15.0% energy savings than scheme 1; the running time of train D3 is 1567.41s in scheme 2, 8.19s more than 1559.22s in scheme 1; the energy-consumption of train D3 is 414.81 kWh in scheme 2, 62.92 kWh more than 477.73 kWh in scheme 1, and 13.2% energy savings than scheme 1. Based on the comparison between train D1 and D3, we understand that the running time of trains D1 and D3 are almost the same. The traction weight of train D1 is heavier that of train D3, but train D3 gets much traction than D1, so the running speed of D3 is at the same level of that of D1. Moreover, using the same control mode, the energy-savings of train D1 and D3 are 15.0% and 13.2% separately, which achieve the energy-efficient goal.

Finally, train D1 and D3 are selected for simulation based on the data of line 2. The speed-distance curve and whole-line running results are shown separately in Figure 8 and Table 7.

As shown in Table 7, the running time of train D1 is 1567.34 s in scheme 2, 27 s more than 2125.88 s in scheme 1; the energy-consumption of train D1 is 331.94 kWh in scheme 2, 28.57 kWh more than 360.51 kWh in scheme 1, and 7.9% energy savings than scheme 1; the running time of train D3 is 2147.92 s in scheme 2, 4.56s more than 2143.36 s in scheme 1; the energy-consumption of train D3 is 640.03 kWh in scheme 2, 37.56 kWh more than 677.59 kWh in scheme 1, and 5.5% energy savings than scheme 1. Based on the comparison between train D1 and D3, the running time difference between two schemes of train D3 is 4.56s, and average energy savings percent is 5.5%, lower than 7.9% of train D1.

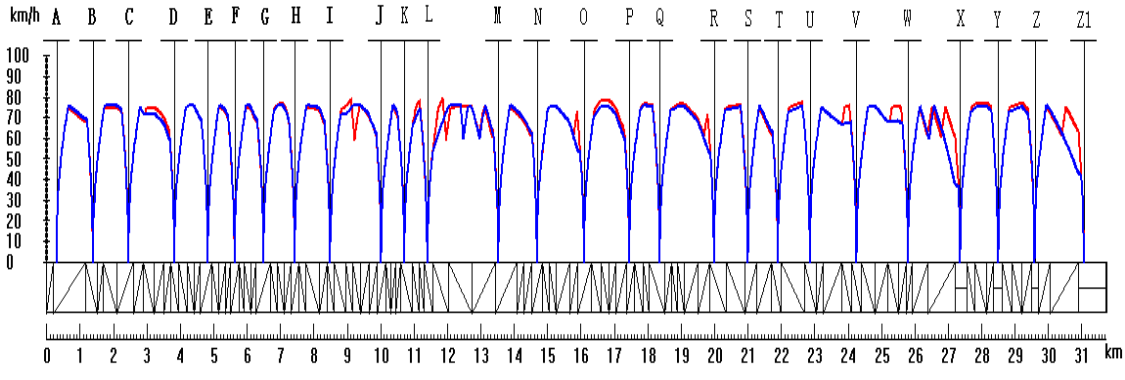


Figure 8. Speed-distance curve comparison on line 2 used energy-efficient handle.

Table 7. Running results of line 2 used energy-efficient handle.

Line	Distance (m)	D1				D3			
		Energy-efficient 1		Energy-efficient 2		Energy-efficient 1		Energy-efficient 2	
		Time(s)	Energy (kWh)	Time(s)	Energy (kWh)	Time(s)	Energy (kWh)	Time(s)	Energy (kWh)
Line 2	30763	2125.88	360.51	2152.88	331.94	2143.36	677.59	2147.92	640.03

As a conclusion, the average energy-saving percent is 10.4% of two trains on two lines.

Conclusion

Using energy-savings study of home and abroad for reference, based on the general calculation system of train running, this paper has done research on the energy-efficient algorithm and train running simulation under the restriction of given time. Using simulation method, this paper gets the running results of two energy-efficient schemes simulated on different lines, traction weights, and multiple unit conditions. Based on analysis of running results, scheme 2 is more reasonable on the train control than scheme 1, and the energy-savings percent 10.4% larger than that of scheme 1. This paper proposes advanced heuristic algorithms based on the existing energy-efficient strategy, simulates using the field data and obtains satisfactory results. However, because of the complexity of a real problem, more work should be done to improve and modify the algorithm based on the train running condition.

Acknowledgement

This research was funded by the National Natural Science Foudnation of China, (No. 60634010).

References

Jin, W., and Wang, Z. (1997). “Study on optimization method of train operation for saving energy.” *Journal of the China Railway Society*, 19(6):58-62.

Wang, Z. (1994). “An optimal operating method of train running for saving energy [J].” *Journal of Southwest Jiaotong University*, 29(3), 275-279.

Howlett, P. G., and Pudney. P. J. (1995). “Energy-Efficient train control. Springer, London.

Chang, C. S, and Sim, S. S. (1997). “Optimising train movements through coast control using genetic algotithms.” *IEE Proc- Electr. Power Appl.,* 144(1), 65-73.
Wong, K. K., and Ho, T. K. “Coast control for mass rapid transit railways with searching methods [J].” *IEE Proc. Electr. Appl.* ,151(3), 365-376.

Mao, B., He, T., Yuan, Z., Liu, H., Zhao, L.,, and Chen, Z. (2000). “A general-purposed simulation system on train movement” *Journal of the China Railway Society,* 2000, 22(1), 1-6.

Liu, J. Ding, Y., and Liu, H.. (2005). “Multi-train movement simulation system for urban rail transit [J].” *Journal of Transportation Systems Engineering and Information Technology*, 5(1), 79-82.

Ho, T. K. Mao, B. H. Yuan, Z. Z., Liu, H. D., and Fung, Y. F. (2002). "Computer simulation and modeling in railway applications" *Computer Physics Communications*, 143, 1-10.

Ding, Y. Mao, B. Liu, H. Wang, T., and Zhang, X. (2004). "An algorithm for energy-efficient train operation simulation with fixed running time." *Journal of System Simulation,* 16(10), 2241-2244.

A Method to Correct the Mileage Error in Railway Track Geometry Data and Its Usage

Haifeng Li[1] and Yude Xu[2]

[1] Associate Professor, [2] Professor, Department of Urban Rail Transit and Railway Engineering, School of Transportation Engineering, Tongji University, China, 4800 Cao'An Highway, Shanghai, CN 201804; PH: (+86) 21-69583721; FAX (+86) 21-69589880; Email: haifengli@tongji.edu.cn, xuyude2000@gmail.com

Abstract

With the novel technologies used in railway track status inspection, more and more track geometry data can be obtained, and they play a great role in ensuring track in good conditions. But, for the same section of the same track, there are also mileage errors between the track geometry data from the different inspection dates, so, in order to eliminate these inverse effects on the analysis results and avoid non-effective track maintenance works caused by the above mistaken data, a new optimization model to correct the mileage error of the track geometry data is put forward in this paper, and a automatic mileage error correction software based on the above model is developed at the same time. Based on a large amount of track geometry data from Shanghai-Nanjing railway line, the model verification and analysis are completed by the above software efficiently, the model is proved to be correct and effective.

Key Words: railway; track status; track geometry data; mileage correction; optimization method; software system

Introduction

Track geometry status has great influence on the safety and comforts of railway train, so the inspection and management of track geometry is paid more and more attentions. The current inspection ways of track geometry include man-made measurement, trolley measurement (static) and inspection car dynamic measurement etc. Though the previous two ways are limited to measurement efficiency and can not meet all the inspection requirement of rapid trains, they are still the effective aided ways; whereas for the track inspection car, it can measure the track geometry status with rapid speed, high efficiency, so this method and its measured data have

been accepted widely and used extensively. But there are some differences between the track geometry data from the above various inspection ways, even for the same inspection tool and for the same section on the same railway line, there are still some difference between different inspection dates. The difference not only shows on the numerical magnitudes of the various track geometry items (for example, profile, alignment, gauge, etc.), but also on the mileage error (WANG Qin, XU Yude and LI Haifeng, etc., 2006). It must be pointed that, a certain extent difference on the numerical magnitude of track geometry items can be accepted, and due to the influence of the track structure characteristics and its work property, the change of the numerical magnitude is limited; but the mileage error is not permitted for the data effective usage in the future. Such mileage error not only shows on the difference between the inspection mile and the actual mile, which may affect the accurate determination of the mileage position for the track inspection data and delay the maintenance work for the severely damaged sections of railway track; but also, shows on the difference between continuously inspected track geometry data at different dates, which are caused by the accumulated errors from the inspection principles, inspection speed, and some system errors of the aided equipments (for example, GPS etc.) etc. and this will cause great influences on the reasonable usage of such track geometry data and the correct evaluation of track geometry status, even on the safety of railway train running.

Considering the above inverse conditions of track geometry data, more and more engineers or researchers, who works on track geometry inspection and data maintenance and management, adopt various methods to reduce the influence from the mileage error. Literatures investigation shows that, there two ways to correct such mileage error, one is to adopt new techniques or new inspection ways, for example, inspection principles improvement or new inspection car invention in Japan, German, Italia etc. (Deng Xutong and Ye Yiming, 2004, Zhao Guotang, 2001) and GPS installed in inspection car in China (LI Qinghong, WANG Qin and YANG Yuanyuan, etc., 2002), the other is to eliminate the mileage error by mathematic models before track geometry data are used, for example, track geometry data processing software with mileage error correction function whose principles are still not reported openly (IMAGEMap, 2004), and LI Lijun's method based on curve characteristics which was introduced and used in the master thesis by YE Yong (YE Yong, 2006).

Further investigation shows that, for the above mathematic ways to eliminate the mileage error, there are still some deficiencies, for example, track geometry data post processing software used in German and China can only done once for the whole inspection data one time and all the data will have to take the same mileage displacement after mileage error corrected, but in fact, different sections have different mileage error displacement; and the mileage error correction based on curve characteristic can only process the mileage error in curve section and can not process the mileage error for the data between curves. So, a new method to calibrate the above mileage error by mathematic model will be introduced and analyzed in this paper, and it can solve the above two problems.

Mileage Error Correction Model

Large amount of track geometry data analysis shows that, from a long time point of view, track geometry status will be deteriorated with the traffic volume increased. But such deterioration process is very slow, if no maintenance work is done, the change trend of track geometry irregularity may be kept relatively stable. So, to make mileage error correction for the track geometry data from a certain period is feasible.

To get two adjacent track geometry data and select the track geometry data from a specified section (for example, 1km in length), and calculate the different value one point by one point, and then change the mileage position and repeat the above calculation process, after all the difference value corresponding to all the possible mileage displacement were calculated, the best mileage error correction value of the track geometry data will be determined finally. This method can be concluded to be an optimization problem on the correction value of x, that is, the minimum of the sum of squares of the difference between the two selected track geometry data in a certain section. The optimization model can be expressed by the following formula 1.

$$\begin{aligned} &\min u = \sum_{i=1}^{N} (Y_{1i} - Y_{2(i+x)})^2 \\ &s.t. \ -200 \le x \le 200, x \in Z \end{aligned} \tag{1}$$

where,

Y_{1i} is the previous track geometry data (or, reference data) of the i point;

$Y_{2(i+x)}$ is the next track geometry data (or, to be corrected data) of the $i+x$ point;

x is the mileage displacement distance while mileage error correction;

N is the total points number for the selected section to be calculated.

Mileage Error Correction Procedures

In order to secure the effectiveness of the mileage error correction for the track geometry data, the procedures of mileage error correction is divided into three steps, that is, the pre-processing of the track geometry data, the building of reference data library for mileage error correction, and correcting data mileage error etc., and this procedure can be described by Figure 1.

It must be explained that, the function of track geometry data pre-processing is to eliminate the obvious errors which may be existing on the track geometry data, such as data zero drift and data lost in some points etc. so as to ensure the reality of the inspected data source; The building of reference data library is to select one set of track geometry data to be the base of data mileage error correction, and it must be built at first according to some principles (for example, time first etc.) and refreshed by the inspection track geometry data if the maintenance works are done on site, so as to keep the consistency of the reference data library in the process of data mileage error correction.

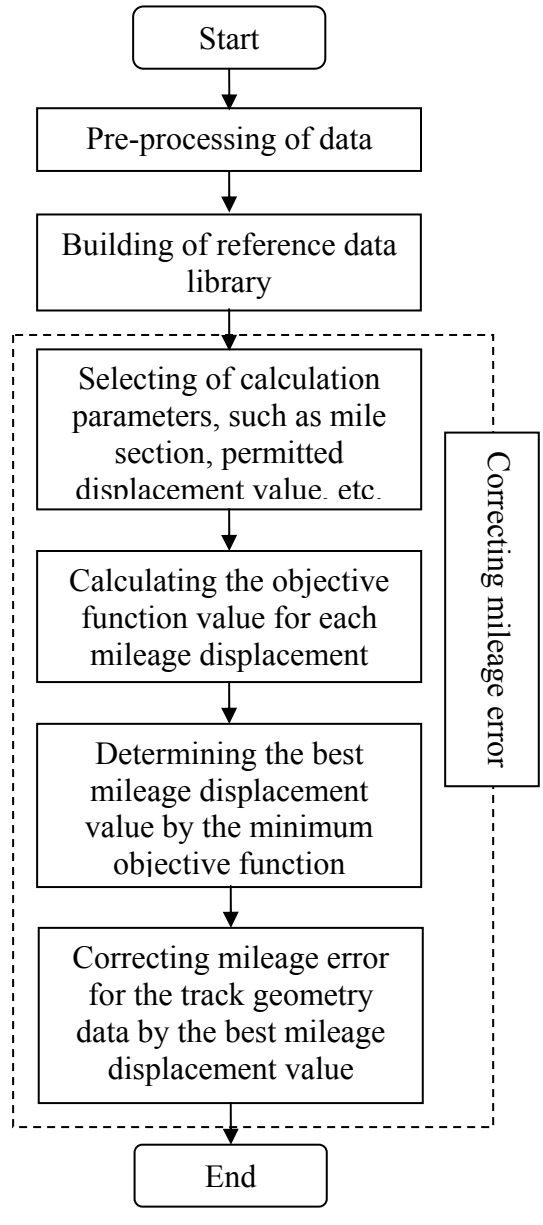


Figure 1. Flow chart of track geometry data mileage error correction.

Verification and Analysis of Mileage Correction Model

In order to increase the efficiency of large amount of track geometry data mileage error correction, a software system is developed by Visual Basic 6.0 based on the above principles, and large amount of track geometry data obtained from inspection car are used to verify and improve the model.
Due to the length limitation, a practical data and its corresponding calculation results are illustrated in the following.

Data Source. Track geometry data are inspected on the railway line from Shanghai to Nanjing on 2008.05.08 and 2008.07.05 respectively, and the data inspected on

2008.05.08 are used as a reference data library and the data inspected on 2008.07.05 are to be corrected. The data from K571+000 to K572+000 are selected to involve the following calculation and verification.

Results. taking the left profile, which is one of the track geometry characteristic expressing items, as a example, the objective functions calculated based on different mileage displacement are showed by Figure 2, and the track geometry data before and after the mileage error correction are illustrated by Figure 3 and Figure 4.

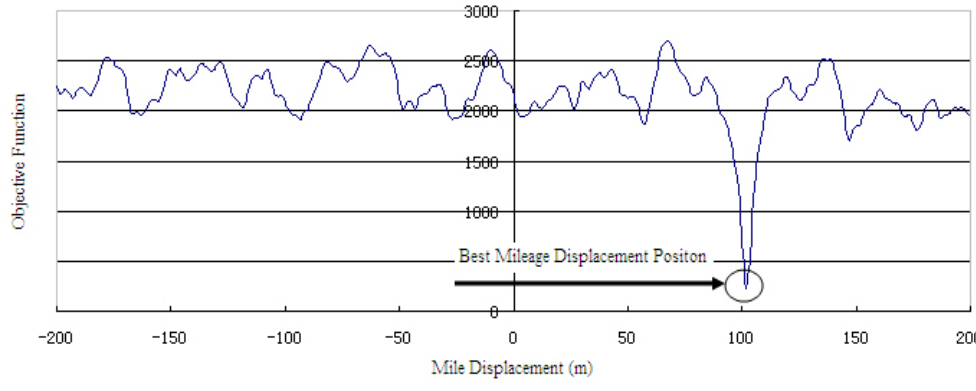


Figure 2. Relationship between mileage displacement and the objective function.

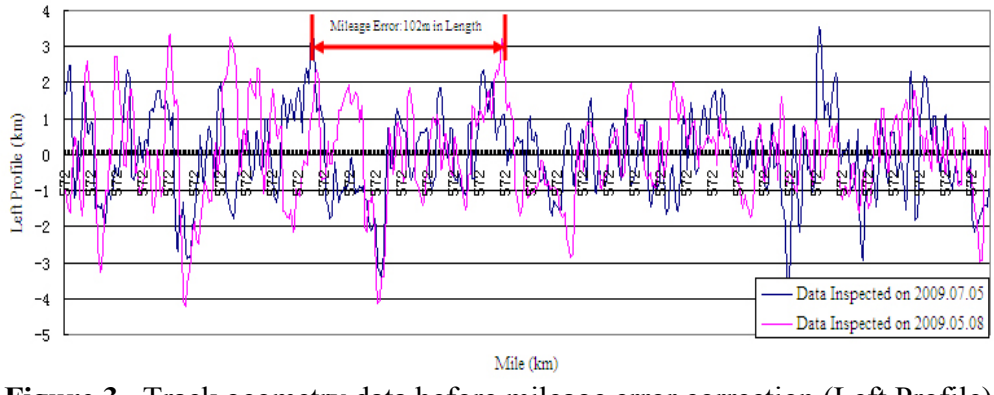


Figure 3. Track geometry data before mileage error correction (Left Profile).

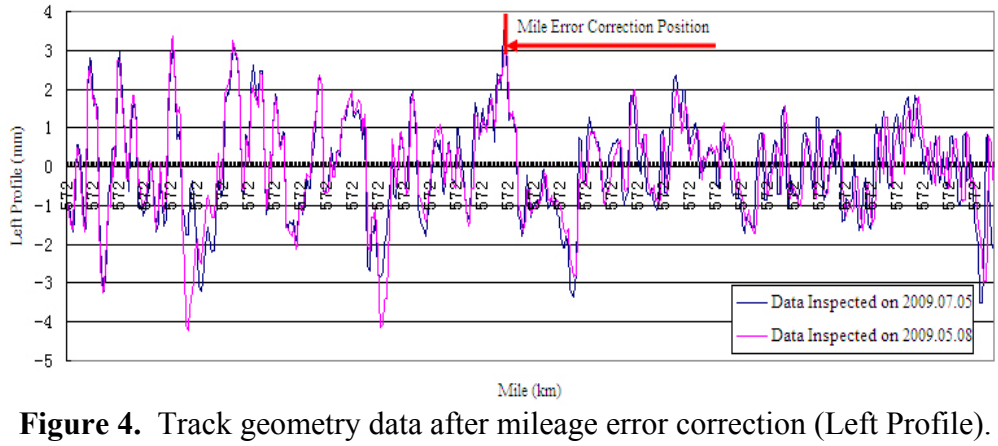


Figure 4. Track geometry data after mileage error correction (Left Profile).

It can be seen from Figure 2 that, when the mileage displacement is 102m, the objective function is up to the minimum, so the reasonable mileage displacement of the mileage error correction is 102m for this case. Comparing Figure 3 with Figure 4, it can be seen that, the selected track geometry data are almost coincident after the mileage error correction, that is, the data inspected on 2008.07.05 are moved 102m backward; but if there are no mileage movement, the selected data are different greatly.

Conclusion

Track geometry data is the foundation of track status evaluation and determination of maintenance works, the mileage error correction can secure the reliability and correctness of the data. A novel optimization model used to correct the mileage error in track geometry data is put forward in this paper and the corresponding software based on this model is developed. The calculation results based on large amount of track geometry data from inspection car show that the model is correct, reliable.

Acknowledge

Thanks for the support of National Natural Science Foundation of China (50878158)

References

DENG Xutong, and YE Yiming (2004). “Quasi-high speed track inspection car’s inspection principles and its application.” *China Railway Publishing House*, Beijing.

IMAGEMap (2004). “Laserail™ Profile and Geometry Measurement System WinDBC User's Guide.” *ImageMap company*, USA.

LI Qinghong, WANG Qin, and YANG Yuanyuan, etc. (2002). “Research on the increasing inspection ability and optimization work performance of track inspection car.” *Shanghai Railway Technology*. Shanghai, 1-3

WANG Qin, XU Yude, and LI Haifeng, etc., (2006). “Research on the track geometry status evaluation expert system.” *Department of Urban Rail Transit and Railway Engineering*, Shanghai.

YE Yong (2006). “Development and research of the integral application system on track inspected data.” *Master thesis of Beijing Jiaotong University*, Beijing.

ZHAO Guotang (2001). “Current status and development of track inspection car.” *China Railway Publishing House*, Beijing.

Train Departure Mode and Line Capacity in Inter-city Railway

Peng Du, Baohua Mao, Hui Zhao, Baoshan Wang and Liu Haidong

School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China; Email: pdu@bjtu.edu.cn

Abstract

Since more and more inter-city railways are being constructed in China, the operational management of inter-city railways is becoming much more important than ever. Based on the infrastructure and operational requirements of inter-city railways, the relationship between train departure mode and line capacity is discussed. Train departure mode is the basis of cyclic timetable development, which must suitable for the operation of inter-city railways. Factors affecting period of train platoon are analyzed and the calculation method of minimum period is constructed. The composition of different types of train platoons is discussed under the contradiction between the usage of line capacity and the level of operational service, which provides a method for the timetable development. Next, the design scheme of a soon-to-build inter-city railway is analyzed by developing cyclic timetables by using the proposed method and the computational results show that in this case study, the designed line capacity is inadequate to operational requirements in the long run. Finally, suggestions are given and improvements of the new design scheme are verified by using this method as well.

Key words: inter-city railway; line capacity; overtaking scheme; periodic timetable

Introduction

Intercity railways are the large mass rail transit systems which usually built in the developed areas or metropolitan areas connecting several cities or districts. Today, there are numbers of inter-city railways have been built or under construction, such

as Beijing-Tianjin Inter-city Railway in Bohai Ring Area, Shanghai-Nanjing Inter-city Railway in Yangtze River Delta and Dongguan-Huizhou Inter-city Railway in Pearl River Delta. These lines are all double-tracked, and trains run on the line can be either high-speed, or normal speed, or both. The overall travel time of the line is about one to one and a half hours. Inter-city railways are designed to provide fast and convenient transit service to passengers traveling between neighbor cities or different districts within a large metropolitan. They have their unique operational characteristics which are different from either metro, where train frequency is high, or railway trunk lines operated by Ministry of Railway, where there are not only passenger trains but also freight trains and train speeds vary greatly. Since there will be many inter-city railways latterly, the efficient and reasonable operation of inter-city railway is of great importance to both practitioners and scholars. As consideration of line capacity always plays an important role in train operation, this paper discusses the relationship between train departure mode and inter-city line capacity and presented a method of capacity calculation. A case study is conducted to verify the method.

Line capacity of Inter-city Railway

Thanks to sophisticated train control technologies, the minimum headway can be squeezed to about 2 to 3 minutes. However, different from common metro systems where there is only one kind of trains, there may be several kinds of trains running on inter-city railway lines which have different speeds. Based on characteristics of passenger flow, the train operation also has peak hours and normal hours during a day. Usually in peak hours, all kinds of trains are in operation, the number of trains reaches the maximum and the usage of line capacity is also the highest. Train diagram during peak hours is the most complicated and is the key to daily operation. Considering line capacity, peak hour train diagram can also be used to test whether the designed capacity is enough to cover the projected passenger volume.

Calculation of line capacity of a metro line is easy because there are no sidings at stations so all trains are running in line one by one without overtaking. The maximum line capacity can be described as the number of trains running through a section during a fixed time span, say one hour. This number is only related with the headway. While for an inter-city railway where there are sidings in several stations for overtaking between express trains and local trains, the calculation of maximum line capacity is not an easy task.

In network-based railway operation, the calculation of line capacity of double track line where there are many kinds of trains, the Method of Deduction Coefficient (MDC)[1] is the most commonly used one. While using MDC, deduction coefficients

of special kinds of trains are decided through comparison among running time of different kinds of trains and different arrangement of overtaking. Deduction coefficient of a kind of train means the number of normal trains to deduct when this certain kind of train is to be operated. The reference value of deduction coefficients can be achieved through large scale survey network wide by railway authorities.

MDC is so important, now it is still the primary method for capacity calculation of railway lines operated with trains of different kinds of speed. However, MDC is a macro level method and the reference values are results of statistical analysis, which suits capacity calculation under conditions of network based operation, but which may not suits the condition of inter-city railway, which is usually operated independently. For such independently operated railway line, analysis needs to be conducted specifically.

It is held that the value of deduction coefficient is affected heavily by the distribution of trains, i.e. the spatial relation among different kinds of trains. However this affection in network operation is hard to be analyzed quantitatively and thus can only described qualitatively. The situation is different in the case of inter-city railway, where there are only several train kinds, the overtaking mode is simple and the train distribution is finite as well. This situation makes it possible to study the affection between different kinds of trains using enumeration method. Ref [2] and [3] analyze the hour capacity and position of overtaking stations on high-speed railway line, which has a mixture of high-speed train and middle-speed train. This situation is similar to inter-city railway line except that the train departure mode does not resemble characteristics of mass transit. In this paper, we propose a calculation method of line capacity under different train departure mode, where there is a mixture of express trains and local trains and level of service quality is given consideration for inter-city railway lines.

In previous literature, the line capacity is express as the number of trains pass through a section within given time span. Different kinds of trains are translated by deduction coefficient to normal trains. Since we use enumeration method rather than MDC to study the line capacity, this expression of line capacity seems unsuitable. Therefore, instead of number of trains, we use time to represent line capacity. The capacity discussed here is defined as the time duration for the departure of given number of trains. The time duration scheduled is the Provided Capacity $C_{provided}$ and the minimum time duration needed is the Needed Capacity C_{needed}. It is obvious that different train departure mode will have different C_{needed} and if a certain C_{needed} exceeds $C_{provided}$ then this train departure mode is infeasible.

Train Departure Mode

In an inter-city railway line where there are both express and local trains, they depart alternatively to suit different passenger needs. We call a combination of these trains a platoon, and it is expressed as *N*+*M*, which means *N* express trains and *M* local trains depart in order. The period of a platoon is defined as the minimum time span starting from the departure time of the first train and ending at the time when the first train of the next platoon can depart. The period of a platoon is the minimum capacity needed to fulfill the departure of the trains included in this platoon. During a certain hour, the departure mode consists of many platoons, the sum of whose periods is the total capacity needed. The period of a platoon changes in different sections which have different train running time. In this paper, we take the longest period to achieve the maximum capacity. The combination of different platoons, different overtaking scheme, minimum headway and running time difference between express and local trains also influence the period. When an inter-city railway line is chosen, train running time and minimum headway are known. Therefore, it is taken that the period of a platoon is only affected by overtaking scheme and the time difference between express train and local train in different sections. In this part, with considerations from level of service, an overtaking scheme needing the minimum capacity is presented, which arranges express train overtake one local train at every possible station.

Considerations from Level of Service. Level of service is affected by many factors, and passengers' waiting time is one of the most important. For mass transit operation with regular headways, it is held that passengers' waiting time is half of the operating headways. Inter-city railways have more than one kind of trains and there may be several headways for each train kind respectively, therefore the evenness of the distribution of headways, i.e. the regularity of headways, is more concerned by passengers than the time of headways. Normally, there are less express trains than local trains, so the even departure of express trains is vital to high level of service.

For example, five express trains and fifteen local trains are scheduled to depart within one hour and the minimum headway is 3 minutes. An extreme schedule is to depart the five express trains first, one by one with the minimum headway, and then the fifteen local trains one by one with the minimum headway. The capacity used in this way is the lowest, but it is obvious that the level of service is also very low and this schedule can not be put into operation. A suitable solution is to divide those trains into five platoons, each of which consists of one express and three locals, and the platoon repeats every 20 minutes. It is no doubt that this schedule is of high level of service since the express trains are evenly distribution within this hour. However, this schedule may be infeasible due to lack of enough line capacity. That is, if this

one-express-three-local platoon can not run through all stations along the line within 20 minutes, this schedule can not be realized. For an inter-city railway line with double track and several overtaking stations, overtaking is the action by which the order of trains are changed thus affecting capacity usage. Therefore, overtaking arrangement is the key to achieving a feasible train departure mode with both high level of service and less usage of capacity.

Overtaking Scheme. The inter-city railway line we are interested in is the one which has both express and local trains. Since overtaking is inevitable, sidings are constructed at several stations paralyzing the trunk tracks. These stations we call overtaking stations, and the rest stations are just normal stations with only two trunk tracks. Considering construction cost, for these overtaking stations, there are only two overtaking sidings, one for each direction, so the number of trains to be overtaken at a station is only one. In this situation, a compromise solutions between service level and capacity usage is to operate trains with the following overtaking scheme, as is shown in Figure 1.

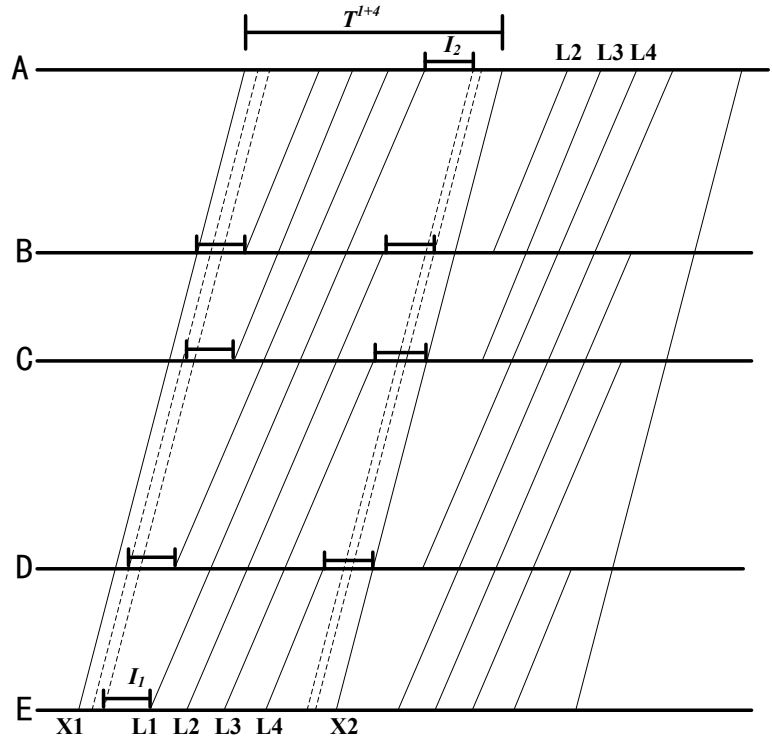


Figure 1. Overtaking Scheme of Platoon 1+4

Due to the speed difference, we arrange the express train overtake local train at every overtaking station. This fixed overtaking scheme can make best use of line capacity under the requirement of high level of service. From the view of train diagram, it can be noted that local trains are separated by express trains, and the number of local

trains between two express trains is the same for every section between two neighboring overtaking stations. Figure 1 is part of the train diagram when applying the fixed overtaking scheme of platoon 1+4, i.e. one express train and four local trains departing consecutively with the headway as narrow as possible. Stations A to E are overtaking stations. Since no overtaking happens at the normal stations, so they are neglected. Solid lines represent trains. X1 and X2 are express trains and L1 to L4 are local trains. It is shown that in each section between two overtaking stations, there are always 4 local trains between express trains. From station E to station A, there are altogether 7 local trains involved between two express trains besides L1 to L4. In this part of the diagram, all 7 local trains are overtaken by express trains except L1, which runs from E to A without being overtaken. The running time of this kind of train is important to calculate the period, so we call the local train without being overtaken *long train*, and the rest *short trains*. Here we consider three kinds of minimum headways, which are the shortest time between any two following trains. I_1 is the minimum time span needed for the departure of a local train following an express train; I_2 is the minimum time span needed for the arrival of an express train following a local train; and I_3 is the minimum time span needed between the departure of two local trains or two express trains. The time span between X1 and X2 is the period of platoon 1+4. A feasible departure mode means the time spans between any two following trains have to be equal or longer than the above mentioned minimum headways in respective situations. To make the calculation method clear, 4 dashed lines are drawn near express trains which will be explained in section 4 of this paper.

Calculation of Period of Train Platoon

For an inter-city railway line with Q+1 stations and Q sections, we first consider the situation of 1+M and when $M \le Q$. It can be seen from Figure 1 that the period of jth combination of M neighboring sections starting from the jth section ($j \le Q - M + 1$) can be calculated as

$$T_{Period}^{1+M,j} = I_1 + I_2 + \sum_{i=j}^{j+M-1} \Delta t_i + (M-1)t_{Stop} + \Omega^{Left} + \Omega^{Right} \qquad (1)$$

where Δt_i is the time difference between local and express trains in ith section, t_{stop} is the dwell time of local train at a station (it is taken as equal at all stations and is not shown in Figure 1), Ω^{left} and Ω^{right} (which will be described later) are additional time offset caused by minimum headway and time difference between express train and local train in certain sections. In formula (1), if I_1, I_2, Ω^{left} and Ω^{right} are aside, then the rest is just the travel time of the long train (running time plus dwell time). For a specific railway line, headways such as I_1, I_2 are fixed as

well. These form the primary part of a period, which is to be completed by additional time offsets.

The period of platoon 1+M is the maximum among all section combinations, so we have

$$T_{Period}^{1+M} = \max\{T_{Period}^{1+M,j}, j=1,2,..,Q-M+1\} \quad (2)$$

When $M>Q$, there are will be $Q-M$ local trains without being overtaken, so it is obvious that

$$T_{Period}^{1+M} = T_{Period}^{1+Q} + (Q-M)I_3 \quad (3)$$

In the same way, when there are more than one express trains running consecutively, i.e. when $N>1$, the period is

$$T_{Period}^{N+M} = T_{Period}^{1+M} + (N-1)I_3 \quad (4)$$

Through formula (1) to (4) the period of a train platoon can be easily calculated out. It is almost the linear function of running time difference between express and local trains Δt_i, dwell time t_{stop} and headways I_x, except there are additional time offsets Ω^{left} and Ω^{right}.

When $\forall\ \Delta t_i \geq I_1$ and $\forall\ \Delta t_i \geq I_2$, the existence of short trains are always feasible, that is, headways are satisfied among all trains both long and shorts. However, if there are sections where Δt_i is less than I_1 or I_2, then there may be the case that the time span between two consecutive trains does not satisfy the minimum headway. Therefore, offset indicators are presented to judge whether or not additional time span is needed and how much to offset the express train in order to satisfy the headway constraint.

Left side offset indicator Ω_i^{left} is defined as the difference between the departure time difference of express and local trains at the ith station, and the minimum headway I_2. This indicator shows whether it is feasible to insert a train at the ith station between the express train and the local train, which departs just behind the express train at the last station. It is obvious that $\Omega_i^{left}=0$ means the express train is followed by the local train with the minimum headway I_2. The departure is feasible and there is no need to offset. $\Omega_i^{left}>0$ means the time span between the express train and the follow train is longer than the minimum headway I_2. There is no need to offset either and the redundancy is Ω_i^{left}. However, $\Omega_i^{left}<0$ means it is infeasible to insert a train at the ith station and some offset has to be taken. To ensure the overtaking scheme feasible

along the line, the express train is to be offset to the left, until for every station Ω_i^{left}>=0.

Left side indicator Ω_i^{left} is related with the difference between express and local train, minimum headway and dwell time. It can be calculated by the following recurrence expressions.
$\Omega_i^{left}=0$

$$\Omega_{i+1}^{left}=\Omega_i^{left}+\sum_{j=1}^{i}[\ \Delta t_j-(I_3-t_{Stop})],i=1,2,..,Q \tag{5}$$

From (5) we can achieve left side indicators of all overtaking stations. And the additional time offsets Ω^{left} is the abstract of the maximum one. After offset of Ω^{left} left-wards, the overtaking scheme is feasible at all stations.

Similarly, right side indicator Ω_i^{right} can also be defined as whether it is feasible to insert a train between the express train and local train just before it at the *i*th station. $\Omega_i^{right}=0$ means the slice between express and local trains equals the minimum headway I_1 and no need to offset. Ω_i^{right}>0 means there is time redundancy after inserting a train and Ω_i^{right}<0 means it is infeasible to insert a train and offset is necessary. To ensure the overtaking scheme feasible along the line, the express train is to be offset to the right, until for every station Ω_i^{right}>=0. Same as the relation between Ω^{left} and Ω_i^{left}, Ω^{right} is also the abstract of the maximum Ω_i^{right}. The recurrence formula is the same except in the inverse order.
$\Omega_i^{right}=0$

$$\Omega_{i-1}^{right}=\Omega_i^{right}+\sum_{j=1}^{i}[\ \Delta t_j-(I_3-t_{Stop})] \tag{6}$$

Through formula (5) and (6), the period of a certain train platoon can finally be calculated out. A train departure mode is arranged by several same or different train platoons, and the minimum capacity needed for a train departure mode can thus be decided by summarizing the periods of train platoons it consists of.

Case Study

We take a case study of an inter-city railway line under the stage of design. The double track line, which is designed to accommodate both 200km/h express train and 140km/h local train, is about 100 kilometers long with 17 stations, including 8 overtaking station, each of which has only one overtaking track for each direction respectively, i.e. only one train can be overtaken each time at an overtaking station. Travel time (running time plus dwell time) of both express and local trains in sections divided by overtaking stations and their difference are shown in Table 1.

Since the two directions are operated separately, in this case we take only the upward direction. It is demanded by passenger volume forecast that in the long term five express and fifteen local trains depart from ending station in one direction in the peak hour. The designed minimum headways are all two minutes.

Table 1. Travel time of local and express trains and the difference (in seconds).

Sec. No.	Local	Express	Difference
1	765	540	210
2	210	135	75
3	375	240	135
4	285	165	120
5	465	270	195
6	450	390	60
7	420	225	195
8	855	645	210
9	255	195	60

Following the calculation method above presented, it will be decided whether the line capacity is enough to operate in a high level of service. Here we consider two scenarios of departure mode in the peak hour. Scenario 1 is the most evenly distributed train departure mode with platoon 1+3 repeats five times in one hour. In scenario 2, twenty trains are divided into three platoons, 2+5, 2+5 and 1+5. Based on data in Table 1, it is calculated that the period of platoon 1+3 and 1+5 are 13’15’’ and 18’00’’ respectively. So, the capacity needed in scenario 1 is 13’15’’ multiple by 5 equals 66’15’’ and the capacity needed in scenario 2 is 20’00’’ multiple by 2 plus 18’00’’ equals 58’00’’. Compared with the capacity provided, 60’00’’, it is obvious that scenario 1 is infeasible while scenario 2 can be realized.

According to formula (2), the period is the maximum among all time span of the combination of three sections, which is also the bottle neck. During calculation of period 1+3 in this case, the bottle neck is section combination 6, 7, 8 or 7, 8, 9. If the railway operator is required to provide the high level service, that is, scenario 1 is to be realized, then the capacity of this bottle neck needs to be improved. Since the time difference accounts the largest part of period, the period can be decreased by a shorter time difference, and this can be achieved by adding an extra overtaking station. In this case, there is a normal station in section 8, so it is suggested this station be upgraded into an overtaking station. With an extra overtaking station, the previous section 8 is broken into two sections and time difference is therefore reduced. After upgrading, the travel time is calculated and listed in Table 2. Using the same method, it is calculated out that the period of 1+3 of the upgraded version is

12’00’’, thus the departure mode of five times of platoon 1+3 can thus be feasible. It has to be noted that this suggestion is only out of operation consideration to make a comparison with the original one. In real project, many more factors need to be taken into account and considerations of construction always play the vital role.

Table 2. Travel time of local and express trains and the difference after upgrading (in seconds).

Sec. No.	Local	Express	Difference
1	765	540	210
2	210	135	75
3	375	240	135
4	285	165	120
5	465	270	195
6	450	390	60
7	420	225	195
8	300	195	105
9	525	450	75
10	255	195	60

Conclusions

Focusing on the inter-city railway operation, we discuss the level of service and usage of line capacity. High level of service usually requires more usage of line capacity. A certain fixed overtaking scheme is presented for inter-city railway with both express and local trains aiming to achieve high level of service with the minimum usage of line capacity. Based on analysis of train diagram, it is designed that trains depart in certain platoons consist of express trains followed by local trains and a calculation method of platoon period is constructed. It is noted that the period is an index of usage of line capacity. The period is mostly connected with the time difference between express and local trains, headways and dwell time of local train, and is affected by the relationship between headway and time difference. In the end the method is tested by a case study of an inter-city railway under the stage of design and it is shown that the method is efficient for the line capacity calculation of inter-city railway.

References

Zheng ,S. D., and Wu, H. L. (1988). *Railway Train Operation*, Beijing: China Railway Publishing House.

Zhu, J. H, , Zhang, X., and Zhang, W. (2009) "The Effect of stoppings to hourly capacity of inter-city railway." *China Railway Science*, 30(4), 108-112.

Ni, S. Q., Zuo D. J., and Wang, C. G. (2005). "The effect of high-speed railway station distribution to line capacity." *China Railway Science*, 26(3), 7-10.

Cao, E. L. (1995). "Distribution principles of high-speed railway stations." *Railway Civil Engineering Journal*, 63(2), 86-89.

Piao, A. H., and Hu A. Z. (1996). Research on Train Organizations of China's High-speed Railway. Journal of Railway, 18(3), 105-110.

Su, S. H., Tian, C. H., and Chen, Z. Y. (2008) "Analysis and calculation of capacity of passenger dedicated line." *China Railway Science*, 29(5), 119-124.

Abril, M., Barber, F., Ingolotti, L., et al. (2008). "An assessment of railway capacity." *Transportation Research Part* E, 44(5), 774-806.

Wagon Flow Organization Modeling Based on Distributed Autonomous Decision-Making

Xuesheng Ji[1], Xi Jiang[2] and Wei Wang[3]

[1] Ph.D. Candidate, School of Traffic and Transportation, Beijing Jiaotong University, The People's Republic of China Ministry of Railways, Beijing 100844, China
[2] Assistant Professor, [3] Ph.D. Candidate, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, Email: jiangxi@jtys.bjtu.edu.cn
School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

Abstract

As the Chinese transport system began to develop rapidly, researchers became highly concerned about the reform of wagon flow organization. The traditional modeling and optimization method of wagon flow organization based on concentrated decision-making doesn't sufficiently support the exploration to the decision process and method of wagon flow organization with the innovation for railway transport mode currently. In this paper， we study the decision-making method of wagon flow organization，which is based on distributed autonomous decision-making in a manner of competing resources according to different wagon flows property in the process of organization. Each distribution agent competes and cooperates with the other according to competing ability in order to make more profit, the system arrived the state of global optimization in the process of conflict and coordination. Furthermore, a multi-agent model on wagon flow organization is presented. The modeling method is tested on experimental computation.
Key Word: railway transport; multi-agent modeling; wagon flow organization

Introduction

The wagon flow organization schedule is the core link between the transportation demand and the transportation resource, and it is also the most basic transportation organization schedule that directly shows the basic characteristic of the transportation organization pattern. As a classic topic in the transportation organization field, many scholars have made widespread researches on the related problems of the wagon

flow organization and gained a lot of achievements (Liang and Lin 2006; Cao et al. 2007; Feng 2003; Huimin 2001; Feng et al. 2003; Böcker et al. 2001). In recent years, along with the great development, a new reform is occurring the in Chinese railway transportation organization pattern, which comes up with a series of new tasks about the researches of the related theories, methods, and technologies in the wagon flow organization. However, the reform of wagon flow organization pattern is firstly reflected in the changing of decision concepts and methods in wagon flow organization. Previous wagon flow organization researches focus more on making the wagon flow organization schedule under a certain transportation organization pattern, so different organization pattern information and its effect on wagon flow organization decision often could not be fully reflected. However, the research on wagon flow organization modeling, which centers on the plan and decision-making process itself for different wagon flow organization concepts and methods, helps to analyze and test the new transportation organization patterns. Therefore, it is important to establish a new model from a new angle for the railway transportation organization researches at the present stage.

Traditional Wagon Flow Organization Modeling

In railway network, there are lots of stations and considerable wagons, which could form a huge number of combination wagons. Consequently, the establishment of the wagon flow organization schedule is a large-scale optimization problem, where researchers focus on the modeling and optimization method on the basis of mathematical programming for a long time. By summing up the previous research results and research methods, it is found that the traditional method of solving the wagon flow organization optimization problem mainly has the following characteristics:

(1) It must focus on maximum use of the efficiency and the capacity of railway, and aims to achieve the least wagon allocation cost (or the related varieties) in model.

(2) Limited by the transportation organization pattern, the optimization model is generally based on a number-fixed consolidating strategy, on the other hand, relies on a specialized coupling in order to meet other special demands.

(3) Due to limitations of the transportation capacity, different requirements of the transportation quality often could not be fully reflected in the previous optimization modeling method.

When the problem of wagon flow organizations is studied under the background of organization pattern reform, it could be seen that previous modeling methods and related models cannot play a very good role. When we use traditional mathematical programming models to describe the optimization problems of wagon flow organization, all the information such as the decision-making procedure and essential factors is fixed in the mathematical models. Once the transportation organization pattern changes, it is difficult for the mathematical programming models to adapt to

the changes of models' structures and components. As a result, it is necessary to come up with a new decision-making method of the wagon flow organization according to the fundamental disciplines of the formation of wagons and the transportation process.

Wagon Flow Organization Modeling Method Based on DAI

Distributed Autonomous Decision-Making Pattern Based on Resource Competition and Cooperation. The key problem of wagon flow organization optimization is to properly distribute and make use of the transportation resources under the demand-oriented. In terms of wagons, the essence of resource distribution is resource competition and mutual cooperation. Considering DAI (distributed artificial intelligence) has an outstanding advantage in solving the overall coordination and consistency problems (Wu and Wang 2004), this paper takes the distributed autonomous decision-making as the guide and studies how to make up models in wagon flow organization optimization.

The basic idea of this research can be expressed as follows: The distributed autonomous decision-making agent is composed of the different kinds of wagons endued with autonomy, various transportation resources, and resource managers. Each distributed intelligence decision-making agent competes and cooperates with each other to pursue maximum profits, and to achieve the integral optimization during the conflicts and coordination. In this way, traditional centralized decision-making mode is converted to distributed autonomous decision-making mode. With different conditions and different qualified profit agents, the same decision-making process can realize different optimization objectives. At this moment, system optimization becomes a dynamic behavior consisting of many agents' distributed decision-making. The wagon flow organization program made in this way is a multi-coordinative optimization result, so it is highly adaptive.

Model Framework for Wagon Flow Organization Decision-Making Based Multi-Agent. The entities referred in the railway wagon flow organization mainly include wagons, trains, transportation resource, transportation resource managers, and transportation environment. The coupling and decomposition among wagons is the key behavior to change the way of wagon operation, so it should be emphasized in the model. Just because of the coupling and decomposition behaviors, which occur in any sections in the entire journey, wagons can be divided into several wagons sections.

Wagon flow organization resolves the problem of competition and cooperation in the distribution of transportation resource, whose solution will carry on around this process. Due to transportation resources limitation and unique railway transports manner by assembling wagons to trains, wagons coupling and decomposition could be regarded as a "choice behavior" to pursue the maximum benefits in the transportation process. That is, wagons are coupled, and then form "wagon alliances." Applying for resources by means of alliance can reduce the resource

occupied cost.

According to the wagon flow organization modeling method based on DAI, the decision objectives can be distributed and realized by decision-making agents through cooperation among other agents. Therefore, the wagon flow organization mission can be divided into several sub-missions, accomplished by mission solving entities which include transportation managers and several wagon flow decision-making individuals. According to the division of missions, each mission could be executed by different roles such as wagon flow decision-maker, train flow decision-maker, resource managers, and many so on. In the multi-agents models, these roles could be acted by related decision-making agents. Within it, train flow can be regarded as the results of the wagons cooperative decision. That is to say, train flows are made up of several wagon sections in some certain spatial paths. In order to emphasize the wagon flow cooperation in the way of alliance to apply for the transportation resource, this paper uses "wagon alliance" agents to represent the role of "train flows." Whereas, the distribution decision-making multi-agents model of wagon flow organization mainly include different wagon flow agents, wagon flow alliance agents, transportation resource agents, resource management agents, and so on. Each decision-making agent endowed with initiative act ability and individual decision-making objectives realizes the optimization of wagon flow organization in conflict and coordination. Thus, we could construct the decision-making model framework of the railway wagon flow organization based on MAS, just as Figure 1 shows. In the MAS system, the wagon flow agent model is the most critical part, and the modeling method will be shown in the following part of this paper.

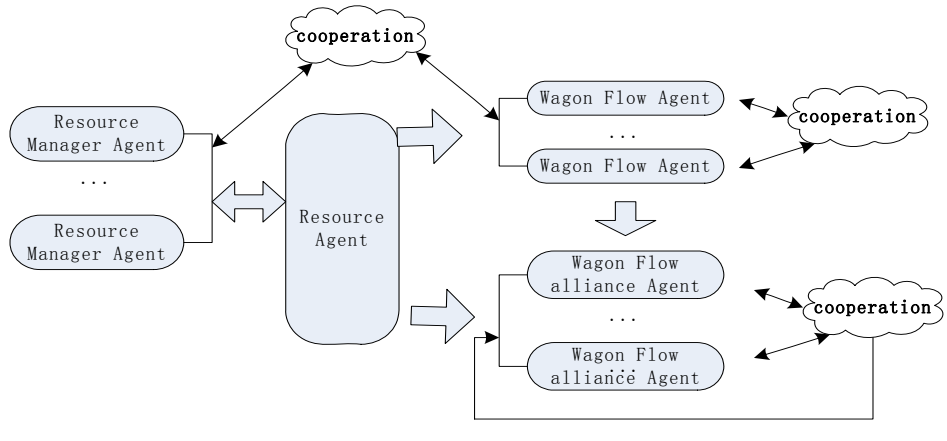


Figure 1. Wagon flow organization modeling framework based on MAS.

Decision-Making Model of Wagon Flow Agent

Decision-making Objectives of Wagon Flow Agent. In distributed autonomous decision-making, the system optimization goal is expressed by an autonomous wagon flow agent with different requirements. In the model, competitive ability is endowed with wagon flow agent and the use cost is endowed with transportation

resource. The decision-making objective of each wagon flow agent is the minimum of occupation consumption of transportation resource while wagon flow agent has enough competitive capital to afford the cost of using transportation resource on the running route.

In the wagon flow organization scenario, the resource consumption of each wagon flow includes the consumption of wagon assembling time in the original station, the wagon coupling and decomposition time in the marshalling station, the running time in line, and the occupation cost of running line. In order to be convenient when making a decision, the consumption is expressed as the generalized cost of train component. Within certain transport capacity, the wagon flow, which is able to afford the train running cost, could be organized and then form a train. Therefore, the competitive capital is introduced as the pre-condition of obtaining transportation resource.

When the wagon flow agent cooperates with each other, the train in one direction contains more than one wagon flow agent. So the consumption related to the train composition and operation could be shared by all wagon flow that is in the same direction. In order to clear such a shared relationship in the interest composition of wagon flow agent, the cost elements can be defined as follows:

The generalized cost Cdirect of train in station. The generalized cost Cdirect of train is defined as the total assembling time of wagon flows, which are coupled to a same "wagon alliance" in station.

The generalized cost Ctrains of occupancy train running line. The generalized cost Ctrains of occupancy train running line is defined as the total occupancy cost of running line for the wagon alliance movement in rail, and can be expressed as the multiplier between the number of open trains N and a generalized cost Cst of single train.

On this basis, the generalized function f_f on the interest of wagon flow is used to describe the decision-making objective constituting the wagon flow agent. For example, the traffic volume of wagon flow agent *j* is *N*, which terminal stations are A and B in network. In a certain organization scenario, some running routes of the wagon flow contain L_k technology stations apart from the terminal station. Wagon flow *j* is coupled in S (L_k+1> *S*> = 0) stations among the L_k technology stations. The set of technology stations in which wagon flow j is coupled is ω_s. Wagon flow j will be consolidated and form the running train in S+1stations—including the origin station A. The set of consolidated station is ϕ_{s+1}. So it will form into S+1 succeeding train flow. The set of running section number is u_{s+1}. In this way, wagon flow *j* must take the corresponding generalized cost of train component for S+1stations, the generalized cost of wagon flow coupling for S technology stations, and the generalized cost of train traveling for S+1 sections, in addition to the generalized cost of train running for S+1 sections. The objective function to minimize the cost of

wagon flow j is expressed as:

$$Min\ G=f_f((\alpha * N * \sum_{i\in\omega_s} C^i{}_{couple-assemble}),(\beta * \sum_{i\in\phi_{s+1}} \alpha_{ij} C^i{}_{direction}),(\lambda * N * \sum_{i\in u_{s+1}} T^{ij}{}_{running}),$$

$$(\delta * \sum_{i\in u_{s+1}} d_{ij} C^i{}_{\text{trrains}}),R_j) \quad (1)$$

Let $C^i{}_{couple-assemble}$ represent the cost of wagon coupling and sharing time in coupled station i;

Let $C^i{}_{direction}$ represent $\boldsymbol{C}_{\mathbf{direct}}$ in coupled station *i*;

Let α_{ij} represent the proportion of wagon flow *j* assembling time cost in $C_{i\,direction}$, the proportion parameter relates many aspects, such as the coupling mode, the composition, and proportion of traffic volume;

Let $T^{ij}{}_{running}$ represent the running time of wagon flow *j* which starts from station *i*;

Let $C^i{}_{\text{trrains}}$ represent the $\boldsymbol{C}_{\mathbf{trains}}$ cost in station *i*;

Let d_{ij} represent the proportion of occupation cost of wagon flow *j* for running line in the cost of C_{itrrains}, the proportion parameters relate the coupling mode;

Let R_j represent the competitive capital of wagon flow *j*. It relates traffic volume, stability, running distance, the type of wagon flow, and social value of the wagon.

Decision-making Model of Wagon Flow Agent Based on Cooperation. In the program of the traffic organization, the interests of the wagon flow agent are mainly reflected in the four kinds of costs. The composition ratio and numerical value of the four kinds of costs are different because of the use patterns of different resource and traffic organization, and there is relevance among them. The generalized costs of traveling by train depend primarily on the choice of running path, and the relationship with other three is more independent. The optimization process arising from the decision-making factors is relatively straightforward in the wagon flow decision-making, while the other three costs formed a close relevance due to the collaboration of the decision-making agent. In the interest composition of the wagon flow agent, the coupling costs and train line costs on wagon flow were time-related costs, while the generalized cost of a train traveling have nothing to do with the time, and its affordability is limited by the competitive capital wagon flow agent. The decision-making of wagon flow agent can be launched from the following three basic directions:

(1) It is determined whether wagon flow agent confers with the resource managers on train compiled conditions according to the competitive capital. The time cost expenses will be lower when train-line cost expenses increase, thus reducing the number of train wagons or adopting the fixed-time strategy with certain frequency.
When the competitive capital of wagon flow could influence hardly the pre-establishing condition of train coupling, the decision-making of wagon flow is how to balance the coupling cost of wagon flow and the generalized cost of train line by the cooperation between wagon flows, so it could reduce the total cost. The

decision-making direction goes to (2) or (3).

(2) The total time charges will be reduced by adopting direct trains and saving the coupling cost.

(3) The total time consumption will be reduces by cooperating with other wagon flows in its route in order to reduce the proportional cost of each train component.

Each wagon flow agent makes its own decisions according to its individual targets. First of all, the wagon flow agent determines the basic decision-making direction of the wagon flow according to its characteristics, experiences, and perceptions of the environment. Then, the wagon flow agent will carry out the interesting assessment of variety of the decision-making program according to an objective of maximum interest in multiple paths within the available path set. At the same time, the decision-making direction in the next step is identified, and the optimization process continues. The decision-making process of the wagon flow agent is shown in Figure 2. As the modeling unit having the characteristic of autonomous and intelligence, the wagon flow will make an action spontaneous in order to achieve the main task in the life-cycle. Their behavior can be divided into three stages, such as the process of information gathering and dissemination, the decision-making of the wagon flow, and decision-making assessment and modification.

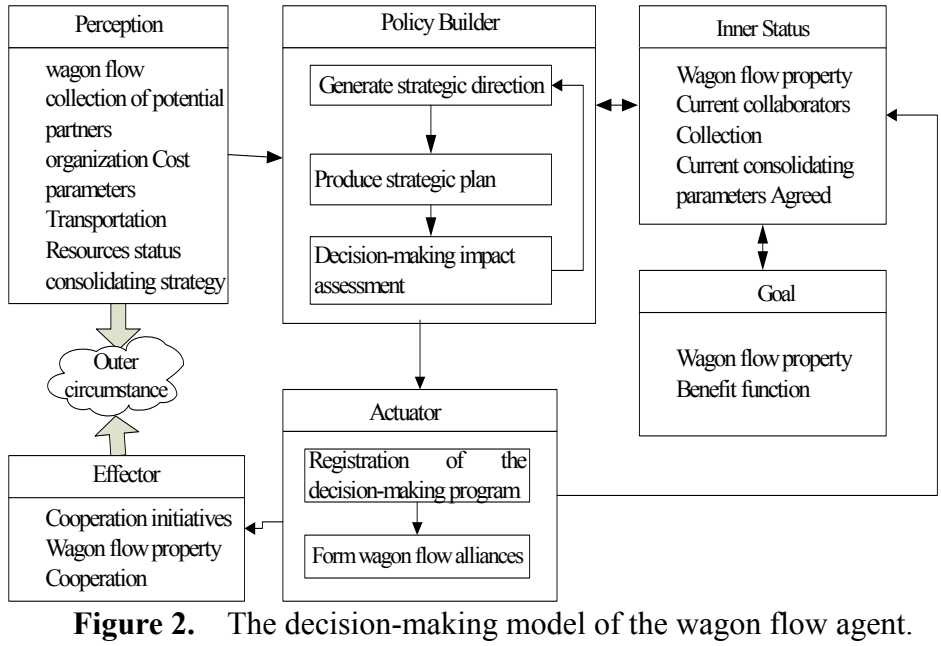


Figure 2. The decision-making model of the wagon flow agent.

Experimental Computation

In this section we describe the test cases in which the multi-agent modeling method is applied to the transport organization scenario under a different consolidating strategy. We consider two transport modes decided by consolidating strategy:

time-fixed and number-fixed. The layout of technical stations A-F and the related information are shown in Figure 3.

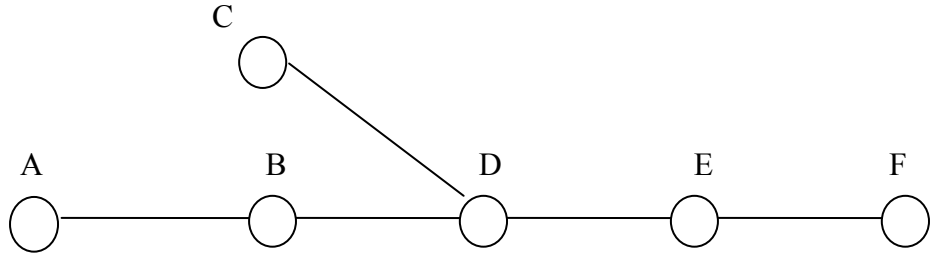


Figure 3. The layout of technical service station.

The competitive capital of wagon flow is divided into three levels according to the stability of various wagon flow, the arriving law, and transportation time limitations. The coefficient R' represents the capital impact to wagon flow interests function. According to the related operating cost and the calculation of time value (Shi Feng，2003), the consistency parameter in interests function is set $\alpha=\beta=0.64$, $\gamma=1$. Wagon flow and related information are as follow:

Table 1. Wagon flow and related information.

Originate	A				B			C			D		E
Destination	B	D	E	F	D	E	F	D	E	F	E	F	F
Wagon flow	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f13	f13
Volume	60	50	45	85	65	50	40	50	35	40	50	40	55
R'	1.5	1	12	1	1.2	1	1.5	1.5	1.2	1.5	1.2	1.5	1.2
Mg						100					200		150
Tg						4					4		3
Tr	A-B				B-D			C-D			D-E		E-F
	3				4			3			3		4

Note: M_s represents the coupling capacity of station; t_s represents train time consumption for making up; t_r represents standard running time of section.

Under the number-fixed strategy, it is assumed that the fixed number of coupling train is 40. Owing to the fixed coupling number, the generalized cost on a train's running line is also fixed, which has no impact on decision-making. So the wagon flow cost includes the generalized cost of train component and generalized cost of train coupling. On the basis of Eq. 1, the interests function G_j on wagon flow agent j is expressed as follow:

$$G_j = (\sum_{i \in S_j} a_{ij} * C_i * M - N_j * \sum_{i \in S_j} \omega_{ij} C^i_{couple-assemble}) * R'_j \qquad j = 1,2,\ldots 13 \tag{2}$$

$$\omega_{ij} = \begin{cases} 1 & \text{wagon flow j is coupled in station i} \\ 0 & otherwise \end{cases}$$

Let R'_j represent the equivalent weight coefficient on the competitive capital of wagon flow j;
Let C_i represent making up parameter in station I. Because the difference of the arriving law of wagon flow and the various operating characteristics on technology station are not to be considered, the making up parameter is set to 12;
Let N_j represent the traffic volume of wagon flow agent j. Its value is shown in Table 1; and
Let S_j represent related technology station set on wagon flow j, including terminal and making up station.

Under the time-fixed strategy, it is assumed that the minimum number of coupling train is 25, and the maximum is 40. On the basis of Eq. 1, the interests function G_j on wagon flow agent j is expressed as follows:

$$G_j = \sum_{v \in Q_j} a_{jv} * (\alpha * 12 * R'_j * N_v / K_v + \gamma * K_v * f_{running} * t^v{}_{running}) + \sum_{i \in S'_j} C^i_{couple-Assemble} * \beta * N_j$$

$$\text{j=1,2,…13} \qquad (3)$$

Let $f_{running}$ represent the generalized cost of train running in unit of time;
Let S'_j represent the making up station set on wagon flow j;
Let Q_j represent wagon flow coalition set in the coalition chain of wagon flow j;
Let N_v represent the total traffic volume on coalition of wagon flow $v \in Q_j$;
Let K_v represent the train frequency after the negotiation of wagon flow coalition $v \in Q_j$ and resource agent;
Let $t^v_{running}$ represent the traveling time of coalition of wagon flow coalition $v \in Q_j$ in section;
Let a_{jv} represent the proportion of traffic volume of wagon flow j accounting for the total traffic volume of coalition of wagon flow $v \in Q_j$.

The decision-making of wagon flow organization is deduced by different strategies according to the proposed decision-making modeling method of distribution and self-government. The result is as follows:

(1) The number of training is 22.625, with a total cost of 5751.7 under the number-fixed consolidating strategy, while the total cost is 5591.7 under time-fixed strategy. Therefore, the cost is lower under the conditions of time-fixed strategy.
(2) Under time-fixed consolidating strategy, the number of training is 25.982, but that is 22.625 under the number-fixed consolidating strategy. The number of train under

time-fixed strategy is more that under number-fixed strategy.

(3) The result shows that the time-fixed strategy exceeds the number-fixed strategy in three aspects, such as fewer wagon flow coupling; reduced number of make-up times, and higher the rate of direct arrival.

Conclusions

In conclusion, there are many advantages by adopting the decision-making approach of distribution and autonomous to model wagon flow organization. First of all, adopting an abstract approach to wagon flow by decision-making agent, it is easy to implement a dynamics merger and decomposition among different wagon flows, and can give the difference between wagon flows. So the demand-based transport organization could be achieved easily. Secondly, the adopted decision-making of distribution and self-governance could influence the local behavior and coordinate activities of decision-making agents in the optimization process by giving different information to the decision-making agent. It facilitates the law within the wagon flow organization and supports the exploration for optimal mode of transport organization. Thirdly, adopting multi-agent modeling, the decision-making agent has more expressive ability and dynamic characteristics, and the system has a higher reusability and flexibility. Finally, the wagon flow organization optimizing is realized in logic distribution through the distribution decision-making method for the basis of solving large-scale problems by distributed and parallel algorithm.

Acknowledgment

This study was jointly funded by Ministry of Education of the People′ s Republic of China (No. 20060004025), and National Natural Science Foundation of China (No. 60736047).

References

Böcker, J., Lind, J., and Zirkler, B. (2001). “Using a multi-agent approach to optimize the train coupling and sharing system.” *European Journal of Operational Research*, 134, 242-252.

Cao, X. , Lin, B. and, Liu, H.. (2007). “Optimization of car flow organization in nonstop train loading base.” *Journal of the China Railway Society.* 29, 17-21.

Niu, H. (2001). “A model for cooperative optimization of heavy and empty traffic organization in railway hubs using genetic algorithm.” *Journal of the China Railway Society.*, 23, 22-26.

Liang, D.,Lin, B. (2006). ”Study on the theory and model of the optimal train formation plan of multi-block trains at technical service stations.” *Journal of the*

China Railway Society, 28, 1-6.

She, F., Fang, Q., Li, X., Mo, H., and Liu, H.Y. (2003). "A bi-level program ming optimization method for the layout of technical service station." *Journal of the China Railway Society.*, 25, 24-30.

Shi, F., Mo, H. H., Li, Xinhua, and Qin J.. (2003). "Optimization method for organization of local wagon flow with district." *Journal of Traffic and Transportation Engineering*. 3, 84-89.

Wu, H. Y., and Wang, X. (2004). "Research on multi agent system and distributed AI." *Microcomputer Development*, 14, 80-84.

Petri Net Modeling of Spatial and Temporal Constraints in Track Allocation Problem at Railway Stations

Wenzheng Jia [1,2], Tinkin Ho[3], Zhenqi Chen[3] and Baohua Mao[1]

[1] Ministry of Education Key Laboratory of Complex System Theory & Technology of Urban Traffic, Beijing Jiao tong University, Beijing 100044, China, Email: eewzjia@gmail.com
[2] China Urban Sustainable Transport Research Centre, China Academy of Transportation Sciences, Beijing 100029, China
[3] Hong Kong Polytechnic University, Department of Electrical Engineering, Hong Kong, China

Abstract

The track allocation problem (TAP) at a multi-track, multi-platform mainline railway station is defined by the station track layout and service timetable, which implies combinations of spatial and temporal conflicts. Feasible solutions are available from either traditional planning or advanced intelligent searching methods and their evaluations with respect to operational requirements are essential for the operators. To facilitate thorough analysis, a timed Colored Petri Nets (CPN) model is presented here to encapsulate the inter-relationships of the spatial and temporal constraints in the TAP. The relationships among the static station track layout, TAP solution and the available buffer times are described by the model and it allows systematic representation of train movement at stations and hence analysis on TAP solutions upon service disturbances.
Key Words: rail stations; track allocation problem; colored Petri nets

Introduction

The track allocation problem (TAP) is a sub-problem of mainline railway service scheduling. With given track layout and service timetables at a multi-track, multi-platform station, a conflict-free path, composed of an arrival route and a departure route, is allocated to each train for traversing and/or stopping at the station.

The essence of solving TAP is to avoid spatial and temporal conflicts due to track layout and timetable respectively while safety restrictions, service requirements for passengers boarding, alighting and transferring are also the constraints. Service disruptions are however inevitable in real-life operation. Robustness of a TAP solution is its capability to absorb disruptions and it is often introduced as buffer time between train services. If a number of feasible solutions to the TAP are available, it is essential to look for the one with the highest robustness while maximizing the utilization of the station capacity. To enable evaluation of the possible solutions to TAP, the spatial and temporal conflicts have to be modeled effectively and the possible buffer times are embodied in the model.

In this paper, a timed Colored Petri Nets (CPN) approach is presented to model the relationships among service timetable, station track layout and buffer times. This model allows systematic representation of train movement at stations and hence simple analysis on the robustness of TAP solutions upon service disturbances.

Station Track Allocation

A mainline railway station is characterized by a number of entry and exit points, track sections and platforms. A train, when timetabled to pass through or stay at the station, has to be assigned a specific route for arrival and another for departure. TAP is to eliminate conflicts among the routes. Various approaches (Zwaneveld, et. al., 2001; Carey, M., and Carville, S., 2003) have been employed to attain solutions to TAP. In general, their objective is to maximize the number of trains to be routed, or to allocate priorities on chosen routes. However, the service quality of the resulting track allocation scheme has rarely been addressed. When robustness is one of the commonly used indicators of service quality, the feasibility solutions for TAP have to be evaluated properly on robustness.

Spatial and Temporal Conflicts. TAP is categorized as a resource allocation problem. With reference to a job shop scheduling problem, a temporal conflict occurs when two jobs must be processed simultaneously and it can be eliminated by assigning parallel machines to the jobs or introducing a buffer time between them. If two jobs need to use the same machine, it is a spatial conflict, and it is resolved by providing a buffer time to absorb the possible disturbances in real-time operation. Evidently, if it is possible to insert more buffer times between jobs, the schedule is more robust spatially and temporally. Resources are however limited, and the required buffer times are not necessarily available to avoid conflicts. As a feasible solution to TAP is embedded with certain buffer times, its robustness is implicit and the evaluation of the solution requires proper modeling of the spatial and temporal conflicts at the station.

An arrival or departure route of a train at a station is composed of sequential track sections, each section is occupied by no more than one train at any one time, or a spatial conflict occurs. An example of a portion of track layout at a station is

illustrated in Figure 1 and it is composed of 10 sections, g1, and g2… g10 despite the simple track layout. From Figure 1, train t1 arrives at track 3 through route r1=<g1, g2, g6> train t2 on track 1 and train t3 on track 4 head for the exit through routes r2=<g7, g2, g1, g3>and r3=<g9, g4, g3>, respectively. Track sections g1 and g2 are the conflict areas for t1 and t2, and g3 for t2 and t3. On the other hand, trains t1 and t3 are free from spatial conflict.

A track section and a train can be regarded as a machine and a job in a job-shop problem. If the processing times of two jobs overlap, a temporal conflict occurs. The timings of the trains using the tracks 1, 2, 3 and 4 are listed in Figure 2. The time diagram of a train is divided into 3 parts. The left-hand triangle denotes the route setup time for the train to stop at or pass through the platform track while the right-hand triangle indicates the release time of the departure route when the train leaves. s(t1, r1) is the setup time for t1 to stop at track 3 through route r1 and s(t2, r2) is the release time for t2 to leave track 1 through route r2. They can be obtained through simple train movement simulator (Ho, T.K.et.al.,2002). The rectangle in the middle is the station dwell time. If a(t1) and d(t1) are the scheduled arrival and departure times for t1, the dwell time is d(t1) - a(t1). Trains t1 and t3 are in temporal conflict, so are t3 and t4.

Buffer Times. To avoid conflicts, buffer time is introduced between two trains which are scheduled to occupy the same track section and its amount reflects the capability to tolerate delays. In TAP, the station layout and service timetables are fixed, different solutions to TAP allow different amounts of buffer times to be inserted and hence their relationships with the track layout and timetables have to be included in the conflict model.

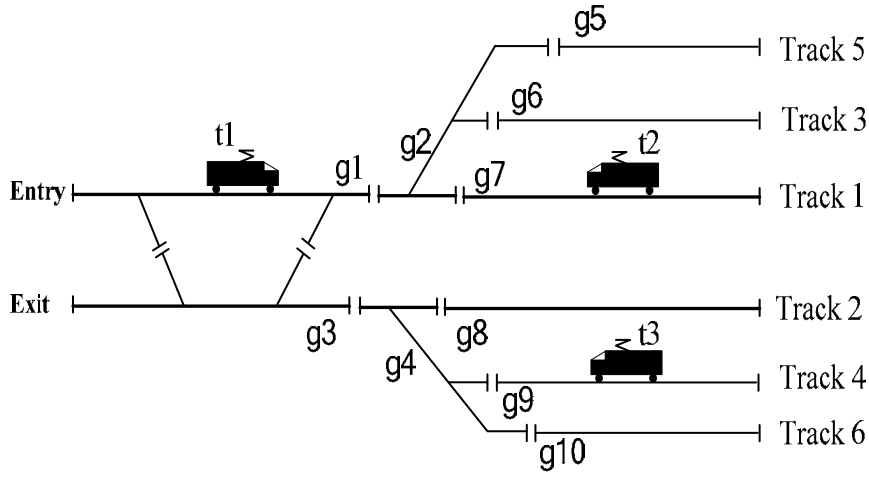


Figure 1. Spatial conflict between trains.

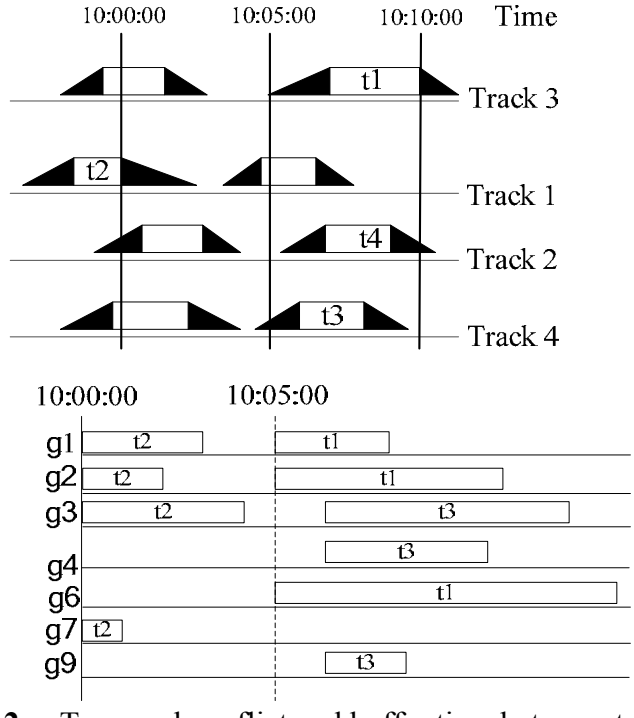


Figure 2. Temporal conflict and buffer time between trains.

Timed Colored Petri Nets Modeling

Petri Nets is a powerful and flexible modeling technique for discrete event systems (Dotoli, M.et.al., 2008). It is particularly suitable for modeling concurrent and asynchronous behaviors and it has found successful applications on job shop scheduling problem modeling (Zhang, H.H. and Gu, M.,2009). Petri Nets has also been adopted to construct a large-scale railway infrastructure (Hagalisletto, A.M.,2007) in which the spatial characteristics of station layouts are modeled. For the modeling of TAP, conflicts, concurrency of train movement and buffer times exist and hence timed CPN approach is employed. The construction and validation of timed CPN model are now well supported by commercially available CPN software tools (Jensen, K., Kristensen, L. and Wells.,2007). The CPN symbols adopted below follows the syntax of CPN ML language (Jensen, K., Kristensen, L. and Wells.,2007).

With the basic CPN model shown in Figure 3, a train and a free track section are denoted by the timed Color Sets P and Q respectively. There are four places and two transitions in the model to represent the process of a train entering, staying and leaving a track section. The tokens in the places are ready to be consumed no sooner than the timestamp indicated (if shown) at the places. Arcs t1 and g1 denote a train and a track section; the output arc of transition 'Occupy' refers to the events that a train occupies the track section, and the value '90' in the arc expression '(t1,g1)@+90' represents the time-span (in seconds) of the event of occupation. The place 'g1 busy' shows the state of the track section being occupied by the train t1. The transition 'Release' denotes train departure from the track section which becomes 'free' afterwards and thus ready to be occupied by the subsequent train.

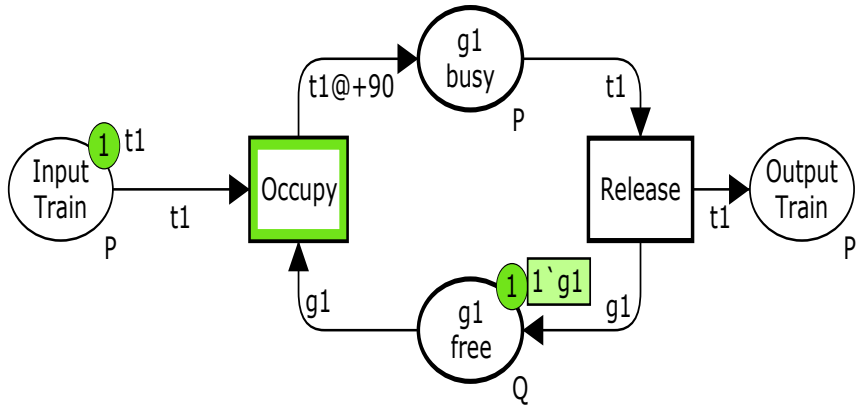
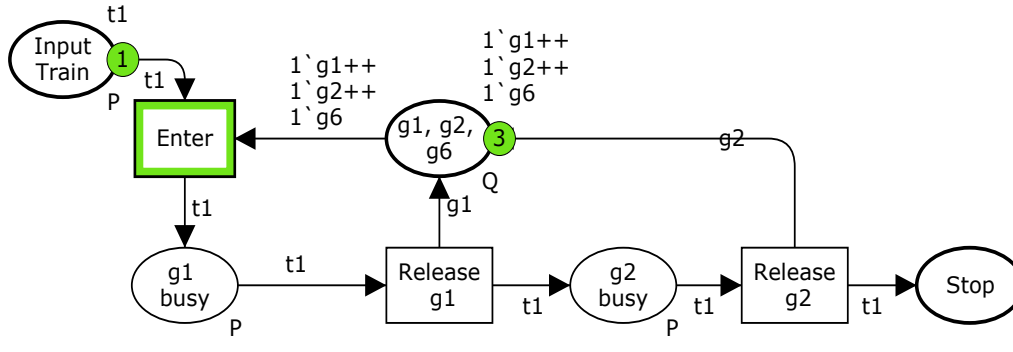
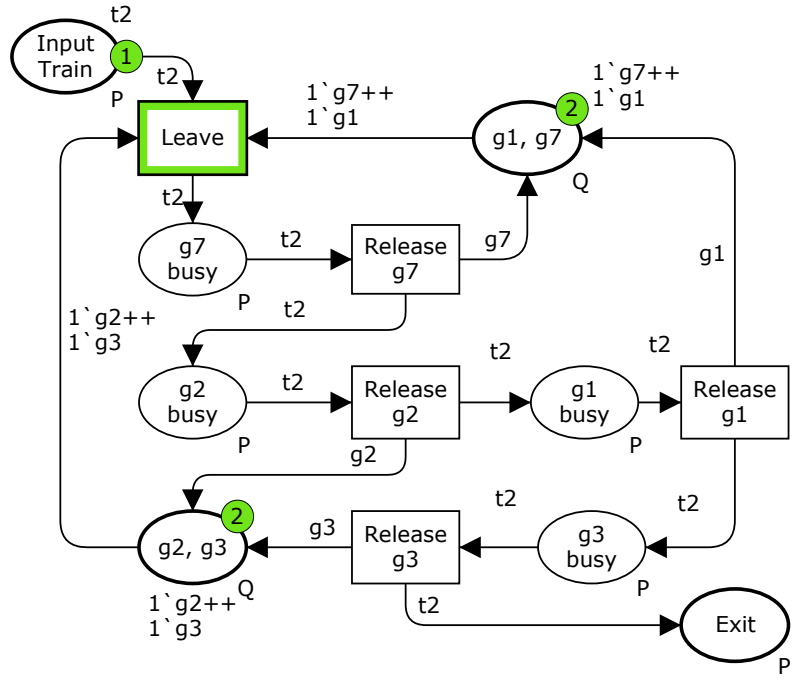


Figure 3. A timed CPN model of a train and a track circuit section.



(a)



(b)

Figure 4. CPN models for trains t1 and t2.

The first step of the CPN model is to represent the movements of individual trains over their designated routes in the basic model. The movements of trains t1 and t2 over routes r1 and r2 in Figure 1 are illustrated in Figure 4 (a) and (b) respectively.

The transition 'Enter' is the event that train t1 starts entering the station via the entry and 'Leave' represents the event of train t2 starts departing from the station. As the track sections g1 and g2 are common to r1 and r2, the places 'g1 busy' and 'g2 busy' and the transitions 'Release g1' and 'Release g2' appear on both Figure 4 (a) and (b). With t1 and t2 running in opposite directions, they proceed through g1 and g2 in opposite sequences. The next step of the model is to combine the CPN diagrams through the interactions at the common track sections, in addition to constraints from the timetable.

If the timetable stipulates that t2 starts its departure 300s ahead of the entry of t1, the combined CPN diagram, as shown in Figure 5, incorporates the spatial and temporal constraints. The starting point is the combined place 'Trains' and the initial marking 1`t1@300+++1`t2@0 represents there are one t1 and one t2 in the place, and the scheduled starting time of t1 at 300s and that of t2 at 0s. The places 'g1 busy' and 'g2 busy', and transitions 'Release g1' and 'Release g2' are now common to both t1 and t2. The interlocking relationships of the track sections reflect the spatial constraints. When timetable and run-times of trains through each track section (for simplicity, 50s for t1 for all track section; and 30s for t2) are given, the temporal constrains between places are taken in. A simple example is that the timestamp on the place 'g7 busy' indicates 30s for t2 to keep g7 from being released.

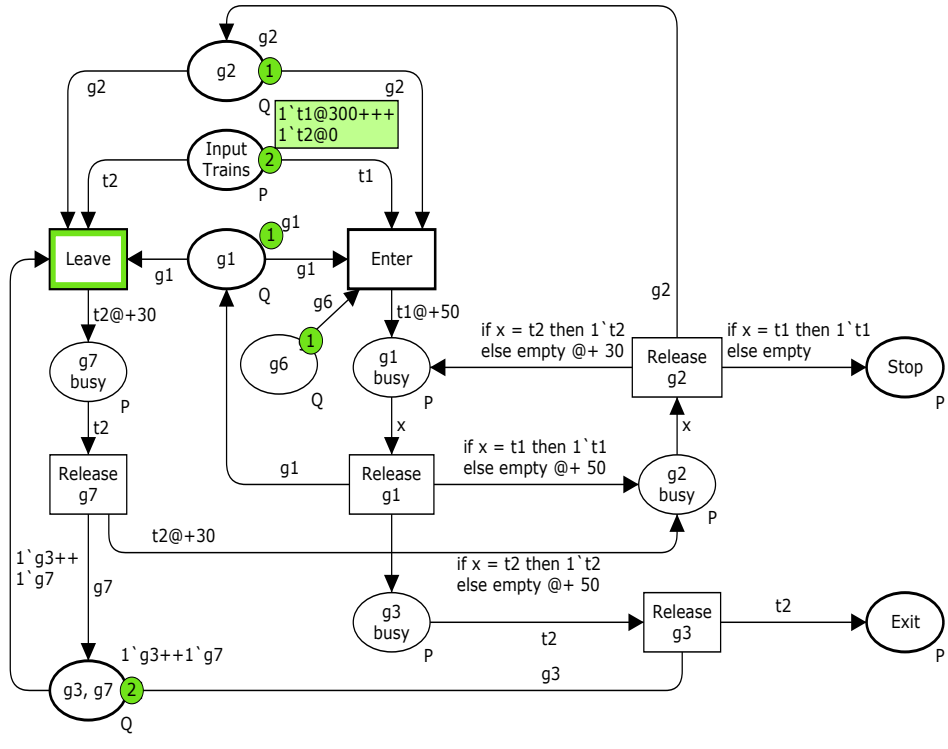


Figure 5. Combined CPN model for trains t1 and t2.

From the transition 'Release g1', the next event to be triggered (either the place 'g2 busy' or 'g3 busy') depends on the 'input' to the transition. To ensure an unambiguous

model, a variable x is introduced in the inscription of the arc from 'g1 busy' to 'Release g1'. The Color Set of x is P defined in Figure 3, representing the train at the place 'g1 busy'. In addition, the out-going arcs from 'Release g1' are also defined according to the value of x. The transition 'Release g2' is also specified similarly.

From Figure 5, t1 is allowed to start only when the scheduled time is reached and the route (g1, g2 & g6) is clear. The latter is now bound the movement of t2 which starts earlier. It is shown on Figure 5 that the transition 'Enter' is cleared to proceed when t2 has gone through g7, g2 and g1 and a total of 90s is elapsed. In other words, the buffer time is implicitly included in the model, i.e. 300s – 90s = 210s. Further, it is also possible to include service disturbances as additional places and transitions in the model.

Conclusions

Timetables and station layout impose spatial and temporal constraints on the TAP. A timed CPN model is proposed to encapsulate both constraints in a systematic depiction while the buffer time also is embedded in the model implicitly. This model enables proper evaluation of possible solutions for TAP with respect to robustness and thus service quality.

Acknowledgments

This work is supported by the National Scientific Foundation of China (70971010), the Research Committee of the Hong Kong Polytechnic University (PolyU G-U500) and Volvo Research and Educational Foundations.

References

Zwaneveld, P.J., Kroon, L.G. and Van hoesel, S.P.M. (2001). "Routing Trains through a Railway Station Based on a Node Packing Model." *European Journal of Operation Research*, 128(1), 14–33.

Carey, M., and Carville, S. (2003). "Scheduling and plat forming trains at busy complex stations." *Transportation Research Part A*, 37(3), 195-224.

Ho, T.K., Mao, B.H., Yuan, Z.Z., Liu, H.D., and Fung, Y.F. (2002). "Computer Simulation and Modeling in Railway Applications." *Computer Physics Communication*, 43(1), 1-10.

Dotoli, M., Fanti, M.P., and Mangini, A.M. (2008). "Real time identification of discrete event systems using Petri nets." *Automatica*, 44(5), 1209-1219.

Zhang, H.H., and Gu, M. (2009), "Modeling job shop scheduling with batches and setup times by timed Petri nets." *Mathematical and Computer Modeling*, 49 (1-2),

286–294.

Hagalisletto, A.M., Bjork, J., Yu, I.C., and Enger, P. (2007). "Constructing and refining large-scale railway models represented by Petri nets." *IEEE Transaction on System, Man and Cybernetics. Part C*, 37(4), 444-460.

Jensen, K., Kristensen, L., and Wells, L. (2007). "Coloured Petri Nets and CPN Tools for modelling and validation of concurrent systems." *International Journal on Software Tools for Technology Transfer*, 9 (3), 213-254

A Rolling Optimization Approach of Train Scheduling of Passenger Dedicated Lines

Ying Wang, Jianrui Miao, Fang Lei and Jun Liu

School of Traffic and Transportation Beijing Jiaotong University, Beijing 100044, China, postdoctor, Email: wangy@bjtu.edu.cn

Abstract

Train scheduling of passenger dedicated lines pay much attention to provide convenience and express service. It is different from train scheduling of exiting railway which is base on determined passenger train lines and mainly schedule freight train lines. This paper builds an integer linear programming model with weighted sum of passengers' travel time and waiting time when transfer between trains as objective, and proposes a rolling optimization approach which solve the model hour-by-hour. The numerical example shows that this rolling approach can solve the problem more efficiently than directly solve the model.
Key Words: passenger dedicated lines; train scheduling; integer linear programming; rolling approach

Introduction

Now, thousands kilometers Passenger Dedicated Lines with different high speed are being built in China. As one of the most challenging and difficult problems in railway operations planning, train scheduling has an important impact on the quality and level of service provided to the public. Also, train scheduling has been drawing the attention of researchers for decades.

Literature Review

For now, domestic researchers mainly focus on the train scheduling problems of existed rail lines (Peng Q.Y., 2001, Shi F., 2005, Zhou L.S., 1998 and Sun Y.,1997). The train scheduling problem of existed lines is to determine the timetable of freight trains with initial departure time and determined timetable of passenger trains. And most of existed researches developed heuristic algorithms to solve this problem.

There are also some researches on train scheduling problem of Chinese High-speed railway (Nie L., 1997, Shi H., 2000 and Ma J.J., 2002) which pay more attention on the determination of qualitative principles, the train scheduling pattern and some heuristic methods which mainly simulate artificial experiences and have low optimization level. However, foreign researchers usually use mathematical programming model to optimize the train scheduling (Bussieck, M.R., 1997, Lindner, T., 2000, Keican G., 2004, Ariano A. D., 2007 and Liu S.Q., 2010). And most of them focus on periodic train scheduling which is different from general train scheduling problem. In these literatures, they usually solve the model by relaxing the constraints and using various branches and bound strategies or directly solve using mathematical software. Besides these, Dorfman (Dorfman M., 2004) also developed a new method to solve the problem which built a discrete event model (DEM) to simulate the train scheduling process and finally generate a timetable when the simulation is end. This method could generate more accurate train schedules and have quick computing speed. However, its optimization level relies on the design of the algorithm whose optimization still needs more research.

However, the train scheduling of the PDLs is different from the existed researches because there are only passenger trains on it. Also, as the scale of the PDLs network is big and its line mileage is long, its train schedule cannot totally use the periodic schedule; aperiodic schedule will still be used in long distance and the scale of the problem will be much bigger. Therefore, this paper focuses on the characteristics of the train scheduling of PDLs and develops a rolling optimization approach based on mathematical optimization to solve the problem.

The Mathematical Model

Variables definition. In the mathematical model the following notation will be used.

T' Start time of maintenance gap.

T'' End time of maintenance gap.

T Set of trains.

F Set of sections.

f_o the start station of section f in down direction, $\forall f \in F$.

f_d the end station of section f in down direction, $\forall f \in F$.

W_i weight of traveling time of train i, $\forall i \in T$.

d_i run direction of train i, $\forall i \in T$, d_i =1 represents the down direction and d_i =2. represents the up direction.

$o(i)$ the origin station of train i, $\forall i \in T$.

$d(i)$ the destination station of train i, $\forall i \in T$.

D_i^U the earliest departure time of train i, $\forall i \in T$.

D_i^L the latest departure time of train i, $\forall i \in T$.

δ_i^s the addition departure time of train i at station s, $\forall i \in T, s \in S$.

φ_i^s the addition stop time of train i at station s, $\forall i \in T, s \in S$.

r_i^{fd} the running time of train i with down direction in section f, $\forall i \in T, f \in F$.

r_i^{fu} the running time of train *i* with up direction in section *f*, $\forall i \in T, f \in F$.

τ_i^s the minimum dwell time of train *i* at station *s*, $\forall i \in T, s \in S$.

I_{ij}^f the minimum time intervals between two successive train *i* and train *j* in section *f*, $\forall i \in T, j \in T, f \in F$.

I_{ij}^{sf} the minimum time intervals between two successive train *i* and train *j* when they depart from station *s*, $\forall i \in T, j \in T, s \in S$.

I_{ij}^{sd} the minimum time intervals between two successive train *i* and train *j* when they arrive at station *s*, $\forall i \in T, j \in T, s \in S$.

I_{ij}^{st} the minimum time intervals between two successive train *i* and train *j* when they pass station *s*, $\forall i \in T, j \in T, s \in S$.

P_{ij}^s the number of passengers transfer from train *i* to train *j* at station *s*, $\forall i \in T, j \in T, s \in S$.

α_i^f binary variable, if train *i* pass section *f* α_i^f =1; otherwise, α_i^f =0, for $\forall i \in T, f \in F$.

Decision variables:

γ_{ij}^f binary variable, if train *i* runs before train *j* in section *f* γ_{ij}^f =1; otherwise, γ_{ij}^f =0, for $\forall i, j \in T, f \in F$, and $d_i = d_j, \alpha_i^f = \alpha_j^f = 1$.

x_i^s binary variable, if train *i* stop at station *s* x_i^s =1; otherwise, x_i^s =0, $\forall i \in T, s \in S$

a_i^s the departure time of train *i* at station *s*, $\forall i \in T, s \in S$.

b_i^s the arrival time of train *i* at station *s*, $\forall i \in T, s \in S$.

Formulation of objective function

Considering the high service level needed by operation of PDLs, this paper mainly considers two objective criteria that are related to service quality.

(1) Minimize the sum of trains' weighted traveling time that is calculated by the sum of every train's traveling time multiplies the weight of its level. The objective ensures minimum stop time unnecessary which is usually caused by trains' overtaking. This objective reveals the high speed of PDLs.

$$\min \sum_{i \in T} W_i \, (d_i^{o(i)} - a_i^{d(i)}) \tag{1}$$

(2) Minimize weighted passengers' transfer time, which is to sum the transfer time by using the transfer number as the weight. This objective reduces passengers' waiting time when transfer at station. This objective reveals the demand of passenger transfer convenience.

$$\min \sum_{i \in T} \sum_{j \in T} \sum_{s \in S} P_{ij}^s (d_j^s - a_i^s) \tag{2}$$

These two objectives together make passengers arrive the destination of their journey as soon as quickly. Considering those two objectives are not conflict and both computed in time, these two objectives can be weighted sum to be one objective. Define ω_1 as the weight of objective (1), and ω_2 as the weight of objective (2), and the objective function can be formulated as:

$$\min \ \omega_1 \sum_{i \in T} W_i \ (d_i^{o(i)} - a_i^{d(i)}) + \omega_2 \sum_{i \in T} \sum_{j \in T} \sum_{s \in S} P_{ij}^s (d_j^s - a_i^s) \tag{3}$$

Formulation of constraints

(1) Lower and upper bound of departure time constraints

According to the train service plan, the departure time of each train at their original station must be in the time span given by train service plan.

$$d_i^{o(i)} - D_i^U \le 0 \qquad i \in T \tag{4}$$

$$d_i^{o(i)} - D_i^L \ge 0 \qquad i \in T \tag{5}$$

(2) Running time constraints

A train's running time in a section includes trains' pure running time, start and stop additional time if it stops at the start station and the end station of the section. This constraint can be formulated as:

$$a_i^{f_d} - d_i^{f_o} - x_i^{f_o} \delta_i^{f_o} - x_i^{f_d} \varphi_i^{f_d} = r_i^{fd} \qquad i \in T, f \in F, d_i = 1, \alpha_i^f = 1 \tag{6}$$

$$a_i^{f_o} - d_i^{f_d} - x_i^{f_d} \delta_i^{f_d} - x_i^{f_o} \varphi_i^{f_o} = r_i^{fu} \qquad i \in T, f \in F, d_i = 2, \alpha_i^f = 1 \tag{7}$$

(3) Dwell time constraints

A train must stop longer than the least dwell time at the station on its travel route.

$$d_i^s - a_i^s \ge \tau_i^s \qquad i \in T, s \in S \tag{8}$$

(4) Headway constraints at stations

To double-track railway, several minimum time intervals between two successive trains need to satisfy when they running, arrive, depart and pass in the same direction. The headway constraints in sections for trains with down and up direction can be formulated as formula (9) and formula (10).

$$a_i^{f_d} - a_j^{f_d} - \gamma_{ij}^f I_{ij}^f \ge 0 \qquad i, j \in T, f \in F, d_i = d_j = 1, \alpha_i^f = \alpha_j^f = 1 \tag{9}$$

$$a_i^{f_o} - a_j^{f_o} - \gamma_{ij}^f I_{ij}^f \ge 0 \qquad i, j \in T, f \in F, d_i = d_j = 2, \alpha_i^f = \alpha_j^f = 1 \tag{10}$$

And the headway constraints at stations can be formulated as follows:

for $\forall i, j \in T, f \in F$, and $d_i = d_j = 1, \alpha_i^f = \alpha_j^f = 1$, they must satisfy these constraints as follows:

$$d_i^{f_o} - d_j^{f_o} - I_{ij}^{f_o f} + M(2 - x_i^{f_o} - \gamma_{ij}^f) \ge 0 \tag{11}$$

$$a_i^{f_d} - a_j^{f_d} - I_{ij}^{f_d d} + M(2 - x_i^{f_d} - \gamma_{ij}^f) \ge 0 \tag{12}$$

$$a_i^{f_d} - a_j^{f_d} - I_{ij}^{f_d d} + M(2 - x_j^{f_d} - \gamma_{ij}^f) \geq 0 \quad (13)$$

$$d_i^{f_o} - d_j^{f_o} - I_{ij}^{f_o t} + M(1 - x_i^{f_o} - \gamma_{ij}^f) \geq 0 \quad (14)$$

for $\forall i, j \in T, f \in F$, and $d_i = d_j = 2, \alpha_i^f = \alpha_j^f = 1$, they must satisfy these constraints as follows:

$$d_i^{f_d} - d_j^{f_d} - I_{ij}^{f_d f} + M(2 - x_i^{f_d} - \gamma_{ij}^f) \geq 0 \quad (15)$$

$$a_i^{f_o} - a_j^{f_o} - I_{ij}^{f_o d} + M(2 - x_i^{f_o} - \gamma_{ij}^f) \geq 0 \quad (16)$$

$$a_i^{f_o} - a_j^{f_o} - I_{ij}^{f_o d} + M(2 - x_j^{f_o} - \gamma_{ij}^f) \geq 0 \quad (17)$$

$$d_i^{f_d} - d_j^{f_d} - I_{ij}^{f_d t} + M(1 - x_i^{f_d} - \gamma_{ij}^f) \geq 0 \quad (18)$$

In these formulas, *M* is an integer big enough which can transform the non-linear headway constraints into linear constraints. For example, formula (11) represents that when the latter train departs the start station of the section, the time interval between the latter train's departure and the former train's departure must larger than the minimum time interval named departure time interval. Similarly, formula (12) and (12) represent that when the latter train arrives the end station of the section, the time interval between the former train's arrival time and the latter train's arrival time must lager than the minimum arrival time interval named arrival time interval. Formula (14) represents that when the latter train passes the start station of the section, the time interval between the latter train's departure time and the former train's departure time must lager than the minimum time interval named pass time interval. Correspondingly, the headway constraints of trains in up direction can be formulated as formulas (15)-(16).

(6) Trains' order constrains

$$\gamma_{ij}^f + \gamma_{ji}^f = 1 \qquad i, j \in T, f \in F, \text{且} d_i = d_j, \alpha_i^f = \alpha_j^f = 1 \quad (19)$$

(7) Maintenance gap constraints

Trains must depart later than the end of maintenance gap and arrive earlier before the beginning of maintenance gap.

$$d_i^s \geq \mathrm{T}' \qquad i \in T, s \in S \quad (20)$$

$$a_i^s \leq \mathrm{T}'' \qquad i \in T, s \in S \quad (21)$$

(8)Transfer time constraints

$$P_{ij}^s (d_j^s - a_i^s) \geq 0 \qquad i, j \in T, s \in S \quad (22)$$

(9) Decision variable constraints

$$x_i^s \tau_i^s \geq \tau_i^s \qquad i \in T, s \in S \quad (23)$$

The above formulas (3)—(23) consists of the train scheduling model of PDLs.

Solution

The mathematical model is a liner integer programming problem. The number of decision variables (mainly the variable γ_{ij}^{f}) increases exponentially with the number of trains and sections and the model is a NP-hard problem. When the number of trains is large, it's really time consuming to solve the model directly. By analyzing the objective and the constraints, this paper concludes that the most difficulty of this model is to determine the order of trains in every running section and avoid being conflict which usually in a small time span. Therefore, this paper proposes a rolling optimization approach that solve the model hours-by-hours.

By splitting the period of the model into several small time periods like 1-3hour each time period (concrete value is determined by the scale of the problem), the problem can be split into several small models with less decision variables which are solved one-by-one. In each period, due to the greatly decrease of train numbers, we could directly use commercial mathematical software to solve the model. In this way, through reasonably setting the period length and the trains considered in each period as well as the processing of period connection, we could solve the model rapidly. As the maintenance gap of PDLs is rectangular, we begin with the end of the maintenance gap to solve the model period-by-period. The three key techniques of this approach are illustrated as follows:

(1) The preprocessing of train distribution in each time period. The train distribution determines the number and type of trains in each period. Define N as the number of trains, N_j as the number of j type of trains, β_{ij} as the proportion of the trains' number in hour i to the total number of trains, n_{ij} as the number of j type of trains in hour i. The value of β_{ij} is known. So the value of n_{ij} can be calculated as $n_{ij} = \beta_{ij} \times N_j$.

(2) The processing of time period connection. Some trains' running time from their origination and their destination may across two time periods. For this situation, these trains are considered in all the two time period. There departure time and arrival time are first determined in the former period. Then, the time determined to be in the latter period are set to be unknown and are solved again in the latter period. Furthermore, if there are passengers needing to transfer from one train in the first period to another train, these two trains are solved in the former period and the determine departure time and arrival time of these two trains are considered in the latter period as input.

(3) The processing of no feasible solution in a time period. If there are many trains in a period, there may be no feasible solution because of headway constraints. If there is no feasible solution in a period, remove trains in this period into next periods until there is feasible solution. Then go on solving the model in the rest period. If there is no feasible solution in the last period, adjust the train distribution and resolve the model from the first period. The flow chart of the rolling approach is shown as Figure 1:

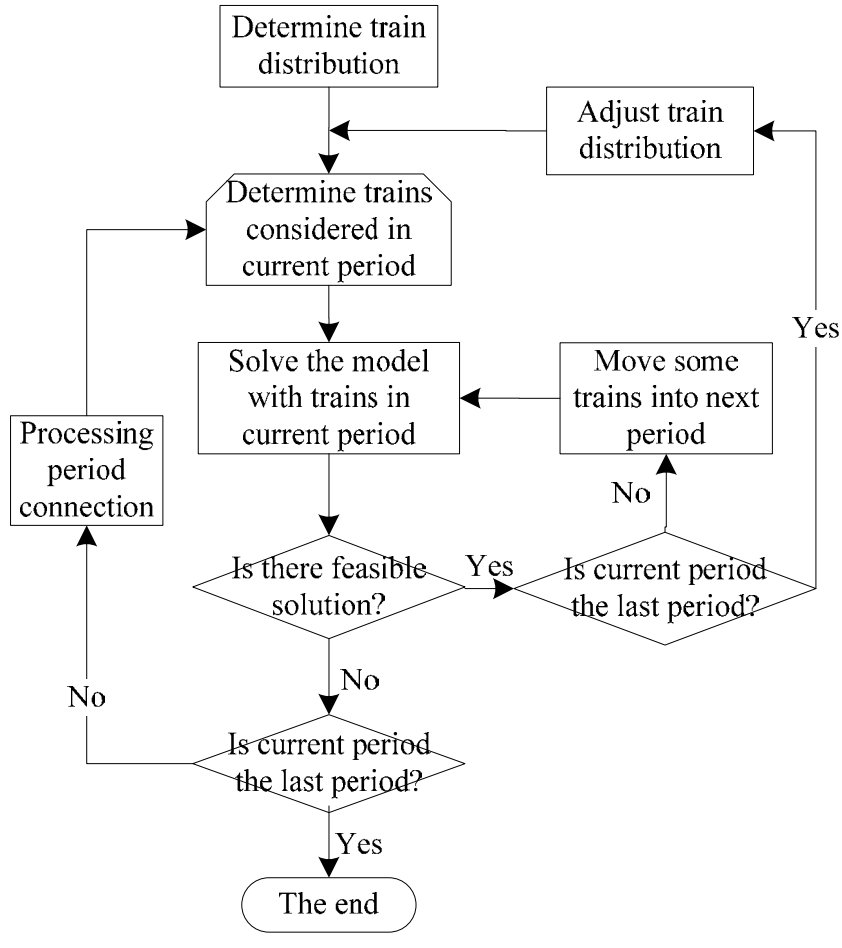


Figure1. The flow chart of the rolling approach.

Numerical Example

The effectiveness of this rolling optimization approach is testified by data generated according the train diagram of PDL from Beijing to Tianjin. The values of parameters are showed in Table 1. And the values of parameters in different section of different type of trains are seemed to be the same for simple. The rolling approach is performed in a personal computer with Intel Pentium4 CPU 2.80GHz processor. The train schedule generated is shown in Figure 2. The performance of the rolling approach is also compared with directly approach (Table 2).

Table 1. Parameters of train scheduling.

N	ω_1	ω_2	T'	T''	τ_i^s	r_i^{fu}
120	1	1	24:00	6:00	5min	22min
r_i^{fu}	δ_i^s	φ_i^s	I_{ij}^f	I_{ij}^{sf}	I_{ij}^{sd}	I_{ij}^{st}
22min	1min	2min	3min	5min	5min	4min

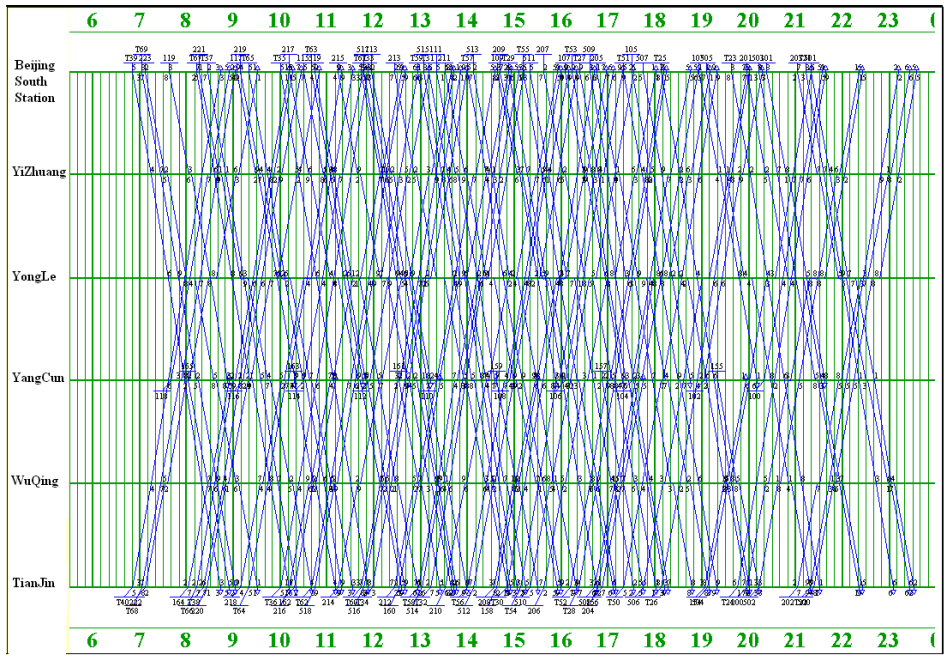


Figure 2. The train schedule generated by this rolling approach.

Table 2. Comparison of rolling approach and directly approach.

	Period length	Computing time	Memory needed	Maximum train number	Objective	Lower Bound
Rolling approach	1 h	6 min	≤32 MB	32	12839 min	12720 min
	2 h	15 min	≤32 MB	46	12839 min	12720 min
Directly approach		＞11 h	＞256 MB	120	unknown	12720

The numerical example shows that this rolling approach can solve the problem more efficiency than directly solve the model. And in the numerical example this approach got the optimization solution the same as directly solving which can be conclude by the integer gap between the objective and the lower bound of the model.

Conclusions

This paper focuses on the characteristics of train scheduling of PDLs, builds a mathematical model and proposes a rolling approach based on mathematical optimization. The numerical example shows that this rolling approach can solve the problem more efficiency than directly solve the model and usually have good performance on optimization.

Acknowledgement

This work was supported by Ministry of Education of the People's Republic of China (No. 20060004025) and National Natural Science Foundation of China (No.60736047). We deeply appreciate the support.

References

Ariano A.D., Pacciarelli D., and Pranzo M. (2007). "A branch and bound algorithm for scheduling trains in a railway network." *European Journal of Operational Research*, 183(2), 643-657

Bussieck, M.R., Winter, T., and Zimmermann, (1997). "U.T. Discrete optimization in public rail transport. Mathematical Programming." 79(1-3), 415–444

Dorfman, M., and Medanic, J. (2004). "Scheduling trains on a railway network using a discrete event model of railway traffic." *Transportation Research Part B*, 38(1), 81-98

Keican G., Ferenc S., and Mohammad J.A. (2004). "A multi-objective train scheduling model and solution." *Transportation Research Part B*, 38(10), 927-952

Lindner, T. (2000). "Train schedule optimization in public rail transport." Ph.D. thesis, *Technische Universitat Braunschweig*

Liu A.Q., and Kozan E. (2009). "Scheduling trains as a blocking parallel-machine job shop scheduling problem." *Computers & Operations Research*, 36(10), 2840-2852

Ma J.J., HU S.J., and XU H. (2002). "Theory study on railway network based on train working diagram of Jing-Hu high-speed railway line." *Journal of the china railway society*, ,24(3), 1-4

Nie L., Hu A.Z., and Guo F.E. (1997). "Study on automatic method of moving train paths on high speed railway train diagram." *China railway science*,18(4), 108-116

Peng Q.Y., Zhu S.N., and Wang P. (2001). "Study on a general optimization model and its solution for railway network train diagram." J*ournal of the china railway society*, 23(1), 1-8

Shi H. (2000). "Analysis on rational applied modes of high-speed train diagram in our country." *Journal of the china railway society*, 22(1), 92-97

Shi F., Li X.H., Qin J., and Deng L.B. (2005). "A timing-cycle iterative optimizing method for drawing single-track railway train diagrams." *Journal of the china railway society*, 27(1), 1-5

Sun Y., Li Z.Z., and Li F.L. (1997). "The optimal original trains distribution for traveling graph of trains." *Journal of Changsha railway university*, 15(4), 80-87

Zhou L.S., Hu S.J., Ma J.J., and Yue Y.X. (1998). "Network hierarchy parallel algorithm of automatic train scheduling." *Journal of the china railway society*, 20(5), 15-21

Improvement of CFG Pile Corrosion Resistance Capability in the Application of Wu-Guang Passenger Transport Special Railway

Kunlin Ma[1,2], Youjun Xie[1], Guangcheng Long[1] and Fei Feng[1]

[1] School of Civil Engineering and Architecture, Central South University, Changsha 410075, China; Email: mark-mkl@163.com
[2] Hunan Industry Polytechnic, Changsha 410208, China

Abstract

With Chinese economy speedup since late 1990s, it was required that large numbers of construction of civil engineering projects, and passenger transport special railway was the results of economic development. CFG (cement fly ash gravel) is one of groundwork reinforced methods widely used in civil engineering. And CFG was also applied in the construction of Wu-Guang passenger transport special railway. However, according to field investigation, it was found that groundwater in Haudu to Shaoguan in Gunagdong province containing lots of sulphate, which can result in corrosion damage of underground CFG pile. Therefore, in order to improve service life of CFG pile, it is necessary to increase the capability of CFG corrosion resistance. In this paper, influence of compositive materials such as water to binder rate (*w*/*b*), fly ash to cement rate (F/C) and quantity of cement on mass change of CFG in sulfate sodium solution, permeability coefficient and compressive strength of CFG were studied. Results show that with the increase of w/c, mass of CFG increase quickly first and then keep steadily. The increase of F/C and cementious addition in *CFG* can decreases the mass increase rate of CFG effectively. The increase of *w/c* and cementious addition increase permeability coefficient of CFG. And there exits an optimal F/C (1.0–1.2) in CFG , making permeability coefficient of CFG decrease to the least at 28 day ages. When CFG is immerged in water and sulfate solution respectively, the compressive strength of CFG immerged in sulfate solution is higher than that immerged in water at 28 day ages. With *w*/*c*, F/C and cementious addition increasing, the compressive strength of CFG decreases. There are acicular ettringite crystals in CFG after immerged in sulfate solution.

Key Words: CFG pile; corrosion; permeability coefficient; compressive strength; passenger transport special railway

Introduction

Cement fly ash gravel (CFG) pile is a kind of groundwork reinforced methods widely used in civil engineering, composed by cement, fly ash, gravel, stone chip, sand and water. According to adjust the cement content and other compositions of CFG pile, the compressive strength of CFG pile is in the range of C5–C25. And there are great differences between CFG pile and blank concrete because of the compositive materials. The compositive materials of CFG pile contains lots of industrial scrap such as fly ash, which replaces much cement in CFG pile, therefore the increase of later strength of CFG is obvious (Zhao, 2004) . The characteristics of CFG pile are as follows: (1) CFG pile can take full of the carrying capability between pile and soil. (2) The elasticity module and the carrying capability characteristics of CFG pile possess the characteristics of rigid pile, which can play notable role in the replacement of compound groundwork. (3) It is easy to construct and then shorten the time of construction. (4) There are not steel bar in CFG pile, and compared to cement content in normal pile used in groundwork the cement content in CFG pile is only 50%–80%, so CFG pile possesses low cost. (5) Compared to the cost of normal pile used in groundwork the cost of CFG pile decrease 40%–60%. (6) CFG pile consumes lots of industrial scraps, showing favorable impact on society. Therefore, because of many advantages of CFG pile, it was widely used in civil engineering projects, including the construction of Wu-Guang passenger transport special railway.

Wu-Guang passenger transport special railway is an important traffic line in China, and it has been open to traffic in Dec. 12, 2009. But in the construction process of Wu-Guang passenger transport special railway, it was found that the groundwater in Huadu to Shaoguan in Gunagdong province containing lots of sulphate, which can result in corrosion damage of underground CFG pile. According to relative researches (Neville, 2004; Santhanam, 2001), when cement-based material immerging in sulfate solution, three kinds of corrosion products will appear-gypsum, thaumasite and ettringite, which can cause the capability decrease of cement-based material. Consequently, to improve the capability of corrosion resistances of CFG pile, in this paper the change characteristics of mass in sulfate sodium solutions, permeability, strength and microstructure of CFG pile under sulfate attack were studied detailed.

Experimental

Testing Specimen. The cement used was Grade of 42.5 ordinary Portland cement according to the Chinese standard. The sand used in the experiment had a fineness modulus of 2.88 and apparent density of 2.67 g/cm^3. The coarse aggregate used in the experiment was crushed limestone with size of 5 mm–25 mm and apparent

density of 2.64 g/cm^3, and crushed index of 7.8%. Graded II fly ash with a density of 2.41 g/cm^3 and specific surface area about 460 m^2/kg was used. All water used in this experiment was deminerialized water. Chemical compositions of cement and fly ash are given in Table 1. Mixture proportion and details of CFG is shown in Table 2.

Table 1. Chemical compositions of cement and fly ash.

	SiO_2	Al_2O_3	Fe_2O_3	CaO	MgO	SO_3	$I.L$
P.O 42.5	24.3	4.8	3.	55.3	4.2	2.2	2.4
FA	52.7	25.8	9.7	3.7	1.2	0.2	2.5

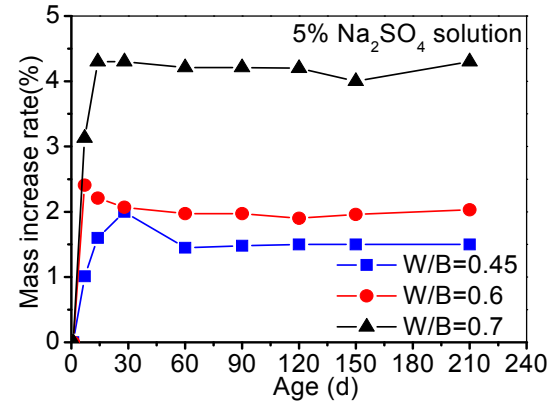
Table 2. Mixture proportion of CFG.

No.	Raw materials (/kg/m^3)					*w/c*	Fly ash to cement (F/C)
	Crushed	Stone	Cement	Fly ash	Water		
A-1	1190	473	177	195	168	0.45	1.1
A-2	1255	500	173	191	218	0.60	1.1
A-3	1190	473	163	180	240	0.70	1.1
B-1	1190	473	230	117	207	0.60	0.5
B-2	1255	500	173	191	218	0.60	1.1
B-3	1255	500	167	198	218	0.60	1.2
B-4	1253	500	143	217	217	0.60	1.5
B-5	1250	500	120	240	217	0.60	2.0
C-1	1347	537	150	166	190	0.60	1.1
C-2	1303	520	163	180	203	0.59	1.1
C-3	1255	500	173	191	218	0.60	1.1
C-4	1423	567	130	143	167	0.60	0.9

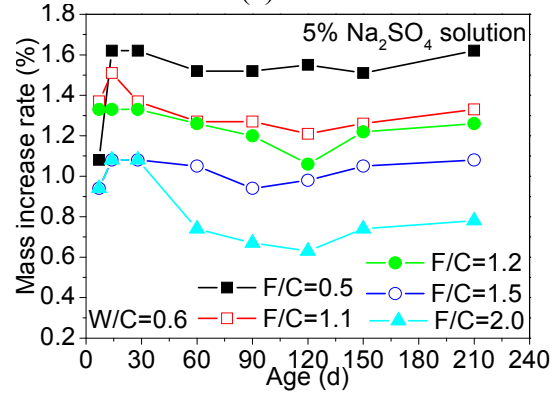
Experimental Procedures. The test of mass change, permeability coefficient and strength according to Chinese standard *GB/T 50081-2002*. The workability of CFG pile was tested according to Chinese standard *GB/T 500800-2002*. SEM was introduced in order to test microstructure of CFG attacked by sulfate solution.

Results and discussion

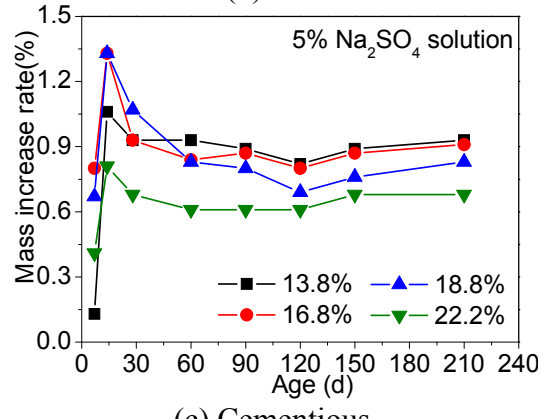
Mass Change. When cement-based materials are attacked by sulfate solution, the mass will increase, resulting in corrosion damage of cement-based material. Figure 1(a)–1(c) show the influence of compositive materials on mass change of CFG pile immerged in different sulfate solution. As can been seen from Figure 1 (a), with the increase of immerging time, the mass of CFG with *w/c* 0.45, 0.6 and 0.7 increases quickly first and then keeps steady gradually. And with the increase of *w/c*, mass increase rate of CFG increase. Porosity of CFG increases with the increase of *w/c*, thus big *w/c* can result in more sulfate solution ingress into CFG. Therefore, corrosion products are on the rise, resulting in the increase of mass.



(a) *w/c*



(b) F/C



(c) Cementious

Figure 1. Influence of compositive materials on mass change of CFG attacked by sulfate.

There are five different F/C on the mass increasing ratio of CFG immerging in 5% sulfate solution, as can be seen from Figure 2 (b), with the increase of immerging time, the mass of CFG increases quickly first and then decreases gradually. But after 120 d, the mass of CFG begin to increase. The increase of F/C can decrease the mass increase ratio of CFG, which means that addition of FA can improve the capability of resistance sulfate attack. Figure 1(c) shows the results of cementious addition on mass change of mass. As can be seen from Figure 1(c), in different cementious

additions the mass change characteristics of CFG are the same. Namely, mass of CFG increase quickly first and then decrease gradually. And the increase of cementious addition can decrease the mass increase rate. Increasing cementious addition can improve the compaction of CFG, so the capability of resistance sulfate solution attack increases.

Permeability. Permeability coefficient is a very important parameter for porous materials. In this paper, influence of compositive materials on permeability coefficient of CFG is studied, and Figure 2(a)–2(c) are the tested results. As can been seen from Figure 2(a), with the increase of w/c, permeability coefficient of CFG increase, which means that the capability of resistance sulfate solution of CFG decreases with w/c increasing. As can be seen from Figure 2(b), that there exists an optimal F/C (1.0–1.2) in CFG can make permeability coefficient of CFG decrease to the least. Figure 2(c) is the tested results of influence of cementious addition on permeability coefficient of CFG. As can be seen from Figure 2(c), the permeability coefficient of CFG increases appreciably with the increase of cementious addition. The phase of aggregate cuts off the permeable channels in CFG, but in the same volume the increase of cementious means to decrease aggregate, and the permeable channels in CFG increase, so permeability coefficient of CFG increases.

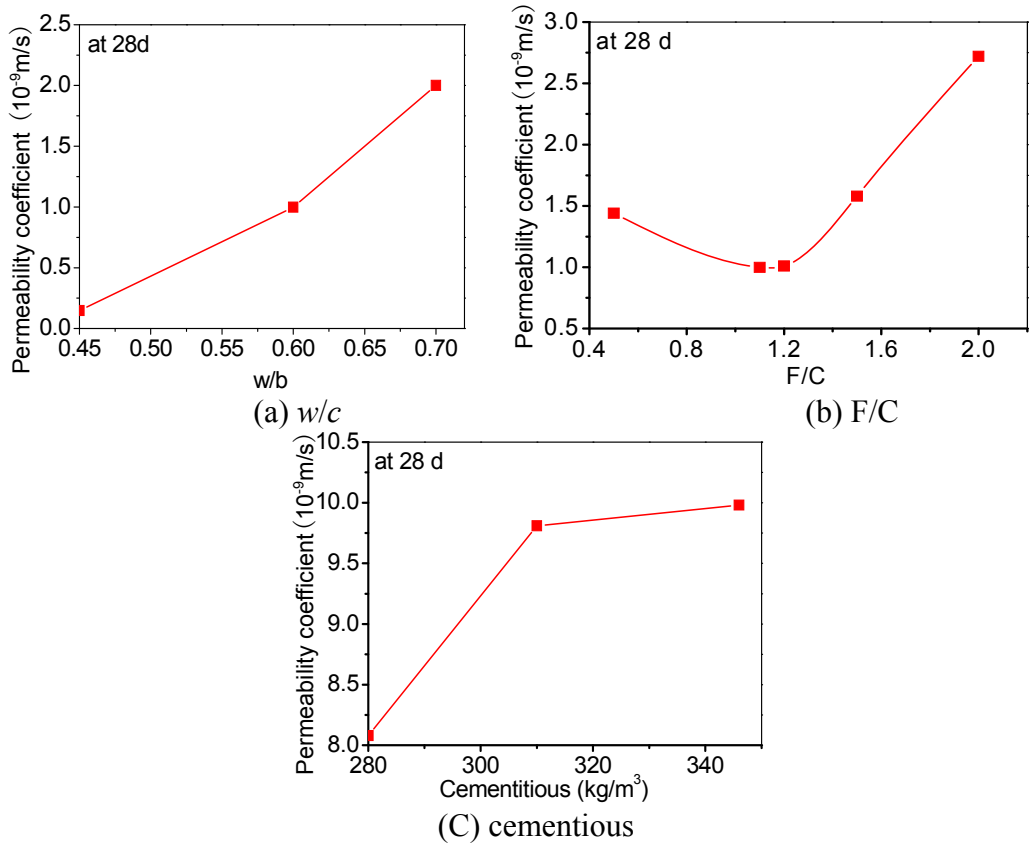


Figure 2. Influence of compositive materials on permeability coefficient of CFG attacked by sulfate.

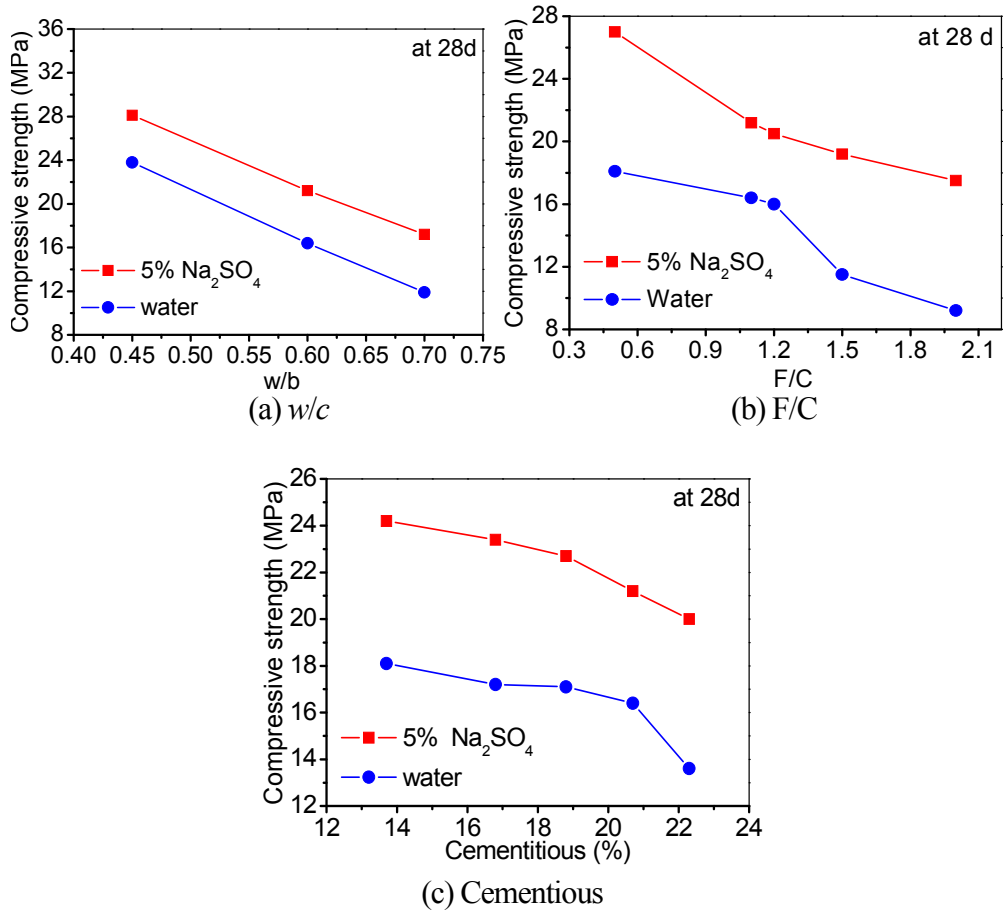
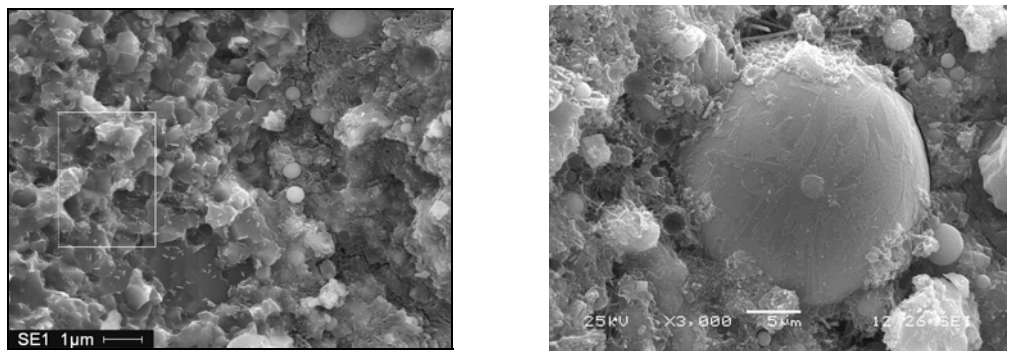


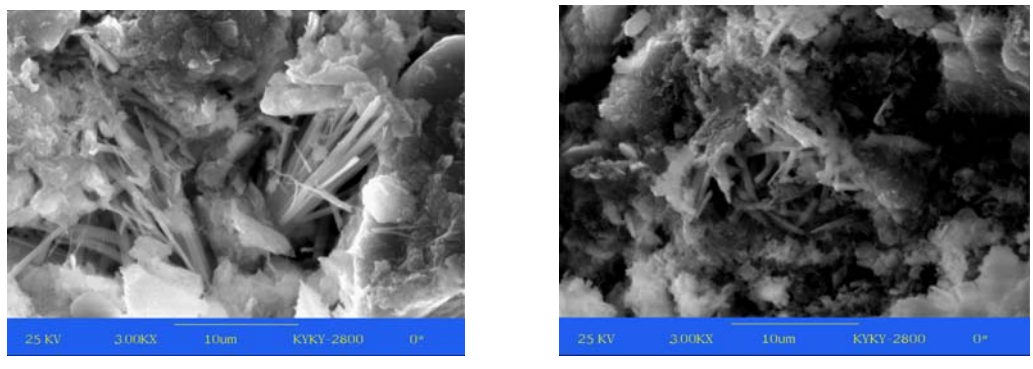
Figure 3. Influence of compositive materials and curing conditions on strength of CFG attacked by sulfate

Strength. Strength loss is a main result resulting from sulfate solution attack, and different curing conditions can also influence the strength of CFG. Figure 3 (a)～(b) show the results of compressive strength of CFG under sulfate solution attack. As can be seen from Figure 3(a), with w/c increasing compressive strength of CFG decreases linearly. As can been seen from Figure 3(b), with F/C increasing compressive strength of CFG decreases slowly. As can be seen from Figure 3(c), with cementious increasing compressive strength of CFG decreases slowly. And the strength of CFG in 5% Na_2SO_4 solution curing is higher than that in water at 28 day ages. This can be explained by that the active fly ash is activated by sulfate when CFG was immerged at 28 day ages.

Microstructure. Figure 4(a) and 4(b) are the microstructure of CFG immerging in water and 5% sulfate solution. As can be seen from Figure 4(a), lot of fly ash particles exist in CFG and some fly ash begin to take hydration. But as can be seen from Figure 4(b), after CFG has been immerged in sulfate solution for 120 d, some



(a) Photos of SEM of CFG immerging in water at 120 d (B-3)



(b) Photos of SEM of CFG immerging in sulfate solution at 120 d (B-3)

Figure 4. Microstructure of CFG in different conditions.

acicular crystals can be found. And according to the shape of these crystals and sulfate attack environment, it is suggestion that these acicular crystals are ettringite (Hobbs 2000; Brown 2002).

Conclusions

With the increase of w/c, mass of CFG increase quickly first and then keep steadily in sulfate sodium solution. The increase of F/C and cementious addition in CFG can decrease the mass increase rate of CFG effectively. The increase of w/c and cementious addition increase permeability coefficient of CFG. And there exits an optimal F/C (1.0–1.2) in CFG , which makes permeability coefficient of CFG decrease to the least at 28 day ages. When CFG is immerged in water and sulfate solution respectively, the compressive strength of CFG immerged in sulfate solution is higher than that immerged in water at 28 day ages. With the increase of w/c, F/C and cementious addition, the compressive strength of CFG decrease. There are corrosion products acicular ettringite crystals in CFG after immerged in sulfate solution.

References

Brown, P., and Hooton, R．D. (2002). “Ettringite and thaumasite formation in laboratory concretes prepared using sulfate-resisting cements.”*Cement and Concrete Composites*, 361–370.

Hobbs, D.W., and Taylor, M.G. (2000). “Nature of the thaumasite sulfate attack mechanism in field concrete.” *Cement and Concrete Research*, 529–533.

Neville, A．(2004). “The confused world of sulfate attack on concrete.” *Cement and Concrete Research*, 1275–1296.

Santhanam, M., Cohen. D., and Olek, J. (2001). “Sulfate attack research-whither now？”*Cement and Concrete Research*, 845–851.

Zhao, W. L., and Shi, R. (2004). “Design of mix Proportion with different strength grades in composite foundation with CFG piles.” *Architecture technology*, 196–200.

Multistage Linear Prediction Model of Track Quality Index

Huan Chang[1], Rengkui Liu[2] and Wei Wang[3]

[1,3] Graduate student, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China
[2] Associate Professor, State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University. Beijing 100044, China, Email: hichanghuan@163.com

Abstract

Both Chinese and foreign researchers have developed linear models to predict track irregularities. These models adopt the Track Quality Index (TQI) as an index to evaluate overall track irregularity. In these models, an averaging method is employed to calculate the slope k of track irregularity over a particular track section and *k* is then used to predict future TQIs. However, complicated factors affect track irregularity. Roadbed state, passing tonnage, train running speed and track structure are all important factors that determine track irregularity changes. Under the joint action of various factors, changes in track irregularity are characterized by nonlinearity, which can be described the so-called Bathtub Curve. Linear prediction models provide a simple but convenient tool to predict the evolution of track irregularities. In a certain sense, these models help to deepen research and can shed new light on the pattern of track irregularity changes. However, these models inevitably have two obvious disadvantages. The first is that no systematic consideration is given to the inherent characteristics of track irregularity changes, while the second is that it is apparently impractical to use a single slope *k* to predict track irregularities for all track sections of a whole railway line. The result is that when linear models are used to predict track irregularity changes, big predictive errors are caused. In the light of the cyclic, exponential and multistage changes in track irregularity, the current authors have developed a multistage linear prediction model for track irregularity. Given the cyclic nature of track irregularity changes, we take a 200 m track section as the unit section for our research. Because track irregularities change at different rates within one track irregularity development cycle, we divide such changes into several shorter stages. Based on our understanding of the three major characteristics of track irregularity changes, we put

forward an improved linear prediction model, namely the multistage linear prediction model (MLPM). The model adjusts the calculation method of *k* to make the predicted results fit better with the actually measured values. The authors finally use track car inspection data from the Beijing-Shanghai Railway Line (Jing-Hu Line) to validate the model and compare the predicted results with the measured data.
Key words: railway; track; inspection data; TQI; prediction

Introduction

To meet the needs of increasing freight volumes and ensure the increase of transport capacities, high speed and heavy trains have become a development trend. To ensure safe railway transportation, both the railway management authorities and experts in various countries have conducted research on track irregularities. Employing the stochastic process theory, Iyengar and Jaiswal have analyzed track surface data obtained from India's two railway lines and concluded that track surface can be modeled as a stationary Gaussian stochastic process (Iyengar, 1995). Alfelor et al. have fitted linear equations to the loaded and unloaded gauge data measured by the Gage Restraint Measurement System (GRMS) to describe the relationship between gauge and each of the factors affecting it (Alfelor, 2001). Hyslip has conducted fractal analysis of track geometry data and developed numerical indices for assessing track condition (Hyslip, 2002). Kawaguchi et al. at Railway Technical Research Institute in Japan have respectively developed degradation models which use standard deviations of track irregularities to predict the standard deviations of both track alignment and surface of 100 m-long track sections(Kawaguchi, 2005). Chen et al. at the China Academy of Railway Sciences have developed an integrated factor method which uses TGC-generated track waveform data to predict track irregularities (Chen Xian-mai, 2006). Professor XU Yu-de et al., at Tongji University in China, has put forward a linear prediction model of track irregularity, which adopts the averaging and least squares methods (Xu, 2005). The predicted result is defined by

$$TQI=TQI_0+kt \tag{1}$$

where TQI_0 is the initial value of TQI; k is the slope; and t is time.

Professor Xu et al. have used TQI to examine track irregularity changes in the Guangzhou-Shenzhen Railway Line. k is calculated through the averaging and least squares methods. However, complicated factors affect track irregularities. Roadbed state, passing tonnage, train running speed and track structure are all important factors that determine changes in track irregularity. Under the joint action of various factors, changes in track irregularity are characterized by nonlinearity. Such nonlinear changes have been shown to match with the characteristics of the Bathtub Curve. Consequently, a single nonlinear model to predict the TQI change process often results in considerable discrepancies with the real world situation. Overall, most existing models tend to consider the entire factors affecting the track irregularities. Factually, the results were not that satisfied. Based on the characteristics of track irregularity changes, we put forward a multistage linear model to fit the process of TQI changes.

Model Building

Data Source. Track inspection cars (TGC) are the main tool used to make dynamic track inspections in China. Currently, 3 main types of TGCs are used by China's railway management department, namely GJ-4 TGC, Enhanced GJ-4 TGC and GJ-5 TGC. Enhanced GJ-4 TGC is the most commonly used one of the three. GJ-4 TGC is a strap down measuring system developed by the China Academy of Sciences. First put into use in 1999, GJ-4 TGC samples and processes track geometry data real time and automatically completes data modification, filtering and geometrical parameter synthesis and displays waveform data and other items on a computer screen. Table 1 shows the track geometrical parameter specifications and accuracies of GJ-4 TGC. Data generated from every TGC are divided into three categories: waveform data, off-limits data and TQI data. Among them, the TQI data, generated over a 200 m interval, are data about overall track irregularity. Table 2 lists the fields of TQI data generated by Enhanced TGC IV.

Table 1. Track geometrical parameters measured by GJ-4 track inspection car and their accuracy.

Item for measurement	Measuring range	Accuracy of measurement
Distance	0–2999.9 km	≤±1‰
Speed	0–160 km/h	≤±0.2 km/h
Gauge	1415–1480 mm	≤±0.8 mm
Curvature	±23o/30 m	≤±0.05o/30 m
Level and off-limit	±200 mm	≤±1.0 mm
Surface	±60 mm	≤±1.0 mm
Alignment	±100 mm	≤±1.5 mm
Twist	±100 mm	≤±1.0 mm
Train body's vibrational acceleration	±9.8 m/s^2	≤±0.1 m/s^2
Axle box's vibrational acceleration	±980 m/s^2	±9.8m/s^2

Note: In China, a chord length of 20m is adopted for surface and alignment measurements.

Table 2. List of fields of tqi data generated by enhanced TGC IV.

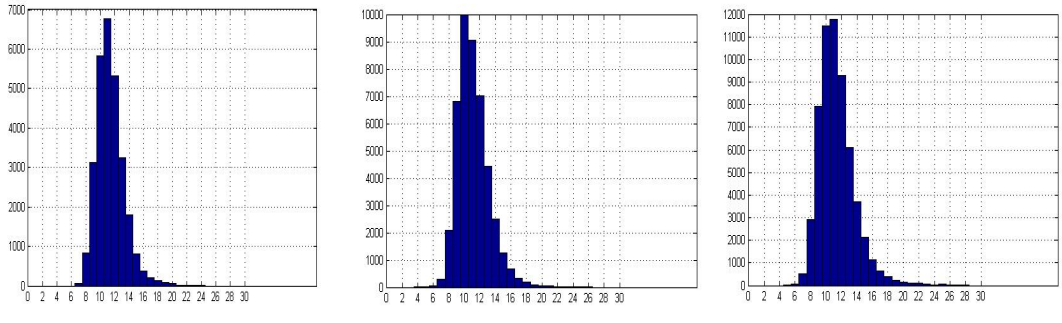
Field	LINENAME	DRECTIONS	KM	METERS
Field	RAL	LPR	RPR	GAGE
Field	LENGTH	SPEED	CURVE	LAL
Field	LEVEL	WARP	TQI	DATE

We have collected the TQI data from the Jing-Hu Line for 2004, 2005 and 2006, and will use them as the basic data for our prediction model.

T*rack Irregularity Changes And Model Building.* Track irregularity changes exhibit three major characteristics: cyclicity, multiage and exponentiality (Xu, 2009). Cyclicity means that for the same track section TQI changes have a similar pattern between two adjacent maintenances. Given this, we take the period between two adjacent maintenances has the time range for studying track state changes.

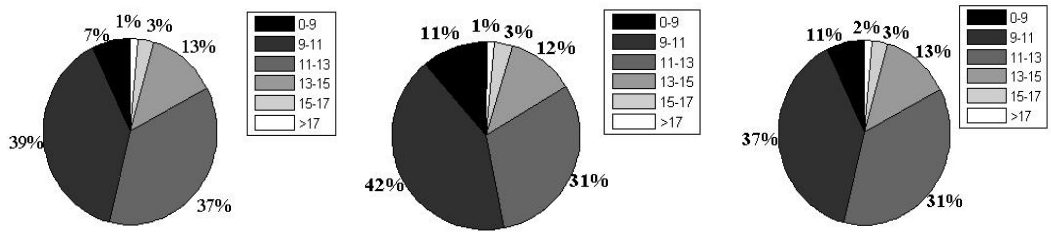
Multistage and exponential changes refer to that within one track state evolutionary cycle TQI changes at different rates in the early, middle and later stages. Given this, we divide TQI values of a selected unit track section into several stages of change. Different linear equations are used to fit the different stages within one track state change cycle. We then get broken lines of TQI changes and establish a multistage nonlinear model as an improvement over the single *k* linear model.

D*ivision of TQI Changes into Different Stages*. It is specified in Railway Line Maintenance Regulations that for trunk railway lines with a running speed of <160km/h, lines with a TQI value of <15.0 are considered as being in a balanced quality state, while lines with a TQI value of > 15.0 should be included into a line maintenance plan. For trunk railway lines with a running speed of 160 km/h, the TQI management value is 10.0. Based on the management values of TQI above, we have analyzed the distribution of TQI data for the Jing-Hu Line in 2004, 2005 and 2006 and found that 95% of all the TQI values fell in the 9.0–17.0 range (Figures 1 & 2). Based on this distribution pattern, we divide TQI values into 4 ranges, namely 9.0–11.0 (25%), 11.0–13.0 (41%), 13.0–15.0 (22%) and 15.0–17.0 (7%). Based on the TQI values identified by Railway Line Maintenance Regulations and the conclusions above, we divide the range of TQIs in trunk lines with a running speed of <160 km/h into 6 stages: stage I (TQI between 0 and 9.0), stage II (TQI between 9.0 and 11.0), stage III (TQI between 11.0 and 13.0), stage IV (TQI between 13.0 and 15.0), stage V (TQI between 15.0 and 17.0), stage VI (TQI above 17.0).



(a) (b) (c)

Figure 1. (a) Distribution of TQI values in 2004 (b) Distribution of TQI values in 2005 (c) Distribution of TQI values in 2006.



(a) (b) (c)

Figure 2. (a) Distribution of TQI values in 2004; (b) Distribution of TQI values in 2005; (c) Distribution of TQI values in 2006.

F*itting by Multistage Linear*. As mentioned above, track irregularity changes show a pattern similar to the Bathtub Curve. Based on the 6 stages, we use different linear equations to fit TQI changes in each stage. We then get a broken line which consists of several line segments which are used to approximate the trend of TQI changes over a unit track section. Let *A,* i.e. $\{TQI\ (t_n)|n=1, 2, 3\cdots\}$ be the time series of the TGC-generated TQI data of the unit track section. For the 6 stages, the TQI time series A is split into 6 stages, as shown in Eq. (2):

$$A=\{TQI(t_{n_0+1}),TQI(t_{n_0+2}),\ldots,TQI(t_{n_1}),TQI(t_{n_1+1}),TQI(t_{n_1+2}),\ldots,TQI(t_{n_2}),TQI(t_{n_2+1}),TQI(t_{n_2+2}),\ldots,TQI(t_{n_6})\} \quad (2)$$

where $(TQI(t_{n_0+1}),TQI(t_{n_0+2}),\ldots,TQI(t_{n_1}))\in[0,9.0)$, $(TQI(t_{n_1+1}),TQI(t_{n_1+2}),\ldots,TQI(t_{n_2}))\in[9.0,11.0)$,⋯, $(TQI(t_{n_2+1}),TQI(t_{n_2+2}),\ldots,TQI(t_{n_3}))\in[15.0,17.0)$, $(TQI(t_{n_5+1}),TQI(t_{n_5+2}),\ldots,TQI(t_{n_6}))\in[17.0,+\infty)$, $n_0=0$.

The different stages of TQI changes are fitted using Equation y_i:

$$y_i=y_{0i}+k_it \quad (3)$$

where i=1, 2, 3⋯ 6; k_i is the slope fitted for each stage, and y_{0i} is the intercept fitted for each stage.

The fitting coefficients are respectively

$$\left.\begin{aligned}\hat{y}_{0i}&=\frac{\sum_{k=n_{i-1}+1}^{n_i}TQI\ (t_k)\sum_{k=n_{i-1}+1}^{n_i}t_k^2-\sum_{k=n_{i-1}+1}^{n_i}t_k\sum_{k=n_{i-1}+1}^{n_i}t_kTQI\ (t_k)}{(n_i-n_{i-1})\sum_{k=n_{i-1}+1}^{n_i}t_k^2-(\sum_{k=t_{n_{i-1}}+1}^{n_i}t_k)^2}\\ \hat{k}_i&=\frac{(n_i-n_{i-1})\sum_{k=n_{i-1}+1}^{n_i}t_kTQI\ (t_k)-\sum_{k=n_{i-1}+1}^{n_i}t_k\sum_{k=n_{i-1}+1}^{n_i}TQI\ (t_k)}{(n_i-n_{i-1})\sum_{k=n_{i-1}+1}^{n_i}t_k^2-(\sum_{k=t_{n_{i-1}}+1}^{n_i}t_k)^2}\end{aligned}\right\} \quad (4)$$

Using Eq. (3), we establish a multistage linear prediction model MLPM. This model improves the linear model by changing a single k into several k values, i.e. k_i, to make linear fit with TQIs in different stages.

Modeling verification

We validate the MLPM model using TGC data from the Jing-Hu Line (over the K238+000–K725+800 section), and the results show that MLPM fits better with the TQI change process. It generates more accurate predictive results than the single *k* model. In this section, we take two track sections as examples to compare the predicted results of the MLPM model and the single k linear model.

Modeling Verification of the K253+800–K254+000 Section. The K253+800–K254+000 section of the Jing-Hu Line is selected for model validation. The MLPM model and the single k linear model are applied to the TQI data from a total of 32 TGC inspections in 2004, 2005 and 2006 to predict TQIs for the track section.

Predicted Results from MLPM. TQI changes in the unit track section are divided into 4 stages. The measured TQI data are used to establish a 4-stage linear prediction model for the track section. Linear fit is then made for each of the 4 stages to derive fitted equations, as shown in Table 3.

We use the model to predict TQIs for April 10, May 15, June 9 and July 11, 2006. These 4 dates were respectively 48, 83, 108 and 140 days from the last maintenance (on February 21, 2006). The equation for Stage 1 (TQI between 0 and 9.0) is chosen for the predictions. The predicted results are given in Table 4.

Predicted Results From the Single k Model. The fitting method described by professor XU is applied to the data collected from the track section to get a linear TQI equation:

$$y=0.0099t+8.2752 \tag{5}$$

Eq. (5) is used to predict TQIs for April 10, May 15, June 9 and July 11, 2006. The predicted results are given in Table 5.

Table 3. MLPM model for the Jing-Hu Line--K253+800–K254+000 section.

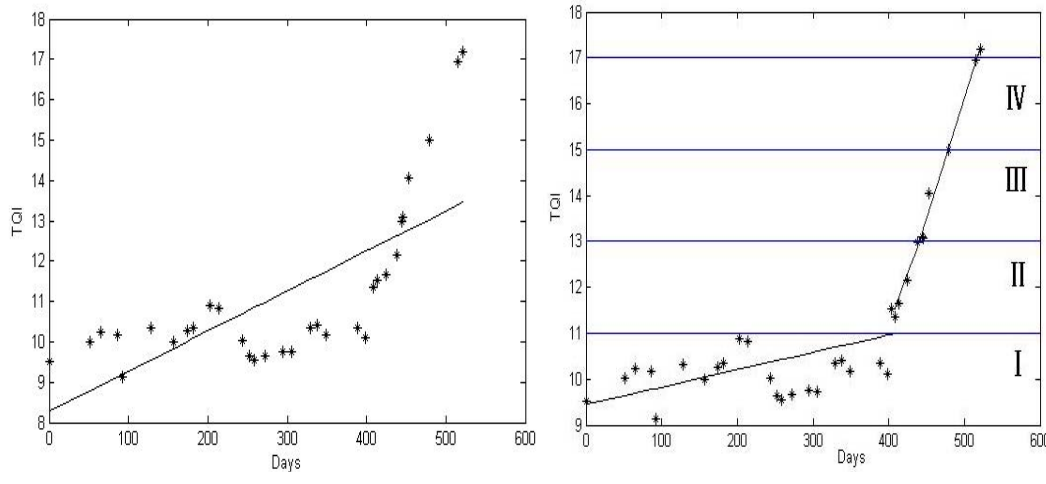
Stage	TQI range	Days from last maintenance	Fitted equation
1	[9.0,11.0)	0–399	$y = 0.0038t + 9.4492$
2	[11.0,13.0)	399–438	$y = 0.0446t - 6.6621$
3	[13.0,15.0)	438–479	$y = 0.054t - 10.0863$
4	[15.0,17.0)	>479	$y = 0.0572t - 10.2564$

Table 4. Predicted and measured TQI values of the Jing-Hu Line –K253+800–K254+000 section.

Date	Days	Predicted value	Measured value
2006-4-10	48	8.7504	10.49
2006-5-15	83	9.0969	10.05
2006-6-9	108	9.3444	10.08
2006-7-11	140	9.6612	10.19

Table 5. Predicted and measured TQI values of the Jing-Hu Line–K253+800~K254+000 section.

Date	Days	Predicted value	Measured value
2006-4-10	48	9.6316	10.49
2006-5-15	83	9.7646	10.05
2006-6-9	108	9.8596	10.08
2006-7-11	140	9.9812	10.19



(a) (b)

Figure 3. (a) Data Fit Diagram with the Single *k* Model for the Jing-Hu Line–*K*253+800–*K*254+000 Section; (b) Data FFit Diagram with MLPM for the Jing-Hu Line--*K*253+800–*K*254+000 Section.

Figure 3 shows the fitted data diagrams of the two models. The predicted values in Tables 4 and 5 are compared using the 2 indices of absolute error and relative error. The results are given in Table 6

Table 6. Comparison of predicted results by the two models

Days from last maintenance	Model	Absolute error	Relative error (%)
48	Single *k* model	–1.7396	16.58
	MLPM	–0.8584	8.18
83	Single *k* model	–0.9531	9.48
	MLPM	–0.2854	2.83
108	Single *k* model	–0.7356	7.30
	MLPM	–0.2204	2.19
140	Single *k* model	–0.5288	5.19
	MLPM	–0.2088	2.05

It can be seen from Table 6 that compared with the predictions by the single *k* model, MLPM predictions for the K253+800–K254+000 section for the 4 days can reduce the absolute errors by 0.8812, 0.6677, 0.5152 and 0.32, respectively and the relative error by 8.4%, 6.6%, 5.1%, and 3.1%, respectively.

Modeling Verification of K*550+800–*K*2551+000 *Section. An analysis of data from 22 inspections in 2004, 2005 and 2006 for the K550+800–K551+000 section shows that the range of TQI values for the section varied from 11.28 to 13.73, covering Stages III and IV. Measured inspection data are therefore used to establish a 2-stage linear prediction model for this unit section. Fitted equations are given in Table 7.

MLPM is then used to predict the TQI values for March 24, April 21, May 15 and June 9, 2006, which were respectively 31, 59, 83 and 108 days from the last maintenance (on Feb. 21, 2006). The equation for Stage I in Table 7 is therefore chosen to make TQI predictions. The predicted results are given in Table 8.

Single k Predicted Results. Again, the fitting method developed by professor Xu is applied to the data collected from the track section to get a linear TQI equation:

$$y=0.0045t+11.8459 \quad (6)$$

The predicted results for the 4 days are given in Table 9.

Table 7. MLPM model for the Jing-Hu Line--*K*550+800–*K*551+000 sections.

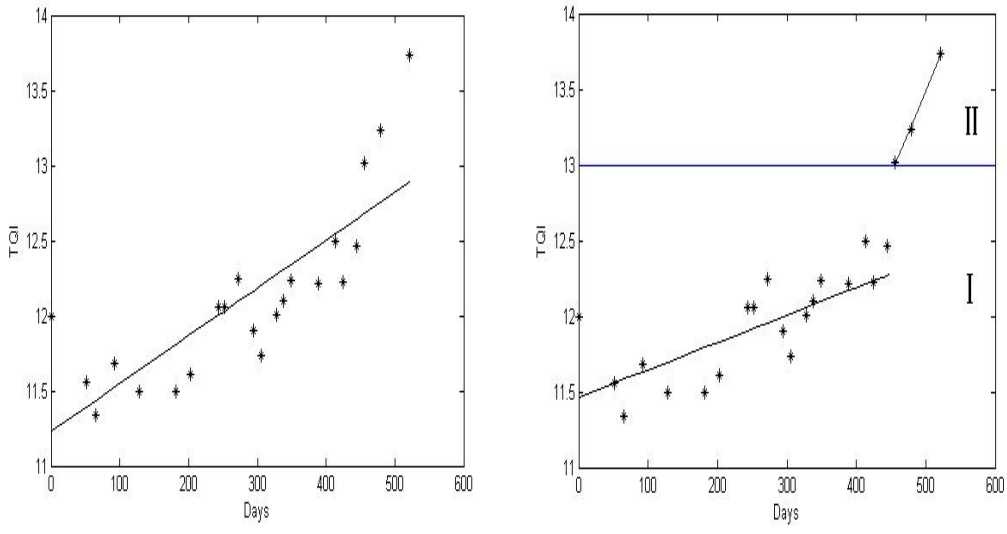
Stage	TQI Range	Days from last maintenance	Fitted equation
1	[11.0,13.0)	0–453	$y = 0.0018t + 11.4664$
2	[13.0,15.0)	>453	$y = 0.0110\ t + 7.9670$

Table 8. Predicted and measured TQI Values of the Jing-Hu Line–*K*550+800–*K*551+000 Section

Date	Days from last maintenance	Predicted TQI values	Measured TQI values
2006-3-24	31	11.5222	11.56
2006-4-21	59	11.5726	11.63
2006-5-15	83	11.6158	11.83
2006-6-9	108	11.6608	11.58

Table 9. Predicted and Measured TQI Values of the Jing-Hu Line–*K*550+800–*K*551+000 Section

Date	Days from last maintenance	Predicted TQI values	Measured TQI values
2006-3-24	31	11.9854	11.56
2006-4-21	59	11.1114	11.63
2006-5-15	83	12.2194	11.83
2006-6-9	108	12.3319	11.58



(a) (b)

Figure 4. (a) Data Fit Diagram with the Single *k* Model for the Jing-Hu Line--K550+800–K551+000 Section. (b) Data Fit Diagram with MLPM for the Jing-Hu Line--K550+800–K551+000 Section.

Table 10. Comparison of predicted results by the two models.

Days from last maintenance	Model	Absolute error	Relative error (%)
31	Single *k* model	0.4254	3.68
	MLPM	–0.0378	0.33
59	Single *k* model	–0.5186	4.45
	MLPM	–0.0574	0.49
83	Single *k* model	0.3894	3.29
	MLPM	–0.2142	1.81
108	Single *k* model	0.7519	6.49
	MLPM	0.0808	0.70

***Comparison*.** Fit diagrams for the two models are plotted in Figure 4. The predicted results by the two models (shown in Tables 8 and 9) are compared using the 2 indices of absolute error and relative error. The results are given in Table 10.

Table 10 shows that compared with the single *k* model LPM can respectively reduce the absolute errors by 0.3876, 0.4612, 0.1752 and 0.6711, respectively, and the relative errors by 3.35%, 3.97%, 1.48%, and 5.80%, respectively. Based on the comparative analysis in Sections 3.1 and 3.2, the multistage model MLPM fits more with the accrual TQI process than the single k model.

Conclusions

This paper, we have developed an improved multistage linear prediction model called MLPM on the basis of the single k linear model. Firstly, Mat Lab is used to

analyze the accumulated TQI data from the Jing-Hu Line. Based on the distribution of TQI values in the K238+000–K725+800 section, 6 stages of TQI changes in one track irregularity change cycle are divided for trunk railway lines with a running speed of <160 km/h. Secondly, based on the characteristics of TQI changes, we put forward MLPM. The model changes the single k into different k_i to predict TQI for different stages. Verification with measured data from the Jing-Hu Line shows that our MLPM fits better with the actual process of TQI changes than the single k model. It produces more accurate predicted results.

Acknowledgements

This Research was fully sponsored by the National Key Technology R&D Program (No. 2009BAG12A10) and the State Key Laboratory of Rail Traffic Control and Safety of Beijing Jiaotong University of China (Contract No. SKL2008K005).

References

Alfelor R.M., Carr G.A., and Fateh M. (2001). "Track Degradation Assessment Using Gage Restraint Measurements." *Transportation Research Record* 1742, 68-77.

Chen X. M., *et al.* (2006). "Integrating Factor Method for Predicting the Developing Trend of Railway Track Irregularity." *China Railway Science*, 27-31.

Hyslip, J. P. (2002). "Fractal Analysis of Geometry Data". *Transportation Research Record* 1785, 50-7.

Iyengar R.N., and Jaiswal O. R. (1995). "Random Field Modeling of Railway Track Irregularities." *Journal of Transportation Engineering*, 121(4), 303-8.

Kawaguchi, A., Miwa, M., and Terada K. (2005). "Actual Data Analysis of Alignment Irregularity Growth and its Prediction Model." *Quarterly Report of RTRI*, 46(4), 262-8.

Lu, Z. W. (2002). *Railway Track Structure and Maintenance*, China Railway Publishing House.

Ministry of Railway of the People's Republic of China. (2008). *Railway Line Maintenance Regulations*, China Railway Publishing House.

Xu, P. (2009). Research on the Variation Characteristics of the Track Dynamic Irregularity in Railway Line, Beijing Jiaotong University.

Xu, Y. D, Wu J. C. (2005). "Analysis on development of track irregularities with linear forecast model." *Journal of Shijiazhuang Railway Institute*, 6-9.

Relation between Railway Track Alignment Irregularity and Gross Passing Tonnage

Ran Guo[1], Baoming Han[3] and Futian Wang[3]

[1] Ph.D. Candidate, [2] Prof., [3] Associate Professor, Department of Traffic and Transportation, Beijing Jiaotong University, Beijing, China; E-mail:jack-guoran@163.com, bmhan@bjtu.edu.cn, wangfutian@jtys.bjtu.edu.cn

Abstract

The authors conducted research on the changing regularity of local alignment irregularities with the gross passing tonnage and built the deterioration model of local alignment irregularities. And after analyzing the data of the Chengdu-Kunming Railway (Cheng-Kun Railway) administrated by the Kunming Railway Bureau, this paper provided forecasts of the development of alignment irregularities of the railway. In the paper, shorter track segment is taken as the study object. During a maintenance period, the trend of the amplitude of local alignment irregularities is increasing the curve line with gross passing tonnage rising. Railway Engineering manages alignment irregularities through four management standards which base on peak values of alignment irregularities. On the basis of the management standard, this paper has researched the changing regularity of alignment irregularities with gross passing tonnage in unit section and built the prediction model of the amplitude of alignment irregularities in unit section. Making use of the actual data of rail lines in our country, the paper has forecasted the development of alignment irregularities of the rail line. Examining combined transportation and dispatch data of the Cheng-Kun Railway of the Kunming Railways Bureau from March 2008 to December 2008, this paper calculated the gross passing tonnage in days and collected the relevant data measured by a track geometry car. Using statistics and probability theory and function, this paper analyzed the relation between the amplitude of alignment irregularity of lots (3 m in length) and gross passing tonnage and built the local alignment irregularities deterioration model to forecast the development of alignment irregularities. After comparing predicting data with actual data and analyzing the error, the accuracy of the model was verified.

Key Words: railway; track alignment; local irregularity; gross passing tonnage; model

Introduction

Track alignment is a basic geometric parameter of rail. It shows the level position change of tracks on both left and right side. The deformation of track alignment could result in more transverse vibration of train as well as the increase of lateral force. Train derailment would happen due to unconstrained lateral force increase. To make sure that the train operates in smooth and safe conditions, it is important to check the track alignment periodically and adjust alignment irregularity timely to ensure the track works as it was intended. A large number of studies on the rail condition changes have been finished by both the researchers at domestic and abroad. However, very few literatures have dealt with the single index, i.e., the change rule of track alignment irregularity. Taking into account strength, distortion, and work conditions of the tracks, the Railway Technical Research Institute of Japan (RTRI) proposed a mathematical model with orbital mechanics method to forecast the track alignment irregularities of rail lines in Japan (Kawaguchi et al. 2005). The objective of this research is to predict track alignment irregularities of China's railway. In order to know the practical railway conditions of China, the observed data of the Cheng-Kun Railway (administered by the Kunming Railways Bureau of China) is used. Our work includes the following correlative parts:

(1) The change of track alignment irregularity is related to several factors, such as the rail condition and operating condition. This paper simplified the prediction model in "Actual Data Analysis of Alignment Irregularity Growth and Prediction Model" (Kawaguchi et al. 2005). The regularity of local track alignment irregularity changing with the variation of the gross passing tonnage in unit section, in a certain condition, was researched. And the peak values of track alignment irregularities can be predicted.

(2) The divided short sections are taken as study objects and different prediction models are built for different sections. The models can predict the track alignment irregularities between two maintenances.

(3) As gross passing tonnage increases, deterioration rate of track alignment irregularity is deferent in prophase, metaphase, and anaphase. The researchers built a multi-phases track alignment irregularity prediction model, through dividing the deterioration of the track alignment irregularity into several phases, combining the management standard in "Ministry of Railway of the People's Republic of China, Railway Line Maintenance Regulations" (China 2008).

Alignment Irregularity Prediction Model

Gross passing tonnage has close relation with the change of track alignment. In Figure 1, the deviation between track alignment irregularity and standard value increases and the trend of track alignment irregularity change is according with an increasing curve, when no maintenance and gross passing tonnage increases. The trend of track alignment irregularity change can be predicted, through monitoring the change of gross passing tonnage and combining inspection data.

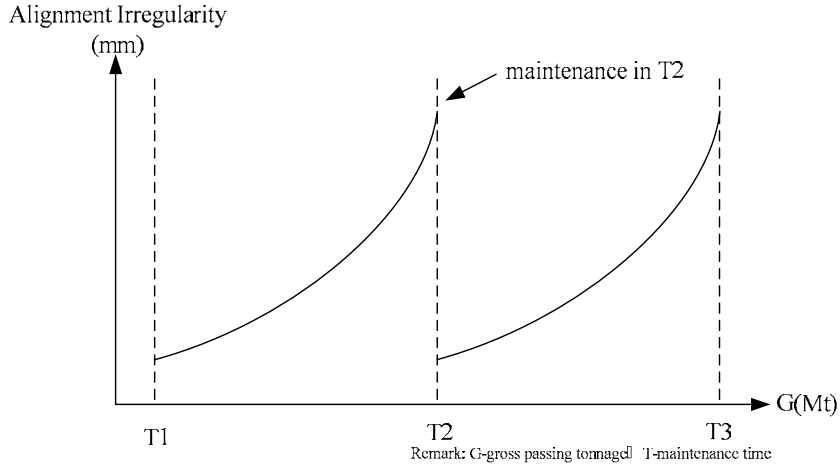


Figure 1. Condition of local track alignment irregularity changing with the variation of the gross passing tonnage.

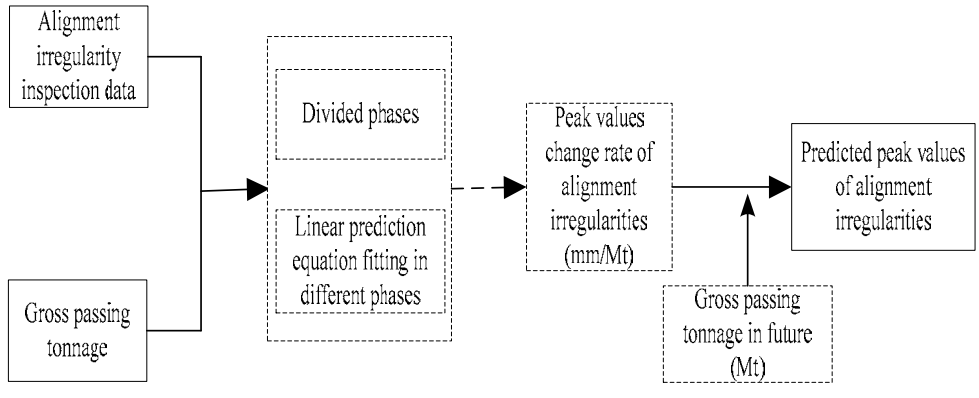


Figure 2. Alignment local irregularity prediction model.

As the features mentioned above, the paper built the following model, detailed in Figure 2. The model shows the change regularity of the peak values of the track alignment irregularities, within unit section, with gross passing tonnage varying. Gross passing tonnage and inspection data of track alignment irregularities are the input of the model. Between two maintenances, the development of track alignment irregularity can be divided into several phases. According to the inspection data, the linear relation between the peak values of the track alignment irregularities and gross passing tonnage, in respective phases, can be fitted by this model. Combining with the gross passing tonnage in future in a certain section, the peak values of the track alignment irregularities can be predicted by the model.

Data Source . The data used in the model is the waveform data of track alignment irregularities from The Cheng-Kun Railway, detected by a GJ-4 track geometry car. The Cheng-Kun Railway is inspected at least twice a month by a GJ-4 track geometry car TGC, which detects four times within a meter (corresponding to four

detecting points and four pieces of data). Table 1 shows the columns in database used to save the inspection data. The inspection data about track alignment of left and right rails is utilized in the model.

Gross passing tonnage is the other input of the model. Currently, gross passing tonnage statistics are annual data. Due to this long period of time for collecting data, there is a large error in the statistics of the gross passing tonnage, thus the statistical result cannot provide accurate data support on the prediction model. In the paper, to calculate gross passing tonnage precisely, gross passing tonnage is counted daily, extracted from the related transportation, and data dispatched from the Cheng-Kun Railway. Gross passing tonnage can be calculated in arbitrary sections during arbitrary periods of time, by daily counting of the gross passing tonnage of the Cheng-Kun Railway by each section. Table 2 shows the clipping of the daily counting results of gross passing tonnage of the Cheng-Kun Railway by each section. The detecting data of the Cheng-Kun Railway (K750+000~K1090+000) detected by track geometry cars and gross passing tonnage from March in 2008 to December in 2008 is selected by this paper.

Table 1. Fields of waveform data detected by a GJ-4 track geometry car.

Fields	Line name	Line id	Mileage	Gauge	Cross level
Fields	Left surface	Right surface	Left alignment	Right alignment	Track twist
Fields	Horizontal acceleration	Vertical acceleration	Speed	Inspection date	

Table 2. Gross passing tonnage of the Cheng-Kun Railway (part).

Number	Line Name	Mileage Scope	Gross Tonnage (t)	Date
1	Cheng-Kun Railway	758.80500-771.083	43243.1	20080301
2	Cheng-Kun Railway	771.08300-784.2855	40278.1	20080301
3	Cheng-Kun Railway	784.28550-798.28971	40278.1	20080301
4	Cheng-Kun Railway	798.28971-806.263	43243.1	20080301
5	Cheng-Kun Railway	806.26300-817.135	43243.1	20080301
6	Cheng-Kun Railway	817.13500-830.44222	43243.1	20080301
7	Cheng-Kun Railway	830.44222-840.65972	40278.1	20080301
8	Cheng-Kun Railway	840.65972-845.552	43243.1	20080301
9	Cheng-Kun Railway	845.55200-851.089	40278.1	20080301

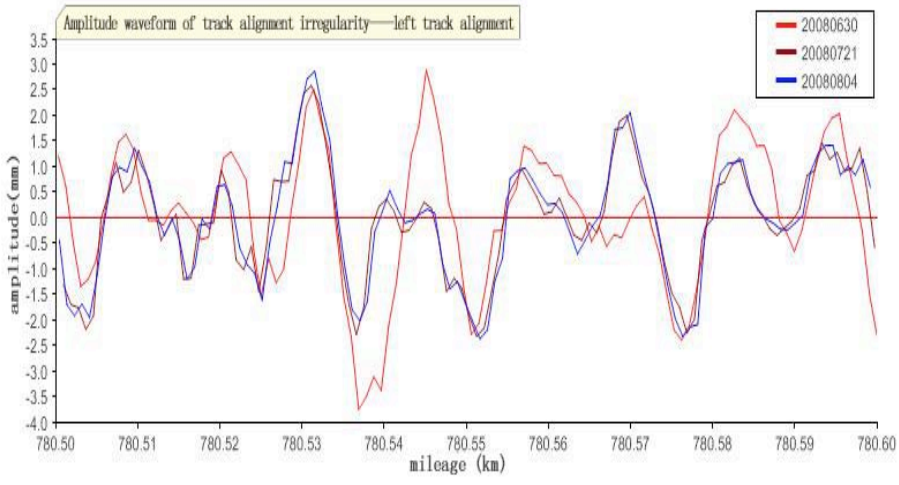


Figure 3. The Cheng-Kun Railway (K780+500~K780+600) alignment irregularity waveforms of amplitude deviation before and after maintenance.

Table 3. Track alignment irregularity dynamic management standards (unit: mm)(wavelength 1.5-42 m).

Railway Permitted Velocity	I Grade	II Grade	III Grade	IV Grade
$v \le 120km/h$	8	10	16	20
$120km/h < v \le 160km/h$	5	8	12	16
$160km/h < v < 200km/h$	5	7	10	12
$200km/h \le v \le 250km/h$	5	7	8	10

Deterioration Phases Division. The interval between maintenances is defined as a maintenance period. In an arbitrary unit section, as gross passing tonnage increases, track alignment irregularities will augment gradually. Adjustments are needed when the deterioration of the track alignment irregularities can cause safety problems. After maintenance, track alignment irregularities have been adjusted and come into a new maintenance period. Figure 3 shows the contrast of the track alignment irregularities fluctuations, to the left rail of the Cheng-Kun Railway (K750+000~K1090+000), detected three times. As it shows, after track alignment irregularities are maintained during 30 June 2008 and 21 July 2008, the peak values of track alignment irregularities have dramatically declined, and track alignment irregularities have come into a new maintenance period.

In each unit section, the period of track alignment irregularities starts at and ends with the time point of this maintenance and the time point of the next maintenance, respectively. The change regularity of track alignment irregularities with gross passing tonnage increases, between maintenances, has been researched.

During a period, the change rates of track alignment irregularity, with gross passing tonnage increase, are wildly different, in prophase, metaphase, and anaphase. In metaphase and anaphase, the change rates significantly accelerate. Based on the

nonlinear feature of change, model divides the development of alignment irregularity peak values into some phases according to the interval of management values divided based on the four grades of management standards of alignment irregularity peak values. Table 3 shows the management standards of track alignment irregularity peak values in China. I grade is maintenance standard, II grade is comfort standard, III grade is temporary repair standard, and IV grade is speed restriction standard. The defect points exceeded III grade must be dealt with immediately.

Taking into account management standards and maintenance requirements mentioned above, the development of the peak value of track alignment irregularity has been divided into the three parts, as shown in Figure 4.

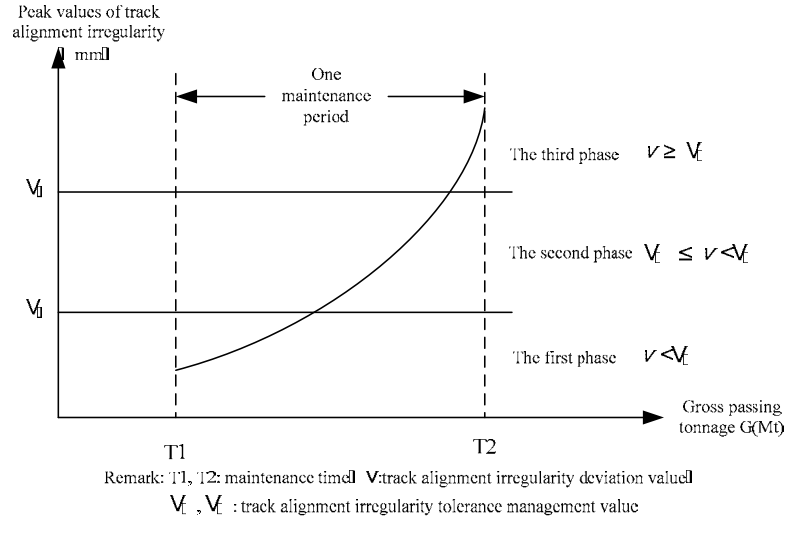


Figure 4. Phases' division of this model.

Model Building. $A(n)=\{(P_1,G_1),(P_2,G_2),...(P_i,G_i),...(P_n,G_n)\}$ is the sequence, consisting of the peak values of track alignment irregularities detected by track inspection car n times prior, and the corresponding gross passing tonnage respectively, in a unit section during the current maintenance period. P_i is the peak value of track alignment irregularity detected at the i th time in the unit section; G_i is the gross passing tonnage from the time of last maintenance to the time of detection at the i th time in the current maintenance period. According to the phases division mentioned above, $A(n)$ can be divided into three parts, as shown in Eq. 1.

$$A(n)=\{(P_{n_0+1},G_{n_0+1}),(P_{n_0+2},G_{n_0+2}),\ ,(P_{n_1},G_{n_1}),(P_{n_1+1},G_{n_1+1}),(P_{n_1+2},G_{n_1+2}),\ ,(P_{n_2},G_{n_2}),(P_{n_2+1},G_{n_2+1}),(P_{n_2+2},G_{n_2+2}),...\} \quad (1)$$

where $n_0=0$.

$$\begin{cases} P_{n_0+j}\in[0,V_I) \\ P_{n_1+j}\in[V_I,V_{II}) \\ P_{n_2+j}\in[V_{II},+\infty) \end{cases}$$

j=1,2,3,...; V_I, V_{II} are track alignment irregularity tolerance in sections.

The linear Eq. 2 can be used to fit track alignment irregularities in different development phases.

$$y_i = y_{0i} + k_i G \tag{2}$$

where, $i = 1, 2, 3$; y_{0i} is the fitting intercept in each phase; and k_i is the fitting slope in each phase.

The fitting coefficients are respectively

$$\hat{y}_{0i} = \frac{\sum_{k=n_{i-1}+1}^{n_i} P_k \sum_{k=n_{i-1}+1}^{n_i} G_k^2 - \sum_{k=n_{i-1}+1}^{n_i} P_k G_k \sum_{k=n_{i-1}+1}^{n_i} G_k}{(n_i - n_{i-1}) \sum_{k=n_{i-1}+1}^{n_i} G_k^2 - (\sum_{k=n_{i-1}+1}^{n_i} G_k)^2}$$

$$\hat{k}_i = \frac{(n_i - n_{i-1}) \sum_{k=n_{i-1}+1}^{n_i} P_k G_k - \sum_{k=n_{i-1}+1}^{n_i} P_k \sum_{k=n_{i-1}+1}^{n_i} G_k}{(n_i - n_{i-1}) \sum_{k=n_{i-1}+1}^{n_i} G_k^2 - (\sum_{k=n_{i-1}+1}^{n_i} G_k)^2} \tag{3}$$

Deterioration progression rates of the peak values of track alignment irregularities, in a certain unit section in a maintenance period, can be obtained by Eq. 3, and denoted by $\hat{k}_i$. Then the linear relation between the peak value of track alignment irregularity and gross passing tonnage in each phase can be obtained by Eq. 2.

The Instance of the Prediction Model

The length of unit section must be identified before the prediction of alignment irregularity using this model. In general, the out-of-gauge length of alignment irregularities are mostly between 1~6 m. The data of the out-of-gauge of track alignment irregularities of the Cheng-Kun Railway (K750+000~K1090+000) detected by a track geometry car in 2008, are counted and the results shown in Figure 5. The places whose out-of-gauge length is less than 6 m account for more than 90%.

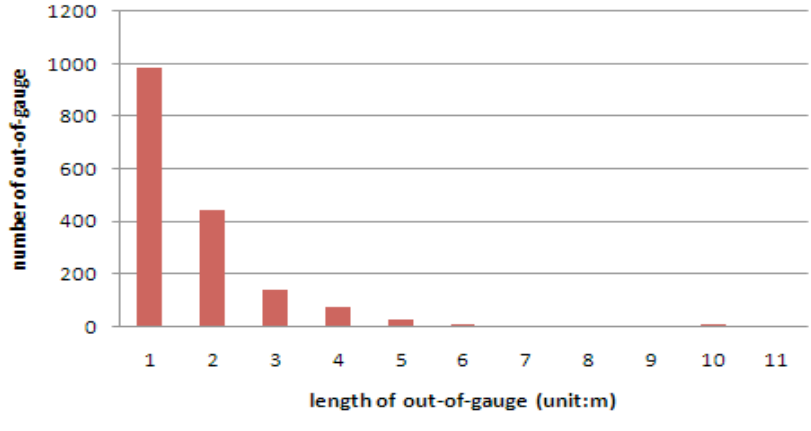


Figure 5. Data of alignment of irregularities out-of-gauge statistics of the Cheng-Kun Railway.

To verify the accuracy of the model, taking a 3 m section of the Cheng-Kun Railway K781+013~K781+016 section for example, the paper used the actual data of the section, from September 17, 2008, to November 10, 2008, to predict the peak values of track alignment irregularities on November 26, 2008, and compared the predicted data with the actual measured data. Table 4 shows the measured data of track alignment irregularities (left track alignment irregularities-LAI and right track alignment irregularities-RAI) detected 4 times, and the corresponding gross passing tonnage (GPT), during the period of maintenance.

According to the phases divided by the model, the development of track alignment irregularities is in the range of the first phase in the section. According to data in Table 4, Eq. 2 and Eq. 3 are solved, as shown in Table 5.

Table 4. The actual detection data of the Cheng-Kun Railway K781+013 ~ K781+016 section.

Date	Peak Value of LAI (mm)	Peak Value of RAI (mm)	GPT (Mt)
20080917	2.368	2.478	0
20080918	2.588	2.539	0.0775791
20081030	2.950	2.770	2.5965153
20081110	3.010	2.620	3.1122934

Table 5. The prediction model of the Cheng-Kun Railway K781+013 ~ K781+016 section.

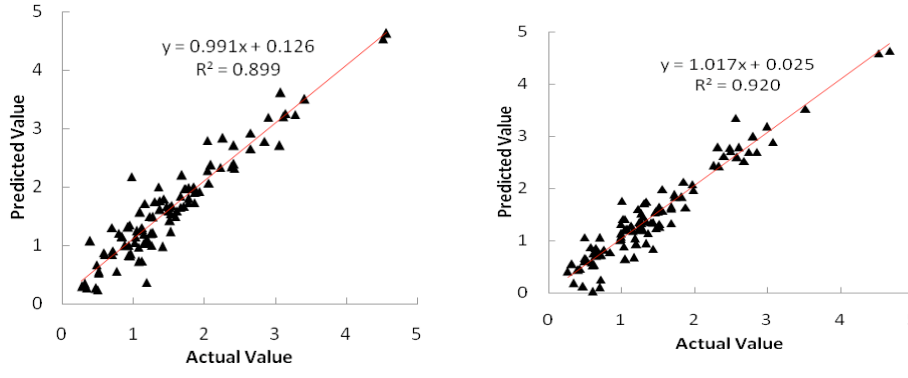
Phase	Item	Fitting Function
The first phase	LAI	$y = 2.403 + 2.023 \times 10^{-1} G$
	RAI	$y = 2.514 + 0.606 \times 10^{-1} G$

Table 6. The predicting outcomes of the Cheng-Kun Railway K781+013 ~ K781+016 section.

Date	Item	Predicted Value (mm)	Actual Value (mm)	Absolute Error	Relative Error
20081126	LAR	3.19	2.90	-0.290	10.00%
	RAI	2.75	2.85	0.1	3.51%

The gross passing tonnage is 3.930611Mt, from September 17, 2008, to November 26, 2008. The peak values of track alignment irregularities on November 26, 2008 are predicted by the model. And the predicting outcomes are shown in Table 6.

As shown in Table 6 the prediction accuracy of the model, for the track alignment irregularities, is ideal. In order to further verify the model, one hundred continuous 3 m sections of the Cheng-Kun Railway K781+000 ~ K781+300 section, were selected for prediction. And the contrast between the prediction data and the actual detection data is shown in Figure 6.



(a) (b)

Figure 6 (a). Left alignment irregularity prediction comparison diagram. (b). Right alignment irregularity prediction comparison diagram.

As shown in the figures above, on November 26, 2008, correlation coefficient R, between the predicted data and the actual measured data of peak values of the track alignment irregularities, on the left and right rail, respectively, is about 0.9. Thus, the prediction accuracy for the track alignment irregularities in sections is ideal.

Conclusions

Track alignment irregularity has crucial effects on train safety operation. As the high-speed and heavy loading railway development, research on the indexes of track irregularity respectively and prediction of track irregularities can be the way to ensure that the railway is smooth and safe. The paper proposed a phased prediction model for the peak value of track alignment irregularity in unit sections, utilizing the waveform data track inspection car detected and gross passing tonnage.

The main features of the model are in the following list:

(1) Within the same period of maintenance, track alignment irregularity change performs the multi-phase and exponential features, with gross passing tonnage increase, in a unit section. Taking the features into account, the phases of the track alignment irregularity deformation are defined, based on the management criterion.

(2) According to the degree of track alignment irregularity deformation in a unit section, with the inspection data, the multi-phase linear equations are fitted, to enhance the prediction accuracy.

The prediction model is implemented by computer programming language. The model is verified with the collected inspecting data of the Cheng-Kun Railway administered by the Kunming Railways Bureau of China. The result shows that the prediction accuracy of the model is relatively high.

References

Chen, X. M., Wang, L., Yang, F. C., Chai, X. S. and Wu, W. Q. (2006). "Integrating factor method for predicting the developing trend of railway track irregularity." *China Railway Science*, 27-31.

Kawaguchi, A., Miwa, M. and Terada, K. (2005). "Actual data analysis of alignment irregularity growth and prediction model." *QR of RTRI*, 262–268.

Luo, L., Zhang, G. M., Wu, W.Q. and Chai, W.S. (2006). "Track Regularity State Control in Wheel/rail System." Beijing: *China Railway Publishing House*.

"Ministry of Railway of the People's Republic of China, Existing Line Raising Speed 200~250km/h Railway Line and Bridge Maintenance Regulations." (2007). Beijing: *China Railway Publishing House*.

"Ministry of Railway of the People's Republic of China, Railway Line Maintenance Regulations." (2008). Beijing: *China Railway Publishing House*.

Sato, Y. (2001). "New track mechanics." *Beijing: China Railway Publishing House*.

Xu, P. (2009). "Research on the variation characteristics of the track dynamic irregularity in railway line." *Master Dissertation*. Beijing Jiao tong University.
Xu, Y. D., and Wu, J. C. (2005). "Analysis on development of track irregularities with linear forecast model." *Journal of Shijiazhuang Railway Institute*, 6-9.

A Short-Range Prediction Model for Track Irregularity of a Shorter Section of Track

Peng Xu[1], Quanxin Sun[2], Qing Li[3], Rengkui Liu[4] and Futian Wang[5]

[1] Ph.D. Candidate , State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University.No.3 of Shangyuan Residence, Haidian District, Beijing 100044, P.R. China. E-mail: suepen_aili@126.com
[2] Prof., [3] Junior Student, [4] Associate Professor, [5] Associate Professor,School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China

Abstract

Railway track is one of the main railway facilities and the basis for trains to operate on. In order to keep passenger and freight trains moving safely, stably and uninterruptedly, the railroads must ensure track roughness low. The launch of high-speed and high-haul trains requires lower roughness of track, and consequently expects the railroads to accurately maintain or repair railway track. In this paper, base on the characteristics of changes in track irregularity, we have developed a mathematical model called SRPM which uses track waveform data generated by track geometry car (TGC) to predict the track irregularity over a short track section with the length of 100m for each day in a future short-range period of time. For verify the effectiveness of SRPM, we applied the model to 25 sets of TGC-generated track waveform data from Beijing-Shanghai railway (Jing-Hu railway) administrated by Jinan Railway Bureau (JRB) to make short-range predictions for track irregularity over each unit section of the line segment. Finally, these SRPM predictions were analyzed in both spatial and temporal dimensions. From the analysis results, we come to the conclusion that SRPM developed in this paper can fairly accurately make short-range predictions for track irregularity over each unit track section of the JRB-administrated Jing-Hu railway.

Key Words: railway; track quality index; prediction; model

Introduction

Track is one of the main railway facilities and the basis for train operations. Safe, stable and uninterrupted train operations primarily depend on lower roughness of track. In China, high-speed and heavy-haul trains have become a development trend, setting higher requirements for track roughness. Consequently, the railway department needs to undertake high-precision railway maintenance. However, the current regimen and practices of railway maintenance in China can no longer meet the lower roughness requirements of the high-speed trains now operating in the country.

Many scholars in China and other countries have conducted research on the quantification and prediction for track irregularity. Kawaguchi et al. (2005) and Morimoto and Miwa (1997, cited in Kawaguchi et al. 2005) at Railway Technical Research Institute in Japan have respectively developed degradation models which utilize standard deviations of TGC-measured track geometry parameters to predict the standard deviations of track alignment and surface over 100m long track sections. Hyslip (2002a, 2002b) has applied fractal analysis to TGC-measured track geometry data and developed numerical indices based on the analysis for use in track condition assessment. Employing the stochastic process theory, Iyengar and Jaiswal (1995) have analyzed track surface data obtained from two railway lines in India and concluded that track surface can be modeled as a stationary Gaussian stochastic process. Alfelor et al. (2001) have fitted linear equations to the loaded and unloaded gauge data measured by Gage Restraint Measurement System (GRMS) to describe the relationship between gauge and each of the factors affecting it. Chen et al. (2006) at China Academy of Railway Sciences have developed an integrated factor method which utilizes TGC-generated track waveform data to predict track irregularities over a single mileage point for the day after the next TGC inspection.

Based on the characteristics of track irregularity changes, we develop a mathematical model called SRPM which uses TGC-generated track waveform data to predict track irregularity over a 100m long track section for each day in a future short-range period of time. In China, the TGC inspection cycles are normally less than 30 days. SRPM can predict track irregularity over unit track sections with the length of 100m for each day within the next TGC inspection cycle. Before verifying the effectiveness of SRPM, we have developed a program for it. We then apply SRPM to the 25 sets of TGC-generated track waveform data from the JRB-administrated Jing-Hu railway (up going) to make short-range predictions for track irregularity over each unit section of the railway segment. Finally, these SRPM predictions are analyzed in both spatial and temporal dimensions to verify the effectiveness of the model.

SRPM

In China, there are two unit lengths of track section, i.e. 200m and 500m, to serve the purpose of mean value management for track irregularities (Ministry of Rail 2006, 2007). To achieve more refined track irregularity management, we borrow the method of track section division and representation in track irregularities mean value management and divide railway tracks into 100m long unit sections along the increasing mileage direction, with the 100m, 500m or 1,000m post as the starting point. The start mileage of a unit track section is used to represent the section. In accordance with the characteristics of the quantification for track irregularity over a unit track section on the basis of standard deviations (Wei 1996), the track irregularity over a 100m long unit section is also quantified by the sum of standard deviations of seven geometry parameters.

Various factors affect track states. These include railway track maintenance, track regimen, transportation and environmental conditions, all with varying degrees of influence, resulting in different process of track irregularity changes over different track sections. To precisely approximate the real irregularity change process, different models should be used for different track sections. In this section, we will show how to build a short-range prediction model for track irregularity over a single 100m unit track section on the basis of the characteristics of track irregularity changes over the unit section.

Steps to Establish SRPM.

Variable nomenclature. For a 100m track section *Sec*, we are concerned with the process of track irregularity over the unit time spell of day (d). Let $\{t_{n,j}\}$ be the time series formed by the dates of each day within the time range under research. When j equals to 0, $\{t_{n,0}\}$ denotes the date of the TGC inspection on the railway line where the track section Sec is located, abbreviated as t_n. When j is greater than 0, $t_{n,j}$ denotes the date of the day after the date of the inspection on the track section. When j is less than 0, $t_{n,j}$ denote the date of the day before the date of the inspection on the track section. Let $S_{n,j}$ be the value of track irregularity over the track section Sec on date $t_{n,j}$.

Process of daily changes in track irregularity over *Sec.* Under the joint effect of various factors, the track irregularity of the unit track section *Sec* changes daily. Based on the transitivity and memory of track irregularity changes (Chen et al. 2006; Xu 2009), $S_{n,j-1}$, the value of track irregularity at $t_{n,j-1}$ is a main input variable which affects the value of $S_{n,j}$ at $t_{n,j}$ Meanwhile, $S_{n,j}$ is also a main output variable at $t_{n,j}$, which affects $S_{n,j+1}$ at $t_{n,j+1}$, as shown in Figure 1.

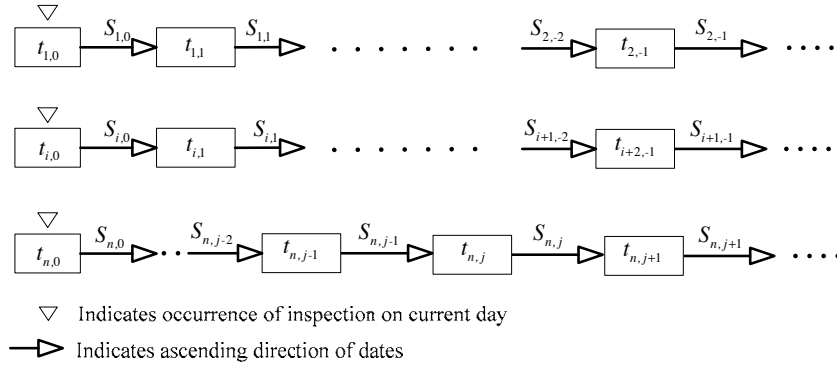


Figure 1. The process of daily changes in track irregularity over the track section *Sec*.

In the process of track irregularity shown in the Figure 1, the railway where the track section *Sec* is located is inspected by TGC at the frequency specified by Ministry of Railways of the People's Republic of China (2006, 2007). $S_{n,0}$, value of track irregularity on the inspection date of $t_{n,0}$ can be calculated from TGC-generated track waveform data according to the formula given by the Ministry of Railways (2006). $S_{n,0}$ is therefore also called the measured irregularity. For $S_{n,j}$, the track irregularity at $t_{n,j}$ between two successive track inspections (wherein), we will develop a mathematical model to make predictions on the basis of the characteristics of track irregularity changes over *Sec*. $S_{n,j}$ is thus also called the predicted irregularity.

Steps to establish SRPM. Affected by various factors such as track condition, transportation and environmental conditions, and track maintenance work, there are cycles of track irregularity change over each unit section, which are separated by the

date of track maintenance (Xu 2009). Practice has proved that within one change cycle track irregularity changes over a unit track section are nonlinear. Because the kinetic energy (determined by the train speed and mass, two key factors affecting track irregularity) acted on a unit track section is small over a short period of time, track irregularity changes over the track section in the absence of maintenance work is also small within such a short period. Based on the above understanding and in accordance with the differential thinking, we divide the process of nonlinear irregularity changes over a unit track section within one change cycle into many short-range change processes, each of which can be approximated by a linear regression equation F (i.e. a segment of a straight line). When a new TGC inspection for the unit track section is performed, we will adjust F based on the characteristics of track irregularity changes as revealed by the newly measured data. With continuous TGC inspections, we can obtain a family of linear regression equations. According to the integral thinking, the nonlinear process of track irregularity over the section within the whole cycle can be approximated by this family of linear regression equations. This family of equations thus forms SRPM, the short-range prediction model for track irregularity over the unit track section.

For the sake of convenience, we assume that the TGC inspection cycle is l (d) days for the railway track where Sec is located. We also assume that there is no maintenance work on Sec during the period from $t_{m,0}$ to $t_{n,l}$ ($m < n$ and generally $n - m = 3$ 或 4) and that the current date is $t_{m,0}$. For the unit track section Sec shown in Figure. 1, we derive a fitted linear regression equation F from the measured irregularity series $\{S_{r,0}, m \le r \le n\}$ of the short period from $t_{m,0}$ to $t_{n,l}$ and use it to approximate the change process of track irregularity over the same period. Then $S_{n,j}$ $(0 < j \le l)$, the track irregularity on each day from $t_{n,1}$ to $t_{n,l}$ will be predicted by using the derived linear regression equation F. The prediction of $S_{n,j}$ is denoted by $\overline{S}_{n,j}$, as shown in Figure. 2.

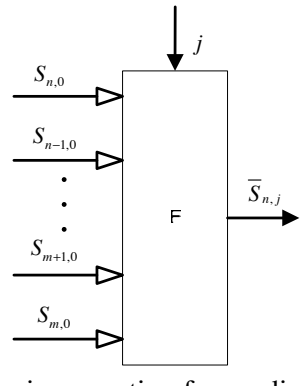


Figure 2. The linear regression equation for predicting the track irregularity on each of the l adjacent days after t_n

When the current date is $t_{n+1,0}$ the $(n+1)^{th}$ TGC inspection has been performed on the railway line where the unit track section Sec is located. To enable SRPM to accurately approximate the real process of nonlinear track irregularity changes, we use the latest measured irregularity $S_{n+1,0}$ to replace the "outdated" measured irregularity $S_{m,0}$, refit the linear regression equation F to the newly-measured irregularity set $\{S_{r,0}, m+1 \le r \le n+1\}$, and then use the refitted F to predict S_{n+1} for each day in the range of $t_{n+1,1}$ and $t_{n+1,l}$. Repeating the above procedure will arrive at an optimized SRPM on a rolling basis.

Establishment of SRPM. As illustrated in the Figure. 2, after the n^{th} TGC inspection on Sec, we use Eq. (1) and the first $(n-m+1)$ adjacent measured irregularities, i.e. $\{S_{r,0}, m \le r \le n\}$, to fit the linear regression equation F and predict the irregularity $S_{n,j}$ for each day from $t_{n,1}$ to $t_{n,l}$

$$\overline{S}_{n,j} = k_n \times j + b_n \tag{1}$$

Wherein: k_n and b_n are estimated by the least squares method using Eq. (2).

$$\left.\begin{aligned} \overline{k}_n &= \frac{-\sum_{i=m}^{n}(d_{i,n} \times S_{t,0}) + \frac{1}{n-m+1}(\sum_{i=m}^{n} d_{i,n}) \times (\sum_{i=m}^{n} S_{t,0})}{\sum_{i=m}^{n}(d_{i,n})^2 - \frac{1}{n-m+1}(\sum_{i=m}^{n} d_{i,n})^2} \\ \overline{b}_n &= \frac{1}{n-m+1}(\sum_{i=m}^{n} S_{t,0}) - \frac{1}{n-m+1}\overline{k}_n \times (\sum_{i=m}^{n} d_{i,n}) \end{aligned}\right\} \tag{2}$$

Wherein: $\overline{k}_n$ and $\overline{b}_n$ are the estimates for k_n and b_n, respectively, $d_{i,n}$ = the number of days from the i^{th} TGC inspection to the n^{th} TGC inspection, $m \le i \le n$.

After the $(n+1)^{th}$ TGC inspection, as described in Section 2.1, the linear equation

F will be refitted using the newly-measured irregularity $S_{n+1,0}$. The refitted F is shown in Eq. (3).

$$\overline{S}_{n+1,j} = k_{n+1} \times j + b_{n+1} \tag{3}$$

Wherein: k_{n+1} and b_{n+1} are estimated according to Eq. (2) using $\{S_{r,0}, m+1 \le r \le n+1\}$.

.

As TGC inspections continue for the railway, the above refitted procedures will produce a family of linear regression equations. This family of equations forms the SRPM for making short-range predictions for track irregularity over the unit track section Sec.

SRPM Application Analysis

We have developed a computer program for SRPM. And then we used the program and GJ-4 TGC-generated track waveform data from the up-going track of the Jing-Hu railway from K714+000 to K724+000 to respectively establish SRPMs for a total of 100 unit track sections of this track segment. Before the execution of the program, we set $n-m$ to 5. It means that track irregularity over a unit track section for each day in the next l days is predicted by a linear regression equation F fitted to the first 6 adjacent measured irregularities. In this section, the SRPM predictions will be analyzed in both spatial and temporal dimensions.

The collected GJ-4 TGC track waveform data. The Jing-Hu railway is a double track. Its segment from K238+000 to K727+000 is administrated by JRB. According to the specifications (Ministry of Railways 2006, 2007), the Jing-Hu railway is inspected twice a month by GJ-4 TGC. GJ-4-type TGC is a strapdown measuring system developed by the China Academy of Sciences. First put into use in 1999, GJ-4-type TGC samples and processes track geometry data real time and automatically, completes data modification, filtering and geometry parameter synthesis and displays waveform data and other items on a computer screen. Table 1 shows the specifications and accuracies of track geometry parameter measurements by GJ-4-type TGC. During the inspection, GJ-4 TGC stores the mileage and the measured parameters data of each sampling point with the suffix as “ste”. In the research and development process of the project, “Railway Engineering and Safety Production Management Information System”, we have collected from JRB 25 sets of GJ-4 TGC track waveform data about the Jing-Hu railway (up going) for the period from September 2004 to December 2005, as shown in Table 2.

Table 1. Track geometry parameters measured by GJ-4-type track inspection car and their accuracy range

Item for measurement	Measuring range	Accuracy of measurement
Distance	0~2999.9Km	≤±1‰
Speed	0~160Km/h	≤±0.2Km/h
Gauge	1415~1480mm	≤±0.8mm
Curvature	±23o/30m	≤±0.05o/30m
Level and off-limit	±200mm	≤±1.0mm
Surface	±60mm	≤±1.0mm
Alignment	±100mm	≤±1.5mm
Twist	±100mm	≤±1.0mm
Train body's vibrational acceleration	±9.8m/s^2	≤±0.1 m/s^2
Axle box's vibrational acceleration	±980m/s^2	±9.8m/s^2

Note: In China, a chord length of 10m is adopted for surface and alignment measurements.

Table 2. The collected GJ-4 TGC track waveform data from the Jing-Hu railway (up going)

Month of TGC inspection	Number of track waveform files	Month of TGC inspection	Number of track waveform files	Month of TGC inspection	Number of track waveform files
Sep. 2004	1	Mar. 2005	2	Sep. 2005	1
Oct. 2004	1	Apr. 2005	2	Oct. 2005	2
Nov. 2004	1	May 2005	2	Nov. 2005	2
Dec. 2004	2	Jun. 2005	2	Dec. 2005	2
Jan. 2005	0	Jul. 2005	2	-	-
Feb. 2005	1	Aug. 2005	2	-	-

Analysis of the SRPM predictions in the temporal dimension. Accordance to the computer output, we have conducted error analysis of the SRPM-predicted values of track irregularity over the 100 unit track sections in the Jing-Hu railway (up going) from K714+000 to K724+000. Table 3 lists the absolute mean $\overline{e^*}$ of the absolute errors and the absolute mean $\bar{e}$ of the relative errors of the 19 SRPM predictions for each unit section in the railway segment.

Table 3. The error analysis results for each of unit track sections in the Jing-Hu railway (up going) from K714+000 to K724+000

Unit Track Section	$\overline{e^*}$ mm	$\bar{e}$ (%)	Unit Track Section	$\overline{e^*}$ mm	$\bar{e}$ (%)	Unit Track Section	$\overline{e^*}$ mm	$\bar{e}$ (%)
K714+000	0.620	7.350	K717+400	0.388	4.690	K720+800	0.320	2.512
K714+100	0.418	4.792	K717+500	0.197	2.540	K720+900	0.454	5.107
K714+200	0.437	4.758	K717+600	0.325	3.724	K721+000	0.737	7.566
K714+300	0.409	4.673	K717+700	0.693	8.012	K721+100	0.554	5.792
K714+400	0.381	4.642	K717+800	0.569	5.884	K721+200	0.566	5.926
K714+500	0.572	5.179	K717+900	0.311	3.375	K721+300	0.770	8.022
K714+600	0.266	3.062	K718+000	0.531	5.937	K721+400	0.672	6.649
K714+700	0.359	3.880	K718+100	0.495	5.556	K721+500	0.614	6.273
K714+800	0.355	2.985	K718+200	0.651	6.673	K721+600	0.551	5.441
K714+900	0.479	4.438	K718+300	0.457	4.945	K721+700	0.540	6.911
K715+000	0.450	3.920	K718+400	0.308	3.134	K721+800	0.539	6.325
K715+100	0.582	5.581	K718+500	0.751	7.573	K721+900	0.380	4.623
K715+200	0.317	3.128	K718+600	0.349	4.118	K722+000	0.468	5.192
K715+300	0.297	3.308	K718+700	0.533	5.275	K722+100	0.362	3.750
K715+400	0.284	3.722	K718+800	0.439	5.015	K722+200	0.446	5.069
K715+500	0.345	3.688	K718+900	0.241	2.997	K722+300	0.565	6.234
K715+600	0.346	3.613	K719+000	0.453	4.692	K722+400	0.609	6.415
K715+700	1.088	9.891	K719+100	0.331	3.590	K722+500	0.404	4.640
K715+800	0.322	3.450	K719+200	0.430	3.960	K722+600	0.698	7.774
K715+900	0.857	8.487	K719+300	0.422	3.097	K722+700	0.692	7.821
K716+000	0.633	6.404	K719+400	0.748	5.199	K722+800	0.590	6.368
K716+100	0.316	3.678	K719+500	0.652	4.851	K722+900	0.706	6.233
K716+200	0.492	5.040	K719+600	0.444	4.002	K723+000	0.642	6.686
K716+300	0.814	8.342	K719+700	0.723	6.350	K723+100	0.382	4.472
K716+400	0.566	5.696	K719+800	0.491	5.155	K723+200	0.375	4.344
K716+500	0.895	8.962	K719+900	0.261	3.045	K723+300	0.316	3.321
K716+600	0.896	8.456	K720+000	0.251	2.910	K723+400	0.621	6.523
K716+700	0.245	2.804	K720+100	0.254	2.765	K723+500	0.351	4.001
K716+800	0.458	4.398	K720+200	0.298	2.876	K723+600	0.367	3.632
K716+900	0.436	4.949	K720+300	0.410	3.738	K723+700	0.647	6.625
K717+000	0.550	5.670	K720+400	0.581	5.403	K723+800	0.391	4.284
K717+100	0.500	4.961	K720+500	0.885	8.110	K723+900	0.607	6.249
K717+200	0.356	4.168	K720+600	0.674	5.465	-	-	-
K717+300	0.411	4.256	K720+700	0.549	4.189	-	-	-

Table 3 shows that the maximum value of $\overline{e^*}$ is 1.088mm, the minimum is 0.197mm and the average is 0.498mm. The maximum value of $\bar{e}$ is 9.891%, the minimum is 2.512% and the average is 5.120%. From the results of analysis in the temporal dimension, we come to the conclusion that SRPM is different for different unit sections and that SRPM can find satisfactory fit coefficients to accurately approximate the process of nonlinear track irregularity changes over each unit section. It is also concluded that SRPM can make fairly accurate predictions for most unit track sections along the Jing-Hu railway (up going).

Analysis of SRPM predictions in the spatial dimension. In this section, we will analyze in the spatial dimension errors in the SRPM-predicted values of irregularity for all the unit sections in the Jing-Hu railway (up going) from K714+000 to K724+000 and assess validity of SRPM in the spatial dimension.

We rank the 100 unit track sections along the 10km railway segment in an ascending mileage order. Let $S_{n,j,p}$ and $\overline{S}_{n,j,p}$ (wherein, $1 \le p \le 100$) respectively denote the TGC-measured value of irregularity and the SRPM-predicted value of irregularity over the p^{th} unit section on the day of $t_{n,j}$. When j equals to 0, $S_{n,0,p}$ is given the measured value of irregularity $S_{n,0}$ over the unit section at $t_{n,0}$. It is therefore also called the measured value of irregularity.

We use the SRPM-predicted values of irregularity and the TGC-measured values of irregularity along this railway segment at $t_{n,0}$ to form a time series $\{\overline{S}_{n,0,p}\}$ of the predicted values of irregularity and a time series $\{S_{n,0,p}\}$ of the measured values of irregularity. Then, using these two time series, we calculate the absolute mean value $\overline{e^*}$ of the absolute errors and the absolute mean value $\bar{e}$ of the relative errors of each prediction for the railway segment (shown in Table 4). And also a point set $\{S_{n,0,p}, \overline{S}_{n,0,p}\}$ which is composed of the TGC-measured and SRPM-predicted values of irregularity, is formed by these two series. Figure. 3 is the scatter plot of the point set.

Figure. 4 shows the superposed waveforms of the TGC-measured irregularity value series $\{S_{n,0,p}\}$ (wherein, $17 \le n \le 25$) and the corresponding SRPM-predicted irregularity value series $\{\overline{S}_{n,0,p}\}$ for the railway segment from August 12, 2005 to December 15, 2005. Figure. 5 (a)-(i) are the superposed waveforms on August 12, 2005, August 22, 2005, September 2, 2005, October 12, 2005, October 22, 2005, November 6, 2005, November 17, 2005, December 6, 2005, and December 15, 2005, respectively.

Table 4. Results of error analysis of each prediction for 100 unit sections on the Jing-Hu railway (up going) from K714+000 to K724+000

t_n	$\overline{e^*}$	$\overline{e}$	t_n	$\overline{e^*}$	$\overline{e}$	t_n	$\overline{e^*}$	$\overline{e}$
	mm	(%)		mm	(%)		mm	(%)
20050311	0.515	5.608	20050617	0.337	3.532	20051022	0.446	4.599
20050319	0.469	4.700	20050709	0.700	6.853	20051106	0.401	4.120
20050409	0.585	5.827	20050721	0.611	6.301	20051117	0.670	6.793
20050420	0.566	5.637	20050812	0.567	5.819	20051206	0.571	6.147
20050519	0.529	5.658	20050822	0.399	4.112	20051215	0.412	4.513
20050528	0.376	3.925	20050902	0.350	3.567	-	-	-
20050604	0.281	2.932	20051012	0.672	6.629	-	-	-

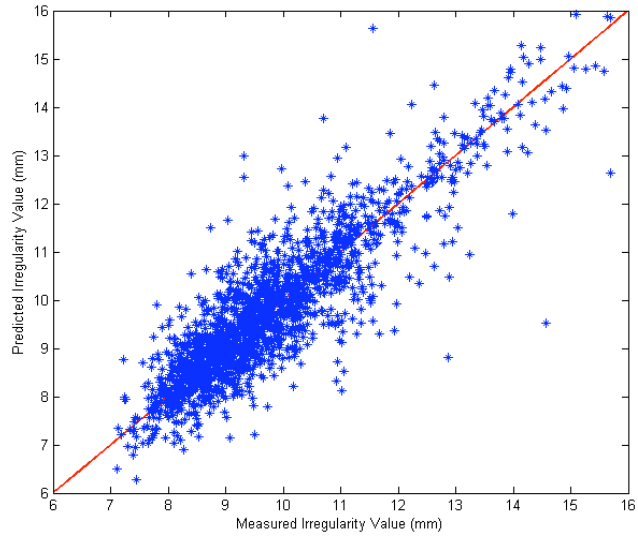
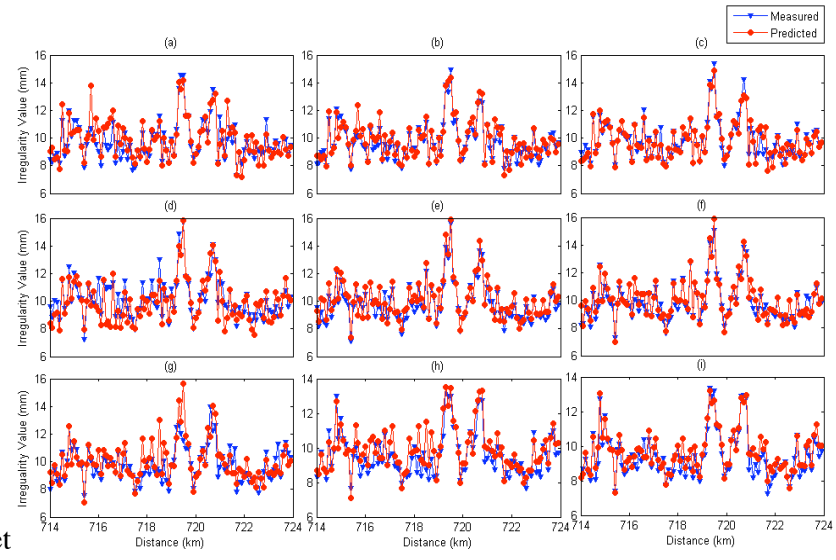


Figure 3. The scatter plot of the point $\{S_{n,0,p}\ \overline{S}_{n,0,p}\}$



set

Figure 4. The superposed waveforms of the measured and predicted irregularity value time series.

Figure 3 shows that most of the TGC-measured and SRPM-predicted irregularity value points are close to the straight line $y = x$. It is clear from Fig. 4 that the waveforms of SRPM-predicted irregularity values are similar to those of the TGC-measured irregularity values. Table 4 shows that the maximum value of $\overline{e^*}$ for the 19 predictions over the 10km railway segment is 0.700mm, the minimum is 0.281mm and the average is 0.498mm and that the maximum value of $\bar{e}$ is 6.853%, the minimum is 2.932% and the average is 5.120%. From the above error analysis results, we are arriving at a conclusion that in most cases the SRPM-predicted irregularity values are close to the TGC-measured irregularity values.

Acknowledgements

Our Research Project was also sponsored by National Key Technology R&D Program with the grant number of 2009BAG12A10.

Conclusions

In this paper, we arrive at the following two main conclusions on track irregularity prediction:

(1) On the basis of the characteristics of track irregularity changes, a mathematical model (SRPM) which adopts the calculus thinking and uses TGC track waveform data has been developed in this paper to predict irregularity value over a 100m long unit track section for each day in a future short-range period of time. The core idea of SRPM is as follows: 1) use the first $(n-m+1)$ (wherein, $n>m$ and generally $n-m=4,5,6$) adjacent measured irregularity values over a unit track section Sec to fit a linear regression equation F to predict values of irregularity for each day within the next TGC inspection cycle; 2) after a new TGC inspection, SRPM will use the newly measured irregularity value to replace the "outdated" measured irregularity value and adjust F to make better irregularity value predictions for each day in the next inspection cycle; and 3) SRPM can repeat the replacement to refit F based on the process of track irregularity changes over the unit section and use the latest refitted F to predict irregularity value for each day within the next inspection cycle.

(2) Using the 25 sets of TGC track waveform data from the JRB-administrated Jing-Hu railway (up going) from September 2004 to December 2005, we have established SRPM ($n-m=5$) for each unit track section along the railway segment. According to the TGC-measured irregularity values about the Jing-Hu railway (up going) from K714+000 to K724+000, the SRPM-predicted irregularity values were analyzed in both spatial and temporal dimensions. From the analysis results, we come to the following main conclusions that: 1) Different unit track sections have different SRPMs; 2) SRPM can fairly accurately approximate the whole process of track irregularity changes over a unit track section; and 3) SRPM can fairly accurately predict irregularity values for most unit track sections in the JRB-administrated Jing-Hu railway (up going).

In future research, we will collect more TGC track waveform data to verify the adaptability of SRPM for other railway in China. We will also conduct research on planning track maintenance according to SRPM predictions under the current railway maintenance regimen. In addition, we plan to turn the current research findings into a functional module of the "Railway Engineering and Safety Production Management Information System" R&D project, and put it into operation in the Jinan and Kunming railway bureaus.

References

Alfelor R.M., Carr G.A., and Fateh M., 2001. Track Degradation Assessment Using Gage Restraint Measurements. *Transportation Research Record*, No.01-2496, 68-77

Chen, X.M., Wang, L., Yang, F.C., Cai, X. S. and Wu, W. Q., 2006. Integrating Factor Method for Predicting the Developing Trend of Railway Track Irregularity. *China Railway Science*, 27 (6), 27-30.

Hyslip, J.P., 2002a. *Fractal Analysis of Geometry Data for Railway Track Condition Assessment*, Ph. D. thesis. University of Massachusetts Amherst, Amherst, MA 01003, USA.

Hyslip, J.P., 2002b. Fractal Analysis of Geometry Data. *Transportation Research Record*, No.02-2771, pp.50-57

Iyengar R.N., Jaiswal O.R., 1995. Random Field Modeling of Railway Track Irregularities. *Journal of Transportation Engineering*, 121(4), 303-308.

Kawaguchi, A., Miwa, M. and Terada K., 2005. Actual Data Analysis of Alignment Irregularity Growth and its Prediction Model. *Quarterly Report of RTRI*, 46 (4), pp.262-268.

Ministry of Railways of the People's Republic of China, 2006. *Railway Line Maintenance Regulations*. P.R. China, Beijing: China Railway Press.

Ministry of Railways of the People's Republic of China, 2007. *Regulations on Existing Speed-raised Railway Line and Bridge Facilities with the Allowable Speed between 200 and 250 Km/h. P.R. China*, Beijing: China Railway Press.

Wei S.B., Yang F.C., and Weng S.D., 1996. *Track Quality Index. China Railway Science*, 17(2), 23-33.

Xu P., 2009. *Research on the Characteristics of the Track Dynamic Irregularity in Railway Line,* M. D. P.R. Beijing Jiaotong University, Beijing 100044, China.

A Method for Organizing Long-Distance Passengers on Railway Passenger Dedicated Lines

Rong Hua[1], Jun Liu[2] and Yun Bao[3]

[1] Ph.D. candidate, School of Traffic and Transportation, Beijing Jiaotong University, and The High-speed Railway Technology Department, Railway Transportation Management Bureau, Ministry of Railway of the People's Republic of China, Beijing 100844, Email: ronghua808@hotmail.com
[2] Prof. [3] Ph.D. Candidate, School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, Email: jliu@bjtu.edu.cn, ybao2003@hotmail.com

Abstract

The organization of long-distance passenger flow plays a vital role in the process of line planning. It can improve the efficiency, capacity utilization, and service quality of railway transportation. The organization of long-distance passenger flow is a hot but hard topic, as it involves so many factors and the relationship among them is very complicated. This paper focuses on the organization mode choice of long-distance passengers under different conditions. First, the basic organization schemes of long-distance passengers are discussed. Then, the influence factors of operating long-distance passenger trains are analyzed. In order to evaluate different passenger organization schemes, a cost function is established. Next, a numerical example is presented and examined. The result of the numerical example shows that the proposed method can offer support for decision-making of railway operators. Finally, as passenger transferring is unavoidable sometimes, methods for improving passenger transferring efficiency are discussed.

Key Words: passenger dedicated line; passenger organization mode; through train; transfer

Introduction

The organization of railway long-distance passengers is an important and difficult issue for railway operators, especially when passenger dedicated line network has not been formed yet in China. In this paper, passenger dedicated lines are not the passenger dedicated lines themselves, but they contain traditional railway lines which connect with passenger dedicated lines. From the experience of other high-speed railway countries, it is essential to operating long-distance passenger trains on high-speed railway lines and traditional railway lines.

Before 1987, the high-speed railway was an independent system in Japan, for the track gauge is 1435 mm on high-speed railway, but it is 1067 mm on traditional railway lines, so passengers, who couldn't get to their destination just traveling by one type of line, have to transfer at the break-of-gauge station. After Japanese railway privatization in 1987, in order to get a better profit of the Tōhoku Shinkansen, East Japan Railway Company (JR) changed track gauge Yamagata traditional railway lines from 1067 mm to 1435 mm and added another track on some segments. In order to fit on the clearances of Yamagata railway lines, JR developed the 400 series high-speed trains: The first Mini-shinkansen type, introduced in 1992 on Yamagata Shinkansen Tsubasa services, made passengers first travel directly between Tokyo and Yamagata, and also made the volume of railway passengers increase 56.7% (Nie et al. 2008).

With a dense railway network, it is very convenient for passengers to transfer between high-speed railway lines in France. But most high-speed trains are still operated on traditional railway lines in order to increase passengers' convenience. When high-speed railway was only 1582 km, the running distance of TGV was up to 7500 km. In Germany, when the length of high-speed railway was only 427 km, the operating distance of Inter City Express (ICE) high-speed trains was up to some 4000 km. (Nie et al. 2008).

It is inconvenient for passengers to transfer on traditional railway lines in China both in space and time. In space, most stations are solid, and the routes of passenger entering and exiting the stations are separated, so passengers who need to transfer have to get out of the station and then get in to it. In time, train connections are rarely considered in the period of train timetabling, and in the period of scheduling, although the connecting time is small, for the high utilization of railway capacity, trains are usually late, so the smooth connections aren't guaranteed. Passengers travelling by traditional railway lines would rather travel directly than transfer. With the construction of passenger dedicated lines, and due to the speed advantage of train units, passengers' choice behaviours have changed. Though transfer is inconvenient

for passengers, the operation practice of Wuhan-Guangzhou passenger dedicated line has shown that some passengers traveled by the trains on traditional railway lines to Wuhan railway station, then transfer to passenger dedicated lines to Guangzhou. So we need to research the methods for organizing long-distance passengers on railway passenger dedicated lines.

Maximizing the number of direct passengers is one main objective of line planning in Europe (Dienst 1978; Bussieck 1996; Goossens et al. 2005). In China, Peng (2007) proposed a method for researching passenger transferring schemes from the perspective of passengers. Since passenger travelling demand is larger than the supply capacity of railway transportation, passenger travelling demand is not met very well, so we need to research the passenger organization mode from the perspective of railway operators in order to meet railway passenger travelling demand, meanwhile, to get a good profit.

The Basic Organization Schemes for Operating Long-Distance Passengers on Passenger Dedicated Lines

There are mainly two basic methods to operate long-distance passengers: directly or by transferring. According to the type of trains and lines, the mainly organization methods of passengers' transportation on passenger dedicated lines are as Figure 1.

The affecting factors of operating long-distance passengers on passenger dedicated lines

The Volume of Passengers. It is the main principle for railway operators to operate trains according to passenger volume, and it should be the operation principle of trains on passenger dedicated lines. Typically speaking, passenger volume is the basis of operating trains. The number of train units and the frequency of train units are determined by passenger volume. Passenger dedicated line is a high value-oriented travelling mode, so railway operators need to subdivide passengers to get the number of high value passengers. However, railway train operation can affect the volume of passengers, for it can affect passengers' choice behaviors, so there is a bidirectional relation between passengers' choice behaviors and railway operation.

Resource Constrains. Transportation resource includes infrastructure and mobile devices. The volume and the using condition of infrastructure play a vital role in operating trains. For example, if lines are not electrified, high-speed trains will not run on traditional railway lines. The volume of train units is another constrain that affects the operating of train units on traditional railway lines. The resource that affects the operation of train units can be summarized as follows:

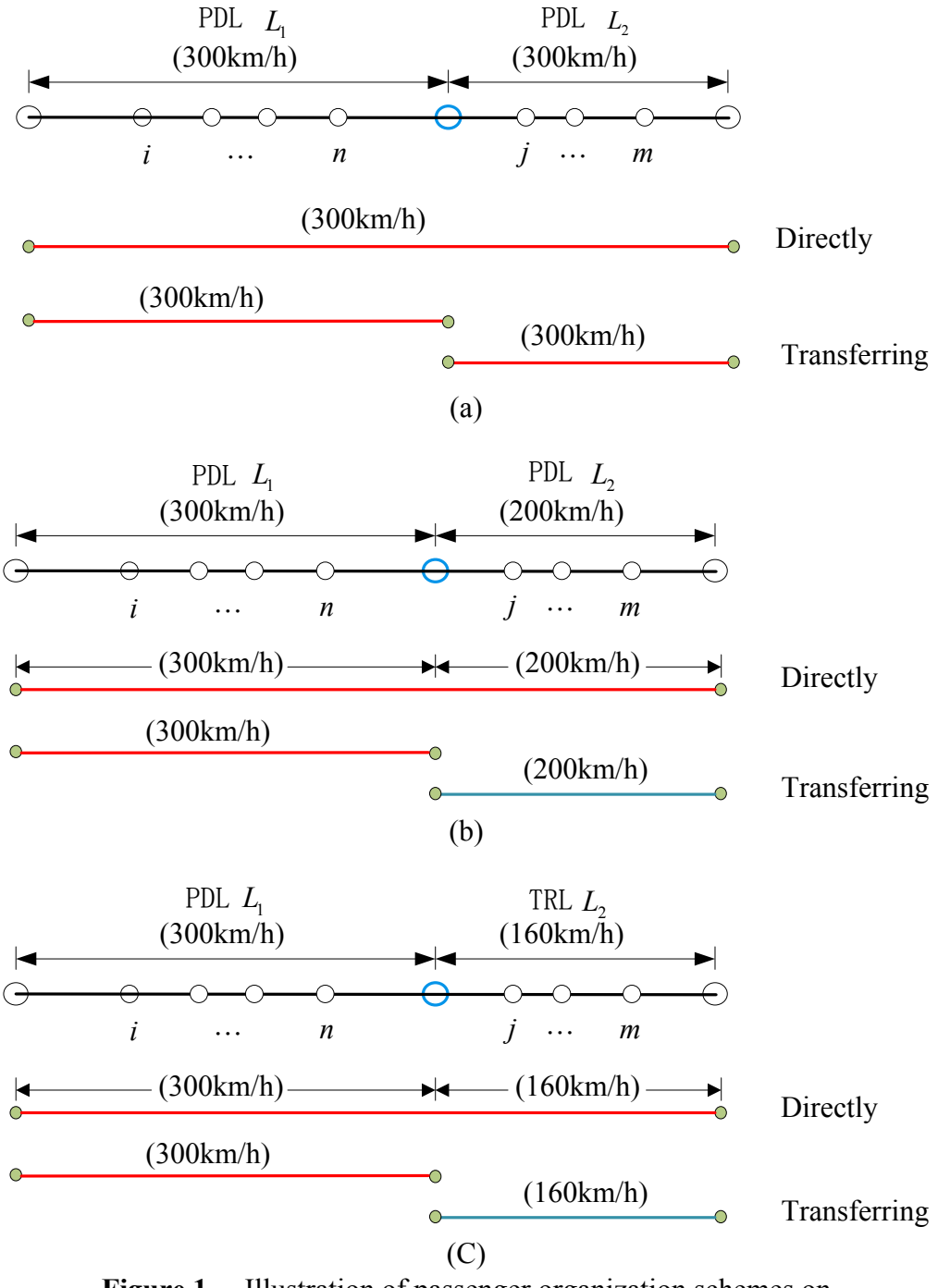


Figure 1. Illustration of passenger organization schemes on passenger dedicated lines

The condition of lines. Electrification and control system of the line determines whether train units run on traditional railway lines. Passenger dedicated lines and traditional railway lines are parallel lines in a transportation corridor. So if train units

run on traditional railway lines, there should be connecting lines which connect passenger dedicated lines and traditional railway lines.

The Condition of Stations. If train units run on traditional railway lines, there should be train unit depots and maintenance depots on traditional railway lines.

The Volume of Train Units and the Maintenance of Train Units. Whether train units run on traditional railway lines depends on the purchasing and productive power of train units. When the volume of train units is limited, organizing passengers to transfer is the most economical choice for railway operators; when the volume of train units is not the constrain, operating train units run on traditional railway lines is favorable to attract more passengers and extend the travelling distance of passenger dedicated lines.

Train units maintenance distance is another constrain. In China, the daily maintenance distance of train units is 3,000 km. So the operation distance of long-distance passenger trains should be less than 3,000 km. Therefore, if passengers' travelling distance is more than 3,000 km, they have to transfer during their trips.

Transportation Operation Mode.

The setting of maintenance windows. Maintenance windows are technical measures to solve the conflict between train operation and infrastructure maintenance. According to the experience of other high-speed railway countries, the maintenance windows are usually set at night from 0:00 to 6:00. As we know, night trains are very popular in China. If we set the maintenance window at night, the night trains must be operated on traditional railway lines, then, the operation condition will be more complex on traditional railway lines, and the capacity of passenger dedicated lines will not be fully used.

Train units travelling speed. Train units stop schemes play a vital role to train travelling speed. If the stop time is 1 min, the additional time for stopping and starting is 3 min and 2 min, and there will be a cost of 6 min at least when high-speed trains stop once. The more times train stop, the more percent of stop time in the whole travelling time is, which will reduce the advantage of high-speed trains.

The cost. Operation cost consists of fixed cost and variable cost. Cost is intimately associated with operation. If high-speed trains run on traditional railway lines, the cost will be significantly increased, for more train units are required, meanwhile, train depots and maintenance depots should be built additionally on traditional railway lines. The control systems of trains on traditional railway lines and on passenger dedicated lines are different. So if train units run on traditional railway

lines, additional control systems should be equipped.

Transportation Operation Methods

There are two basic methods for operating long-distance passengers: operating through trains and organizing passengers to transfer. Through trains can be classified into two types: (1) high-speed train units first run on passenger dedicated lines and then run on traditional railway lines; (2) Low-level train units fist run on traditional railway lines then run on passenger dedicated lines. Passengers who travel by transferring also can be classified into two types: (1) transferring between train units on passenger dedicated lines; (2) transferring between train units on passenger dedicated lines and trains on traditional railway lines.

Passenger dedicated lines are constructed among economic developed cities, so there would be enough passengers. So the operation methods are determined by the cost of railway operators. Therefore, we can build a function that can offer a method for railway operators to choose the most proper method to operate long-distance passengers.

The Cost of Operating through Trains.

Operation cost. Operation cost consists of fixed cost and variable cost. Fixed cost has no influence on the operation, so we don't consider it here. Variable cost consists of the cost per train-km, train stop cost. The cost per train-km is the consumption for each train per kilometer. It includes depreciation expense, energy consumption, material cost, maintenance cost, worker allowances, and so on. Train stops increase the demand of train units. For through trains, we only consider the cost of per train-km.

$$C_1 = \sum_u \sum_e C_{(u,e)} \cdot L_e \cdot \frac{f}{M_u \cdot \rho_u} \tag{1}$$

u - train type;

e - line type;

$C_{(u,e)}$ - the cost per train-km for train u running on line e;

L_e - train running distance on line e;

f - passenger volume;

M_u - the capacity of train u;

ρ_u - the occupation of train u.

Passenger collection cost. The frequency of long-distance through trains is lower than passengers choosing other trains to transfer. So passengers have to wait until there are enough passengers to operate a train. During the waiting process, some passengers may choose other trains or other transportation modes, such as highways or airplanes. So passenger collection cost is the cost of the lost passengers for waiting.

$$C_2 = \sum_u \sum_e R_{(u,e)} \cdot L_e \cdot f' \tag{2}$$

$R_{(u,e)}$ - the fare for each passenger and per train-km of train u running on line e;

f' - the reduced volume of passengers for collection, the value has something to do with train frequency;

The additional cost for train units running on traditional railway lines. The additional cost for train units running on traditional railway lines is caused by the increase volume of train units for train units running on traditional railway lines.

$$C_3 = \sum_u C_{(u)} \cdot n \cdot \lambda \tag{3}$$

$C_{(u)}$ - the cost for purchasing or produce a train units;
n- the increase number of train units for train units running on traditional railway lines;
λ - the probability of train units takes some task.

Therefore, the total cost for operating long-distance trains is:

$$C = C_1 + C_2 + C_3 \tag{4}$$

The Cost for Organizing Passenger to Transfer

Train operation cost

$$C_1' = \sum_u \sum_e C_{(u,e)} \cdot L_e \cdot \frac{f}{M_u \cdot \rho_u} \tag{5}$$

Transferring cost. Transferring cost is the cost caused by the reduced of passengers for transferring, as we know, transferring may increase travelling time, meanwhile, the risk caused by transferring makes passengers prefer to travel directly rather than by transferring. Transferring cost is affected by fare, transferring waiting time. If the

fare and waiting time of transferring is lower than passenger travelling directly, there would be more passengers to transfer.

$$C_2' = \sum_u \sum_e R_{(u,e)} \square L_e \square f'' \tag{6}$$

f'' - the reduced volume of passengers for transferring.

Transferring organizing cost. Transferring organizing cost is the cost for organizing passengers transferring at stations. The cost mainly consists of commission charges and service fees at the station. We assume that the value of transferring cost is δ for each passenger.

$$C_3' = \delta f \tag{7}$$

Therefore, the total cost for organizing passengers transferring is:

$$C' = C_1' + C_2' + C_3' \tag{8}$$

when $C \le C'$, operating through trains is profitable for railway operators.

when $C > C'$, railway operators should organize passengers to transfer.

Numerical example

Figure 2 is a railway network, in which, line A-C is a passenger dedicated line, and other lines are traditional railway lines. The fare for each passenger per train-km is shown Table 1, and the cost of per train-km is as Table 2. The capacity of train units on passenger dedicated lines and trains on traditional railway lines is 1220 P (passenger). We assume that the occupation of train units is 90%, and the occupation of trains on traditional railway lines is 85%. The organizational cost for each passenger is 0.0381 Yuan (Deng 2007). Meanwhile, we assume that the volume of train units is not a constraint for train units running on traditional railway lines, so we don't need to consider the additional cost for train units running on traditional railway lines (C_3). The reduced passengers for collection to operate long-distance through trains occupy 12% of the total passengers.

Taking passengers' organization of segments A-D as an example, when operating through trains, the cost is

$$C = 1398007 \quad (\text{Yuan})$$

when organizing passengers to transfer, the cost is

$$C' = 1225063 \quad (\text{Yuan})$$

Therefore, passengers between A-D should be organized by transferring.

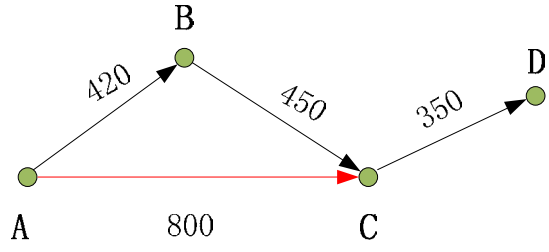


Figure 2. Railway network in numerical example

Table 1. The fare of each passenger per train-km (unit: Yuan).

Train type	Passenger dedicated line	Traditional railway line
High-speed train	0.45	0.16
Middle-speed train	0.30	0.14
Train on traditional line		0.12

Table 2. The cost of per Train-km (unit: Yuan/train-km).

Train type	Passenger dedicated line	Traditional railway line
High-speed train	169.5	264.2
Middle-speed train	127.1	184.1
Train on traditional line		92.4

Table 3. The volume of passengers.

OD	Passenger volume	
	The volume of direct passengers	The volume of transferring passengers
A—B	3000	
A—C	10000	
A—D	6000 (via A-C)	4800 (via A-C)
B—C	4000	
B—D	6000	5200
C—D	8000	

Methods for Improving Passengers' Transferring Efficiency

According to the operation experience of other high-speed railway countries, operating through trains is attractive, but in some situations, it is unavoidable for passengers to transfer. For example, at the beginning of the operation of a passenger dedicated line, the shortage of train units makes it impossible to organize train units run on large-scale traditional railway lines, so passenger transferring is unavoidable. In this situation, how to organize passengers to transfer is an outstanding issue for

railway operators. The situation of passenger transferring on traditional railway lines is that passengers usually have to wait many hours even more than a day. If passengers' transferring situation on passenger dedicated lines is the same as it on traditional railway lines, passenger dedicated lines will lose the speed advantage, and will be with a low passenger volume, therefore, the construction of passenger dedicated lines will not achieve the aim to improve railway transportation capacity. The best method for organizing transferring passengers is to achieve a smooth connection in space and time. In space, methods should be taken to minimize passengers' walking distance. Passengers transferring should be arranged at the same station, even on the same platform. Japanese passenger high-speed train organization mode-"relay trains" should be introduced in China. For example, we can provide relay trains between Beijing and Lhasa where passengers transfer in Lanzhou. Passengers alight at Lanzhou from the train from Beijing to Lanzhou, and then go to the other side of the same platform to board on the linkage train from Lanzhou to Lhasa, which is prepared in advance. This transferring mode only takes 3~4 min (Chen 2007). In time, the connection time between trains should be less than 30 min.

Conclusions

This paper addresses passengers' organization methods, based on the analysis of the factors that affect the organization of long-distance passengers on passenger dedicated lines. An evaluation function for different passenger organization schemes was established from the perspective of railway operators. Meanwhile, the function was verified by a numerical example. The result showed that the evaluation function can provide support for railway operators to make decisions. As sometimes it is unavoidable for passengers to transfer, the methods for improving passengers' transferring efficiency were discussed.

The paper is only an initial step from the perspective of railway operators. Further studies can be conducted to study passengers' choice behaviors, and the relationship between the decision made by railway operators and the choice behaviors of passengers.

Acknowledgments

The work described in this paper was jointly supported by Science and Technology of China, Program for National High-Technology (863) Research (No.2007AA11Z220), and State Key Lab of Rail Traffic Control & Safety, Beijing Jiaotong University (No.RCS2009ZT009). We would like to give our sincere thanks to them!

References

Bussieck, M. R., Kreuzer, P., and Zimmermann, U. T. (1996). "Optimal lines for railway systems" *European Journal of Operational Research*, 96, 54-63.

Chen, H. (2007). "Research on the linkage of transportation between passenger's dedicated lines and existed railways" *Master thesis*, Beijing.

Deng, L. (2007). "Study on the optimal problems of passenger train plan for dedicated passenger traffic line." *Ph.D thesis,* Changsha.

Dienst, H. (1978). "Linienplanung in Spurgeführten Personenverkehrmit Hilfe eines Heuristischen Verfahrens." *PhD thesis,* TU Braunschweig (in German).

Goossens, J. H. M., van Hoesel, C. P. M., Kroon, L. G. (2005). "On solving multi-type railway line planning problems." *European Journal of Operational Research*, 168, 403-424.

Nie, L., Zhao, P., Jia, L., Zhou, L. (2008). "The operation technology of passenger dedicated lines." *The press of Beijing Jiaotong University.*

Peng, Q. (2007). "Transportation organization of passenger special line." *Science Press.*

The Operating Conditions of Long-Distance Passenger Trains on Railway Passenger Dedicated Lines

Yun Bao[1], Rong Hua[2] and Jun Liu[3]

[1] Ph.D. candidate, [3]Professor, School of traffic and transportation, Beijing Jiaotong University, Beijing 100044, Email: ybao2003@hotmail.com, jliu@bjtu.edu.cn
[2] Ph.D. candidate, School of traffic and transportation, Beijing Jiaotong University, and The High-speed Railway Technology Department, Railway Transportation Management Bureau, Ministry of Railway of the People's Republic of China, Beijing 100844, Email: ronghua808@hotmail.com

Abstract

Long-distance passenger trains are very popular in other high-speed railway countries and on traditional railway lines in China. However, the organization of long-distance passenger trains can be complicated on passenger dedicated lines in China, for it is affected by many factors. In this paper, we focus on the analysis of the factors to develop the conditions for operating long-distance passenger trains on railway passenger dedicated lines. Since different factors have different degrees in affecting the operation of long-distance passenger trains, we discuss the combination effect of different factors. Based on the analysis, we propose the conditions for operating long-distance passenger trains on passenger dedicated lines, including train units travelling distance, and train departure and arrival time windows. Finally, we discuss the relationship between train stop schemes and train departure and arrival time windows of the Beijing-Shanghai high-speed railway line.
Key Words: passenger dedicated line; long-distance passenger train; travelling distance; departure and arrival time window

Introduction

With the construction of high-speed railways, the organization of long-distance passenger trains is an important issue to be solved. Long-distance passenger trains are convenient for direct passengers, since they don't have to transfer during their trips. Direct passengers are the passengers that can travel from their origin station to their destination station without having to change trains during their journey (Goossens et al., 2004). To maximize the number of direct passengers is one main

objective of line planning in Europe (Dienst, 1978; Bussieck et al., 1996). The other is to minimize the cost (Claessens et al., 1998; Goossens et al., 2004, 2005).

From the experience of other high-speed railway countries, it is essential to operating long-distance passenger trains on high-speed railway lines. For one thing, it is very convenient for passengers. For the other thing, it can extend the running distance of high-speed trains.

Long-distance passenger trains are very popular on traditional railway lines in China. There are 2669 passenger trains running on traditional railway lines, and the distribution of long-distance passenger trains are shown in Table 1. From the data, we can see that the long-distance passenger trains occupy more than a half in the total railway passenger transportation.

Table 1. Long-distance passenger trains on traditional railway lines in China.

Train running distance (km)	The number of trains	Percentage (%)
1000–2000	1073	40.2
2000–3000	309	11.6
3000–4000	53	2
>4000	8	0.3
>1000	1443	54

Therefore, from the experience of other countries and the operating experience of passenger trains on traditional railway lines in China, we can conclude that it is essential to operate long-distance passenger trains on passenger dedicated lines in China. There are many factors that affect the operating of long-distance passenger trains, and the affecting degree of each factor is different, so we need to research the degree of each factor and the combinational effect of different factors. The remainder of the paper is organized as follow. Section 2 analyzes the main factors that affect the operating of long-distance passenger trains on passenger dedicated lines. In section 3, the running distance and train departure and arrival time windows of long-distance passenger trains will be proposed. Meanwhile, the relationship between train stop schemes and train departure and arrival time windows will be discussed in the research background of Beijing-Shanghai high-speed railway line. Section 4 presents concluding remarks.

The Factors Affecting the Operation of Long-distance Passenger Trains on Passenger Dedicated Lines

Passengers Demand Forecasting. Passengers' purchasing power determines the type of train units to operate, and passengers' volume determines the number and frequency of train units running on passenger dedicated lines. So passengers demand forecasting plays a vital role in operating long-distance passenger trains, especially the forecasting of short-term travelling demand. With the rapid changes of the market,

the conventional passenger travelling demand prediction methods may produce large relative warps. For example, the predicted passenger travelling demand would be 2 billion passengers a year between Beijing-Shanghai railway corridor by considering the development of economy, growth of population, and other factors. But for the high fare of high-speed trains and the discount of airline fare, a lot of passengers may choose to travel by air, or for the high-speed of trains and lower fare compared to travel by air, high-speed trains may induce more passengers. So passengers' choice behaviors should be modeled in the process of long-distance passenger demand prediction.

The Operation Mode of High-speed Trains on Passenger Dedicated Lines. The type of trains and how railway operators organize the trains on passenger dedicated lines are the issues of operation mode. It determines train units type, train units running distance, train units utilization and maintenance. If the lack of train units is the major limitation for operating passenger dedicated lines, high-speed trains should only be operated on passenger dedicated lines. In the future, China will produce enough train units, at that time, high-speed trains should run on traditional railway lines as high-speed trains in other high-speed railway countries, like Germany, France, etc.

The Maintenance Distance of Train Units. There are five degrees of high-speed train unit maintenance: OR, R3, R2, RI, and DI, and their respective maintenance distance is 1200 thousand km, 240 thousand km, 60 thousand km, 12 thousand km, and 3 thousand km. So the operation distance of long-distance passenger trains should be less than 3 thousand km.

The Setting of Maintenance Windows. Maintenance windows are technical measures to solve the conflict between train operation and infrastructure maintenance. There are mainly two types of maintenance windows: rectangle-shaped and V-shaped. Other types of maintenance windows are developed by these two types. By far, the recommended maintenance window for passenger dedicated lines is rectangle-shaped window. The advantage of rectangle-shaped window is that the maintenance is not affected by train operation, so the working efficiency is much more effective, and the working condition is much safer. However, during the maintenance process, the two lines are out of work, so it affects the utilization of railway capacity. According to the experience of other high-speed railway countries, the maintenance windows are usually set from 0:00 to 6:00 at night. Considering safety, efficiency, and the utilization of railway capacity, we propose that rectangle-shaped maintenance window should be adopted on passenger dedicated lines, and the time should be at night. Night trains could be operated on traditional railway lines during the maintenance time.

The Rationality of Train Departure and Arrival Time. Heetal. (2006) proposed the concept of convenience to judge the rationality of train departure and arrival time. They made some conclusions based on questionnaires. The research was from the perspective of customers, but didn't consider the setting of maintenance windows. Train departure and arrival time windows are associated with train types, train stops,

and the time at each stop station. The travelling time of the train is as follows:

$$t = S/V + \sum_{i=0}^{n}(t_{ah} + t_i + t_{as}) \quad (1)$$

S —distance from origin to destination;
V —train speed;
n — train stops;
t_i —the stop time at station $i(i = 1, 2, \cdots, n)$;
t_{ah} —additional time for train stopping; and
t_{as} — additional time for train starting.

If the time maintenance window is between 0:00 and 6:00 at night, the rational departure time of trains should not be earlier than 6:00 and later than $(24:00 - t)$; and the rational arrival time of trains should not be earlier than $(6:00 + t)$ and later than 24.

Train Stop Schemes. Train travelling speed on passenger dedicated lines is mainly affected by train stop schemes and different speed restrictions along the route. The stop schemes (train stops, and the stop time at each stop station, etc.) play a vital role to train travelling speed. Therefore, high-speed trains that have a very high maximum operating speed might still achieve a relatively low average speed. Japan and France provide the fastest services in the world at average speeds of around 260 kph (Takagi, 2005). The Nozomi service on the Tokaido line, which stops only at the main stations along the route, achieves an average speed of 206 kph between Tokyo and Osaka (Givoni, 2006). This is significantly lower than the maximum speed achieved by these high-speed trains. Meanwhile, according to the experience of Japanese high-speed railway, operators rarely operate non-stop trains and frequent stop trains simultaneously in order to reduce the heterogeneity of timetable.

The Utilization of Line Capacity. The number of long-distance passenger trains is restricted by the utilization of line capacity. The utilization of line capacity is affected by train rescheduling. In order to guarantee a good condition for train rescheduling, the utilization of line capacity should be controlled under some level.

The Operating Cost. Operating cost is determined by operating modes. If high-speed trains run on traditional railway lines, the cost will be significantly increased, for it needs more train units, meanwhile, train depots and maintenance depots should be built on traditional railway lines additionally.

Passenger Fatigue Distance. Travelling is a process of strength consumption. People will be tired when their travelling time is several hours, and the feeling will become worse with the travelling, so passengers need additional time to recover. The general researches consider passenger recover time from fatigue and add it in the whole travelling time. Passenger recover time is associated with travelling time and circumstances, and travelling circumstances is associated with the type of trains. Peng et al. (2007) gave following equation for passengers recovering from fatigue:

$$g_u(t) = M/[1 + a\exp(-bt)] \tag{2}$$

$$a = \sum_u (a_u t_u) / \sum_u t_u \tag{3}$$

$$b = \sum_u (b_u t_u) / \sum_u t_u \tag{4}$$

M —the limited recover time, in general, M is 15h;

a_u —nondimensional parameter, when train type is u, and $t = 0$, the recovery time is $M/(1 + a_u)$;

b_u —the strength coefficient of recovery time for one travelling hour, $b_u > 0$, the unit is h^{-1}.

when M is 15 h, $a = 59$, and $b = 0.28$, the recover curve is as Figure 1.

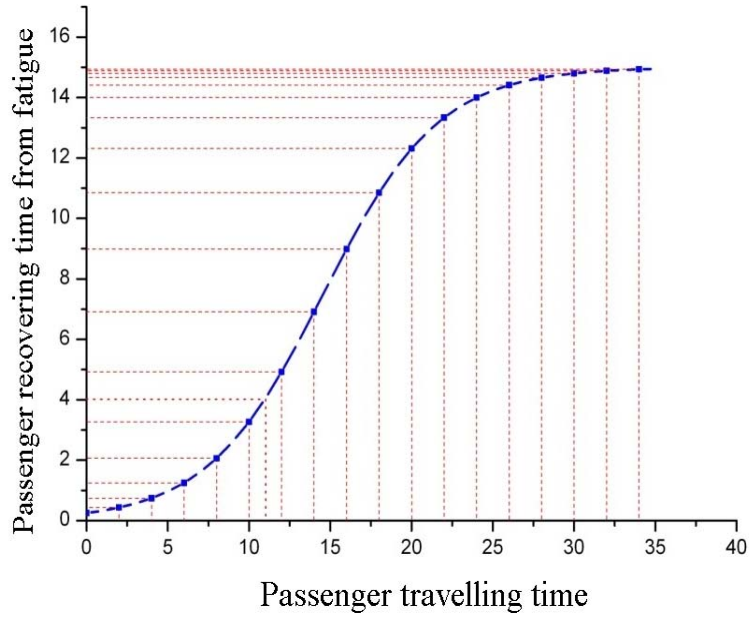


Figure 1. The relationship between travelling time and passenger recovering time on high-speed railway lines.

From Figure 1, we can see that passenger recover time increases with travelling time. When the travelling time is less than 8 h, the increase of recovering time is slower than increasing of travelling time more than 8 hours. When travelling time is up to some hours (about 16h), the increase of recovering time is slower too. So the travelling time on passenger dedicated lines should be kept to less than 8 hours. According to the type of trains and the demand of travelling comfort, the travelling distance should be kept within bounds. If we assume that the train travelling speed is v, affected by passenger fatigue recovering time, the maximum travelling distance should be:

$$S_{\max} = v \times 8h \tag{5}$$

If the train's technical speed is 350 km/h, the coefficient of the train speed is 0.85 (it is derived from the average speed and the operation speed of high-speed trains in Japan and France), the distance should be less than 2380 km ($350\text{km/h} \times 0.85 \times 8\text{h}$).

However, there are still shortcomings in this approach, for we only considered train types, but don't consider seat types. The recovery time is associated with train seats, and it will be different for passenger with a sleeping berth and a normal seat.

The Operating Conditions of Long-distance Passenger Trains

The Travelling Distance of Long-distance Passenger Trains.

The maximum travelling distance of long-distance passenger trains. The travelling distance is affected by train unit maintenance distance, the setting of maintenance windows, passenger fatigue distance, train stop schemes, and so on. The daily train maintenance distance is 3 thousand km, so the operation distance of long-distance passenger trains should be less than 3 thousand km. As affected by passenger travelling fatigue recovering time, the operation distance of trains should be no more than $S_{\max}$.

Peng, et al. (2007) researched the maximum train operation distance on passenger dedicated lines, under constrains of a maintenance window, train stop time, the utilization of line capacity, and other factors. The maximum operation distance is:

$$S'_{\max} = \frac{1440 - T \cdot \beta - \sum T_i}{60} \cdot \gamma \cdot V \cdot \alpha \tag{6}$$

T —time of maintenance window, unit: min; β —the coefficient of adjacent maintenance windows, empirical value is between 1.2 and 1.7; $\sum T_i$ —the sum time of train stop; γ — the utilization of line capacity; V — train speed, unit: km/h.

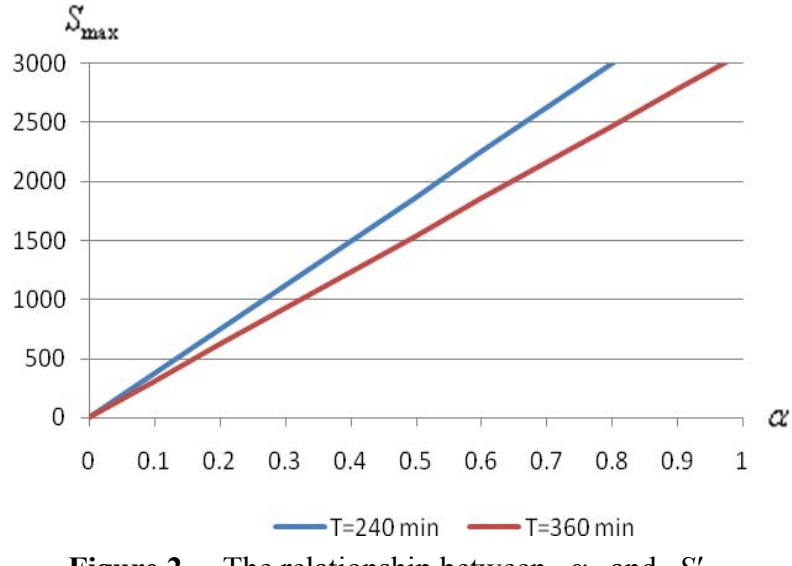


Figure 2. The relationship between α and $S'_{\max}$.

α - the affecting coefficient of trains. The value of α is determined by: (1) the ratio of the same or similar operation distance of trains. The bigger the ratio is, the smaller α is; (2) the equilibrium of operating the same or similar distance of trains, especially the difference of operating the same or similar distance of trains between peak hours and off-peak hours. The bigger the difference is, the smaller α is. The relationship between α and $S'_{\max}$ is as Figure 2.

Therefore, the maximum train operating distance on a passenger dedicated line is:

$$d_{\max} = \min\{3000km, S_{\max}, S'_{\max}\} \tag{7}$$

The running distance of train units on traditional railway lines. Considering the high cost of train units, if train units run on traditional railway lines, we suggest that the running distance of train units on traditional railway lines should be less than $d_{\max}/2$. The running distance of train units on traditional railway lines is mainly determined by profit. The main revenue comes from tickets, and it is determined by passenger volume, passenger fare per kilometer, and passenger travelling distance.

$$P = \sum_k \sum_e \sum_u R(u,e) \cdot L_e \cdot f_k \tag{8}$$

k —passenger travelling scheme, 1 - direct travelling (passengers remain on the same train for the entire trip), 2 - indirect travelling (passengers transfer to another train(s) before they reach their destination);

e —line type, 1 - passenger dedicated lines, 2 - traditional railway lines;

u —train type, 1 - train units, 2 – normal trains on traditional railway line;

$R_{(u,e)}$ —per kilometer fare for each passenger travelling by train u on line e;

L_e —travelling distance on line e;

f_k —the total passenger volume of scheme *k*.

The operating cost consists of two parts: fixed cost and variable cost. The fixed cost has no effect on the passenger operating mode, so we don't need to consider it. The variable cost includes unit distance price $C_{(u,e)}$ and passenger transferring organization fee (ε Yuan for each passenger).

$$C = \sum_k \sum_e \sum_u C_{(u,e)} \cdot L_e \cdot \frac{f_k}{M \cdot \rho} + \varepsilon \cdot f_2 \tag{9}$$

If we assume that P_d is the revenue of operating long-distance direct passenger trains, and P_t is the revenue of organizing passenger to transfer. C_d and C_t are the cost of operating long-distance direct passenger trains and organizing passengers to transfer.

$$P_d = (R_{(1,1)} \cdot L_1 + R_{(1,2)} \cdot L_2) \cdot f_1 \tag{10}$$

$$P_t = (R_{(1,1)} \cdot L_1 + R_{(2,2)} \cdot L_2) \cdot f_2 \tag{11}$$

$$C_d = (C_{(1,1)} \cdot L_1 + C_{(1,2)} \cdot L_2) \cdot \frac{f_1}{M_1 \cdot \rho_1} \tag{12}$$

$$C_t = (C_{(1,1)} \cdot L_1 \cdot \frac{f_2}{M_1 \cdot \rho_1} + C_{(2,2)} \cdot L_2 \cdot \frac{f_2}{M_2 \cdot \rho_2}) + \varepsilon \cdot f_2 \tag{13}$$

M_u - train capacity;

ρ_u - train occupancy.

If we assume $f_1 = f_2 = f$, then

$$(R_{(1,2)} - R_{(2,2)} + \frac{C_{(2,2)}}{M_2 \cdot \rho_2} - \frac{C_{(1,2)}}{M_1 \cdot \rho_1}) \cdot L_2 > -\varepsilon \tag{14}$$

when $R_{(1,2)} - R_{(2,2)} + \frac{C_{(2,2)}}{M_2 \cdot \rho_2} - \frac{C_{(1,2)}}{M_1 \cdot \rho_1} > 0$,

$$L_2 > \frac{-\varepsilon}{R_{(1,2)} - R_{(2,2)} + \frac{C_{(2,2)}}{M_2 \cdot \rho_2} - \frac{C_{(1,2)}}{M_1 \cdot \rho_1}} \tag{15}$$

when $R_{(1,2)} - R_{(2,2)} + \frac{C_{(2,2)}}{M_2 \cdot \rho_2} - \frac{C_{(1,2)}}{M_1 \cdot \rho_1} < 0$,

$$L_2 < \frac{-\varepsilon}{R_{(1,2)} - R_{(2,2)} + \frac{C_{(2,2)}}{M_2 \cdot \rho_2} - \frac{C_{(1,2)}}{M_1 \cdot \rho_1}} \tag{16}$$

Therefore, the running distance of train units on traditional railway lines d_c is:

$$\text{when} \quad R_{(1,2)} - R_{(2,2)} + \frac{C_{(2,2)}}{M_2 \cdot \rho_2} - \frac{C_{(1,2)}}{M_1 \cdot \rho_1} > 0, \quad L_2 < \frac{1}{2} d_{\max}, \quad L_2 \le d_c < \frac{1}{2} d_{\max} \tag{17}$$

$$\text{When} \quad R_{(1,2)} - R_{(2,2)} + \frac{C_{(2,2)}}{M_2 \cdot \rho_2} - \frac{C_{(1,2)}}{M_1 \cdot \rho_1} > 0, \quad L_2 > \frac{1}{2} d_{\max}, \quad d_c = 0; \tag{18}$$

$$\text{When} \quad R_{(1,2)} - R_{(2,2)} + \frac{C_{(2,2)}}{M_2 \cdot \rho_2} - \frac{C_{(1,2)}}{M_1 \cdot \rho_1} < 0, d_c \le \min\{\frac{1}{2} d_{\max}, L_2\}. \tag{19}$$

Train Departure and Arrival Time Window on Beijing-Shanghai High-speed Railway Line. The departure and arrival time windows are associated with the setting of maintenance windows, train speed and train stop schemes, and so on. We take the Beijing-Shanghai high-speed railway line as an example. The overall distance of the line is 1318 km, and there are 21 stations. We assume that the maintenance windows are rectangle-shaped, and the maintenance time is from 0:00 to 6:00 at night. Train speed is 350 km/h, and the stopping time is 6 min (stopping time is 1 min, and additional time for stopping and starting is 3 and 2 min) at each stop station. Therefore, the departure and arrival time windows are shown in Table 2.

From Table 2, we can see that: 1) train stops have significant effect on train departure and arrival time. The travelling time of trains that stop at each station is 114 min more than non-stop trains. As passenger volume at each station is different, the average stopping time would be more than 1 min, so the total stop time will take a larger part in the total travelling time. 2) the departure and arrival time windows is

affected by the setting of maintenance windows. If the maintenance window is rectangle-shaped and the time is from 0:00 to 6:00 at night, the earliest train arriving at Shanghai is at 9:46 a.m. If the transportation condition is congested in the destination city, the golden time for business men is gone in the morning. It implies that the night trains on high-speed lines are necessary.

Table 2. The departure and arrival time windows under different stop schemes on the Beijing-Shanghai high-speed line.

Stops	The departure time window	The arrival time window	Stops	The departure time window	The arrival time window
0	6:00-20:14	9:46-24:00	10	6:00-19:14	10:46-24:00
1	6:00-20:08	9:52-24:00	11	6:00-19:08	10:52-24:00
2	6:00-20:02	9:58-24:00	12	6:00-19:02	10:58-24:00
3	6:00-19:56	10:04-24:00	13	6:00-18:56	11:04-24:00
4	6:00-19:50	10:10-24:00	14	6:00-18:50	11:10-24:00
5	6:00-19:44	10:16-24:00	15	6:00-18:44	11:16-24:00
6	6:00-19:38	10:22-24:00	16	6:00-18:38	11:22-24:00
7	6:00-19:32	10:28-24:00	17	6:00:18:32	11:28-24:00
8	6:00-19:26	10:34-24:00	18	6:00-18:26	11:34-24:00
9	6:00-19:20	10:40-24:00	19	6:00-18:20	11:40-24:00

From the above analysis, the setting of maintenance and train stop schemes are the main factors which determine train departure and arrival time on the Beijing-Shanghai high-speed railway line.

Conclusions

This paper analyzes the main factors that affect the operation of long-distance passenger trains on passenger dedicated lines. The combination effects of different factors are discussed based on the analysis of each factor. Then, the conditions for operating long-distance passenger trains on passenger dedicated lines are proposed, including train units running distance on passenger dedicated lines and on traditional railway lines, and train departure and arrival time windows. Finally, in the research background of Beijing-Shanghai high-speed railway line, the relationship between train stop schemes and train departure and arrival time windows is shown.

Acknowledgements

This study was jointly funded by National Natural Science Foundation of China (No.60736047), State Key Lab of Rail Traffic Control & Safety, Beijing Jiaotong University (No.RCS2009ZT009) and Beijing Jiaotong University (No.141078522). The authors deeply appreciate the support.

References

Bussieck, M. R., Kreuzer, P., and Zimmermann, U.T. (1996). "Optimal lines for railway systems," *European Journal of Operational Research*, 96, 54-63.

Claessens, M. T., van Dijk, N. M., and Zwaneveld, P. J. (1998). "Cost optimal allocation of passenger lines," *European Journal of Operational Research*, 110, 474-489.

Dienst, H. (1978). "Linienplanung in Spurgeführten Personenverkehrmit Hilfe eines Heuristischen Verfahrens," Ph.D. thesis, *TU Braunschweig* (in German).

Givoni, M (2006). "Development and impact of the modern high-speed train: a review," *Transport Reviews*, 26(5), 593-611

Goossens J. W, van Hoesel, S. and Kroon, L. G. (2004). "A branch-and-cut approach for solving line planning problems," *Transportation Science* 38, 379-393.

Goossens, J. H. M., van Hoesel, C. P. M., and Kroon, L. G. (2005). "On solving multi-type railway line planning problems," *European Journal of Operational Research*, 168, 403-424.

He, Y., Zhang, H., Mao, B., Chen, T. (2006). "Multi objective bi-level programming model of making train working plan for passenger-only line," *Journal of the China railway Society*, 28(5), 6-10.

Peng, Q. (2007). "Transportation organization of passenger special line". *Science Press*.

Takagi, R. (2005). "High speed railways: the last 10 years," *Japan Railway and Transport Review*, 40, 4-7.

Safety Management System in Civil Aviation: Our Perspective

Leping Yuan[1], Ruishan Sun[2] and Dongbin Li[3]

[1,2] Research Institute of Civil Aviation Safety, Civil Aviation University of China, Tianjin 300300, China, E-mail: lpyuan@hotmail.com, sunrsh@hotmail.com
[3] Air Traffic Management Department, Civil Aviation Management Institute of China, Beijing 065001, China, E-mail: ldbpjcd@hotmail.com

Abstract

Safety Management System (SMS) is currently being advocated by civil aviation community, and has been introduced in China in the last few years. This paper covers perception of SMS, discusses some fundamental issues concerned. It tries to present a clear view of SMS, furthermore, hopefully leads to a better understanding and implementation.
Key Words: civil aviation; safety; SMS; safety risk management; state safety program

Introduction

Safety Management System as defined by International Civil Aviation Organization (ICAO), an organized approach to managing safety, including the necessary organization structures, accountabilities, policies and procedures, is now being strongly pushed by civil aviation community. By now ICAO annex 1, annex 6, annex 11 and annex 14 have been revised respectively, which mandate the use of SMS in air carriers, air navigation service providers (ANSPs) and aerodromes. In 2006, ICAO historically released its first edition of Safety Management Manual (DOC9859) for the guidance of SMS implementation. SMS is placed in a key position in ICAO global safety improvement strategy (ICAO 2006).

In the past few years, China civil aviation safety has remained well-behaved, and even better than ever. Statistic shows that air accident rate was 0.23 per million flight hours in China over the period 1998-2007, which is far better than the world average 0.33. And up to the present China has flown safely and continuously over 18 million hours, which makes a new domestic record of aviation safety.

However, things are still bothering the industry. China is one of the states with fastest growth in air transportation. According to ICAO statistics, in 2006 the total turnover volume of air transportation reached 30.6 billion ton-km ranking 2nd in the world. Rapid growth in air transportation has brought huge safety pressure to the industry, safety assurance and improvement of air transportation system is more pressing than ever.

SMS is a strategic choice made by China. By now under the leadship of the CAAC, relevant work is proceeding in an orderly manner, and preliminary progress has been achieved. In 2007, the CAAC released its General Implementing Scheme on Safety Management System, which outlined the frames and requirements of SMS.

Moreover, relevant guidance materials regarding SMS implementation for air carriers, ANSPs and aerodromes have been released subsequently. Today the community is making joint efforts to establish SMS with Chinese features in civil aviation.

Where does it from?

Before we come to reveal the nature of SMS, it's better to understand its original background, to answer the question why is SMS? Apparently public expectation on aviation safety forces the community to innovate continuously. SMS is a rational model for an organization managing safety, and it is also a probe of pursing better safety performance in aviation.

Traditional "regulate-oversight" model in aviation safety management exposed its inadequacy in accident prevention. Compliance with regulation alone did not guaranty 100 percent of safety, and a proactive way is expected for management enforcement. Meanwhile safety philosophy evolves all the time, from technology to human factor and now organizational perspective, which requires people addressing safety issues in a larger context with a systematic approach. SMS is created adopting a proactive and systematic approach to match the development of safety management. For the past decades, PDCA, TQM and QMS (and derivative OHSMS and EMS of course) have archived great success around the world, which provide models to run a management system, SMS employs PDCA cycle and deals with safety issues like those in quality, environment, and finance (Stephen etc 2004).

SMS that shares common resemblance with the one in aviation， is already existed in marine, chemical industry and other high risk industry (Fang and Fang Jin 2004), which means experience can be learnt for higher safety performance.

Upon the whole, SMS is an integration of management system principle and system safety best practice, and manages safety with more systematic and proactive approach.

SMS Structure

Although different organization superficially holds different view about SMS, common understanding is actually accepted. Generally SMS can be grouped into 3 main sections:

- SMS Foundation (SMSF) includes safety police, safety objectives, organization structure & responsibility, document system and other resources supporting safety management activities;
- Safety Management Program (SMP), refers to all kinds of safety management activities for the realization of preset safety objectives;
- SMS Evaluation, Supervision & Improvement (SMSESI), refers to those activities for the purpose of continuous improvement of SMS.

Each organization would have their own philosophy in management, so does safety management. Safety objectives and safety policy are the explanation of safety management philosophy, and answer the question of how safe it would be and by what means to achieve it. By combining with other resources, it forms the foundation of SMS.

SMPs are those activities launched for safety performance improvement, including risk management, safety information management, incident investigation, safety training etc, all these activities involved are for the purpose of achieving safety goals. SMSESI is designed to control SMS operation, and assess its effectiveness in order to make improvement. Safety inspection, safety performance management, safety auditing, safety management review are this kind of tools. A continuous improvement is addressed through the regular effective application of these tools.

A positive organization safety culture is the ultimate goal of SMS, meanwhile good safety culture will definitely benefits SMS itself, and leads to a prominent safety performance.

Safety Risk Management

Safety risk management is a core business of SMS, it's a formalized process of identification, analysis and elimination (and/or mitigation to an acceptable or tolerable level) of hazards, as well as the subsequent risk, that threaten the viability of an organization, simply say risk management is a method whereby an organization examines and treats with its deficiency (Luo, 2004). Risk management origins from the insurance industry, and focuses on protecting against catastrophe in the early stages. When insurers came to realize that their results were enhanced by encouraging customers to exercise reasonable care and by rewarding good performance, risk management evolved from natural intuition into a more formal process of communication of the controls to influence outcomes. Now risk management is widely adopted in different industries in the world.

Basically, risk management can be regarded as two-step process i.e. risk assessment and risk control. The task of the former is about identification and analysis of factors that would harm the system. Risk assessment is similar to safety assessment. Today we usually weight safety of the system by measuring its risk, so safety assessment is equivalent with risk assessment in this sense. We employ both terms in Guidance Manual on Safety Management System for ATM System, but safety assessment aims at support of decision-making before a proposed change happens and risk management mainly focus on daily operation.

We can definitely say that SMS is not established without an effective risk management procedure, but how to carry out risk management effectively is a question deserves to consider.

An organization is suggested to integrate risk management into operation, take various effective means to identify all risks exist in or around the system, set up risk database and establish comprehensive dynamic procedures for continuous monitoring of risk, ensure that risks are in the controllable state. A network for the purpose of risk information acquirement and communication will benefit the credible risk assessment (whereas a qualitative approach if appropriate), as well as risk monitoring.

SMS Implementation

The establishment of SMS can generally be divided into three stages:
- Planning and preparation phase,
- Initial operation phase, and
- Continuous operation phase.

The main tasks of planning and preparation phase include knowledge reservation, support of the CEO & working staff, appointment of safety management representative, and making plan. Support from decision maker is of paramount importance for implementation, which is the motivation of all staff and the force of gathering each department involved.

A well-developed overall action plan is another major issue. Survey of present safety management and comparison with SMS requirement should be made to determine what have already been in existence and what needs to construct or improve. This process is usually named as “gap analysis”, and outputs a detailed action plan in which tasks, objectives, responsibilities and privileges are clearly stated, and time schedule and progress requirement be well arranged to ensure that all tasks carried out in an orderly manner.

Upon completion of planning and preparatory work, each task should go as it planned. Any change in operation should be documented, related procedures and manuals should be revised to form complete hierarchical document system. Running base on the new procedures or manuals can be called "a trial run", it’s aiming at the

detection of operation problems, accumulation of evidence of SMS effectiveness for the sake of system amelioration.

Safety management often compares to fighting a guerrilla war, which implies there are no final victories (Reason 2001), so SMS is facing the same situation as well, which means continuous improvement is needed. Safety audit and management review are important means to perfect SMS. Conduction of regular audit is to ensure sustainable and healthy operation of SMS in civil aviation entities, and audit methods can be developed from QMS. Of course, an external audit will normally benefits the system; such mechanism should be put in place in order to push SMS forward.

Management review is a higher level assessment of SMS conformity, effectiveness and suitability, and is organized by the CEO, to determine the direction of SMS and find solutions for those crucial safety issues. Through the regular performing of safety auditing and management review, a self-improving SMS would continually move forward.

State Safety Program

The State Safety Program (SSP) is an integrated set of regulations and activities aimed at improving safety. We define SSP including the following elements: Safety Plan (SP), Safety Information Management (SIM), Safety Resources (SR), Safety Organization (SO), Safety Regulation & Standard (SRS), Safety Oversight (SO), Safety Risk Management (SRM), Safety Education & Training (SET), Safety Investigation(SI), Safety Technology(ST), Emergency & Response (ER), Safety Culture(SC).

It is the duty of the regulator to set the SSP, and SSP forms the superstructure of state civil aviation safety management.

All the elements should be clearly specified in the SSP. Air operators, ANSPs and aerodromes build SMS under the frame of SSP and each element of SMS is fully compatible with SSP.

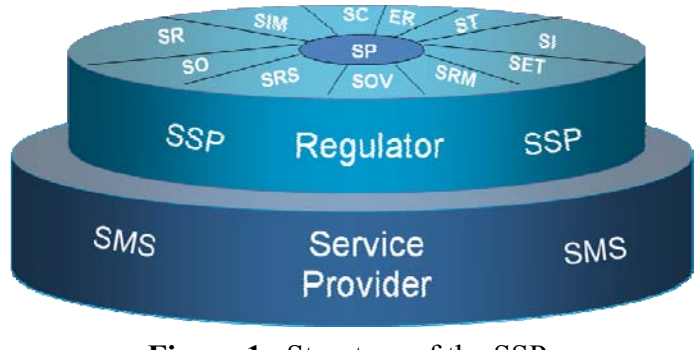


Figure 1. Structure of the SSP.

Conclusions

The establishment of SMS is the strategy for safety management upgrade and ultimately long-term safety, and substantial progress has been obtained. However challenge still remains, the effectiveness of a genuine application of SMS will certainly be a dream without real understanding. There must be a shift of mind for the difference between traditional discipline and the advanced one, and development in both theory and practice of safety management.

Implementation of SMS under an appropriate regulatory regime is suggested, the SSP should be put in place to help the service provider implement SMS. It is certain that along with the deep going of SMS, it will bring a new evolution to air transportation system.

References

International Civil Aviation Organization. (2006). *Safety Management Manua*l, Montreal.

Stephen, C., Yu, K., Hunt, B. (2004). "A fresh approach to safety management systems in Hong Kong." *TQM Magazine*, 210-215.

Fang Quangeng, Fang Jin (2004). "FSA and its applications to the safety of ships." *Navigation of China*, 1-15.

Luo, Y., Fan Y. X., and Ma X. C. (2004). *Risk Analysis and Safety Assessment,* Chemical Industry Press, Beijing.

Reason, J. (2001). "In search of resilience." *Flight Safety Australia*, 26-28.

A Slot Allocation Algorithm for the Flights

Tiangang Zhang[1], Ronglu Sun[1], Xiaoyun Hou[2] and Rentian Yue[3]

[1] Lecturer, School of Mechanical and Electronic Engineering, Tianjin Polytechnic University, No. 63 Chenglin Road, Hedong District, Tianjin,300160, P. R. of China
[2] College of Airport Engineering, Civil Aviation University of China
[3] College of Air Traffic Management, Civil Aviation University of China

Abstract

A rational slot allocation for flights is very important to ease the congestion in airspaces or landing airports. Based on the analysis of slot allocation algorithms of foreign air traffic flow management systems and with the combination of the domestic air traffic flow management situation, a flight slot allocation algorithm suitable for China is proposed. Test of the algorithm indicated that the proposed flight slot allocation algorithm is reasonable, effective and practical, which satisfied the principles of overall optimum, fair and reasonable for ground delay flights.
Key Words: air traffic flow management; slot allocation; ground delay

Introduction

Air traffic congestion is the most critical issue that Chinese Air Traffic Management (ATM) has to face. Capacity of airports and airspace far exceeds the constant growth in air traffic demand, resulting in ever-increasing flight delays. The time and management cost troubles both airlines and passengers.

As the core theory of air traffic management system, the system operation and management algorithms can analyze air traffic situations, evaluate and predict the capacity of air traffic systems, and put forward rational and practical flow management suggestions. The operation and management algorithms are the critical factors to ensure practical value of the system.

Considering lots of constraints of airspaces and airports, the specialists allocate the rational take-off slots for flights, which play a great role in easing the congestion in airspaces or airports.

Researchers have studied the problem of slot allocation for the flights from different points of view, and different mathematical methods have been applied to establish a number of mathematical models. Domestic scholars have made a more complete review on this issue (Traffic Management Project Study Group 1999).

The core algorithms of the current resource allocation for flights used by the Enhanced Ground Delay Program (GDPE) of American are Ration-by-Schedule (RBS) and Compression algorithms. When ground delay occurs, the RBS algorithm at first allocates the limited time slot to the affected airlines, which is the initial distribution of resources, then re-adjusts the time slot allocation resources based on the latest information of flight delays and cancellations provided by airlines,. This is a re-distribution of resources (Dong 2007); Ball et al. 2000).

Computer Assisted Slot Allocation (CASA) is the algorithm used by the Europe air traffic flow management system, which is a heuristics method that allocates slots in accordance with the principle of FCFS (First Come First Serve) priority (Traffic Management Project Study Group 1999).

The Expected Departure Clearance Time(EDCT)Program that is used by Japan is the main active traffic management strategy and is similar to GDP. This program uses linear programming to optimize the departure times for flights in the regional airport under flow control (Zhang et al. 2004).

Because of the complexity of the air traffic flow management problem, it is very difficult to draw conclusions in line with actual requirements by relying only on mathematical methods. For this reason, air traffic flow management systems in Europe and the United States both added in heuristic method, which is effective for specific conditions of air traffic flow management problems. Our air traffic environment and air traffic control procedures are different, so it is difficult to use the algorithm of air traffic flow management and expert database directly. It is necessary to put forward the method taking into account the actual conditions in our country.

Based on the analysis of the slot allocation algorithms abroad, considering the present situation of national air traffic flow management, we have put forward the slot allocation algorithm suitable for China. Based on statistics and predictions for the traffic flow, capacity, the timely detection of emerging and potential traffic problems, it can propose a variety of programs and the corresponding effect to solve the traffic flow management problems and support them to make a final decision. Particularly, the ground delay program uses the slot allocation algorithm to allocate the overall optimum, fair and impartial manner take-off slots for the flights required ground delay.

Slot Allocation Problem Description

Air traffic flow management (ATFM) is a daily pre-tactical filter intended to regulate

scheduled flights across controlled airspace or arrive at the congested airport. Its aim is to limit the number of flights in a given airspace or airport over a given period. This planning will be precisely scheduled by the air-traffic controllers in real time.

The ATFM problem is described in terms of:

- *Flights*: A flight starts from one airport at a specified time, follows a predefined route at a fixed speed, and arrives at another airport.
- *Sector*: The airspace is divided into sectors crossed by the routes of the flights. A sector is a 3D polyhedra, usually a vertical cylinder endowed with a capacity expressed as the maximum number of flights entering the sector during a time period. The partition of the airspace, i.e., the number and shape of the sectors, varies during the day. The capacity of the same sector may also change at given times. We call sector-period a sector for particular a period of time and capacity.

The constraints of the problem are the capacities of the sectors and airports. There are several degrees of freedom to satisfy these constraints. In this paper, we focus only on the ground-holding policy: each flight may be delayed at departure. ATFM is in charge of solving the problem on a day-by-day basis. The difficulty of the problem does not come from the complexity of its constraints but from its size.

The objective of the slot allocation problem is to reduce the delays. There are several ways to achieve this purpose: reducing the total sum of the delays (utility), the maximum delay (equity), the average delay, etc. In this paper, we are more interested in the implementation of the solution than the cost function.

Mathematical Models

To achieve the allocation of flight slots, the following mathematical models were established.

For a single flight,

$$T_{SCH} = T_{IGT} - TAXI \quad (1)$$

where $T_{SCH=}$Schedule arrival time, $T_{IGT=}$Schedule block in time, TAXI=Taxi time.

The set of the influenced flights is

$$F = \{F|F_{ARRD}=D\} \cap ((\{F|A<T_{SCH}<B\} \cup \{F|A<T_{ETA}<B\}) - (\{F|A<F_{SCH}<B\} \cap \{F|F_{ETA}<A\})) \quad (2)$$

where F=The set of the influenced flights, D = Destination airspace/airport, F_{ARRD}=Destination airspace/airport of the flight, A = Begin time, B=End time, F_{ETA} = Estimated arrival time of the flight.

The set of the active flights is

$$F_{Active} = \{F|T_{ETD}<(C+H)\} \cup F_{InFl} \quad (3)$$

where F_{Active}=The set of the active flights, F_{ETD}=Estimated departure time of the flight C=Issue time, H=Exempt departure time, F_{InFl}=Flying flights.

The set of the flights departure from exempt airports is

$$F_3=\{F|F_{OriAirport}\in EmptOriApt\} \quad (4)$$

where F_3=The set of the flights departure from exempt airports, $F_{OriAirport}$=Original airport of the flight, EmptOriApt=Exempt airports.

The set of the exempt flights is

$$F_1=F_2+F_{Active}+F_3+F_{international} \quad (5)$$

where F_1=The set of exempt flights, F_2=Exempt flights inputted, $F_{international}$=International flights.

The intended departure time of the flight is

$$T_{INTD}=T_{INTA}-T_{ETE} \quad (6)$$

where T_{INTD}=Intended departure time, T_{INTA}=Intended arrival time, T_{ETE}=Estimated time of en route.

Slot Allocation Algorithm

Principle of the Algorithm. The "ground-holding" policies base on the fundamental fact that airborne delays cost much more than ground delays, because the former include fuel, maintenance, depreciation, and safety costs. Thus, the motivation underlying ground-holding policies is that one may hold an aircraft on the ground before take-off so that, when the aircraft arrives at its destination, it will not have to wait in the air before landing.

Assuming infinite departure capacities eliminates airborne delays, and the pure 0-1 integer programming formulation of the static deterministic multi-airport GHP can be determined (Vranas et al. 1992).

Heuristic Algorithm

Our algorithm has the characteristic such as improving the priority partition method, rules for slot allocation algorithm, distribution rules for each 15-minute capacity when hour capacity is m, and generating rules for the slot value sequence under different capacity values.

The procedure of the algorithm is: Firstly, determine which flights should consider the airport congestion ground delay program. Secondly, create arrival slots. Thirdly, assign flights to the arrival slots based on each flight's original scheduled time of arrival. At last, calculate each flight's controlled time of departure (CTD) by subtracting the estimated time en route (ETE) from the controlled time of arrival (CTA).

Figure 1 shows the flow chart of the slot allocation method we designed for the flights.

The procedure is specifically described as follows:

(1) Read complete control parameters.
Obtain the name of the control element, the code of the element, the begin time, the end time, the exempt take-off time, the impacting airports, the exempt airports, and the exempt flights.
(2) Filter out the affected flights according to Eq. 2.
(3) Filter flights by exempt status.
Get international flights from the flight plan or 4D track, according to Eq. 3 calculate the active flights, according to Eq. 4 calculate the exempt airport flights, and according to Eq. 5 calculate the exempt flights.
(4) Create available slot resources.
Algorithm generates arrival slots for each control time period. If the hour capacity is m, the distribution rules of each 15-minute capacity are showed in table 1. Generating rules of the slot value sequence under different capacity values are shown in Table 2, and if it is more than 15 we deal with it as 15. If the capacity is 0, it does not generate a slot. Eventually the number of slots is exactly the same to the capacity value that the traffic management center specified.
(5) According to the flight status, we set five groups of flights with different priorities.
The first group: priority flights. The second group: the flights had been controlled. The third group: the successive flights without delay (delay criterion: take-off time is at least 15 minutes greater than the original departure time). The fourth group: no priority and non-controlled flights. The fifth group: the delayed successive flights. The first group flights have the highest priority, and the fifth group flights have lowest priority.
If the specialists can collaborate with the airlines, the flights can be sorted one after another according to the schedule time of arrival. If the specialists can not collaborate with the airlines, the flights successively sequenced according to the estimated time of arrival. Estimated time of arrival is calculated by the latest arrival time of the 4D flight path: if flight track can be obtained, we can get estimated time of arrival by flight track; if flight plan can be obtained without flight track, we can get estimated time of arrival by flight plan; if schedule can be obtained without flight plan, we can get estimated time of arrival by schedule.
(6) Allocate the take-off slots based on the slot allocation algorithm.
The algorithm allocates slots for the five groups of flights with different priorities. The first group flights were given the highest priority in the allocation of take-off slots, and it makes direct use of the time slot which estimated time of arrival occupied as much as possible. The second group flights have second priority, making use of controlled slots directly as much as possible. The third group flights have third priority to make full use of the slots they occupied. The last two group flights will be the final allocated slots from the remaining slots. Slots are arranged in order to generate the corresponding controlled time of arrival. Each of the flights uses the slot where its estimated time of arrival is located as much as possible. A flow chart of the slot allocation method is shown in Figure 1, and the specific processes can be described as follows:

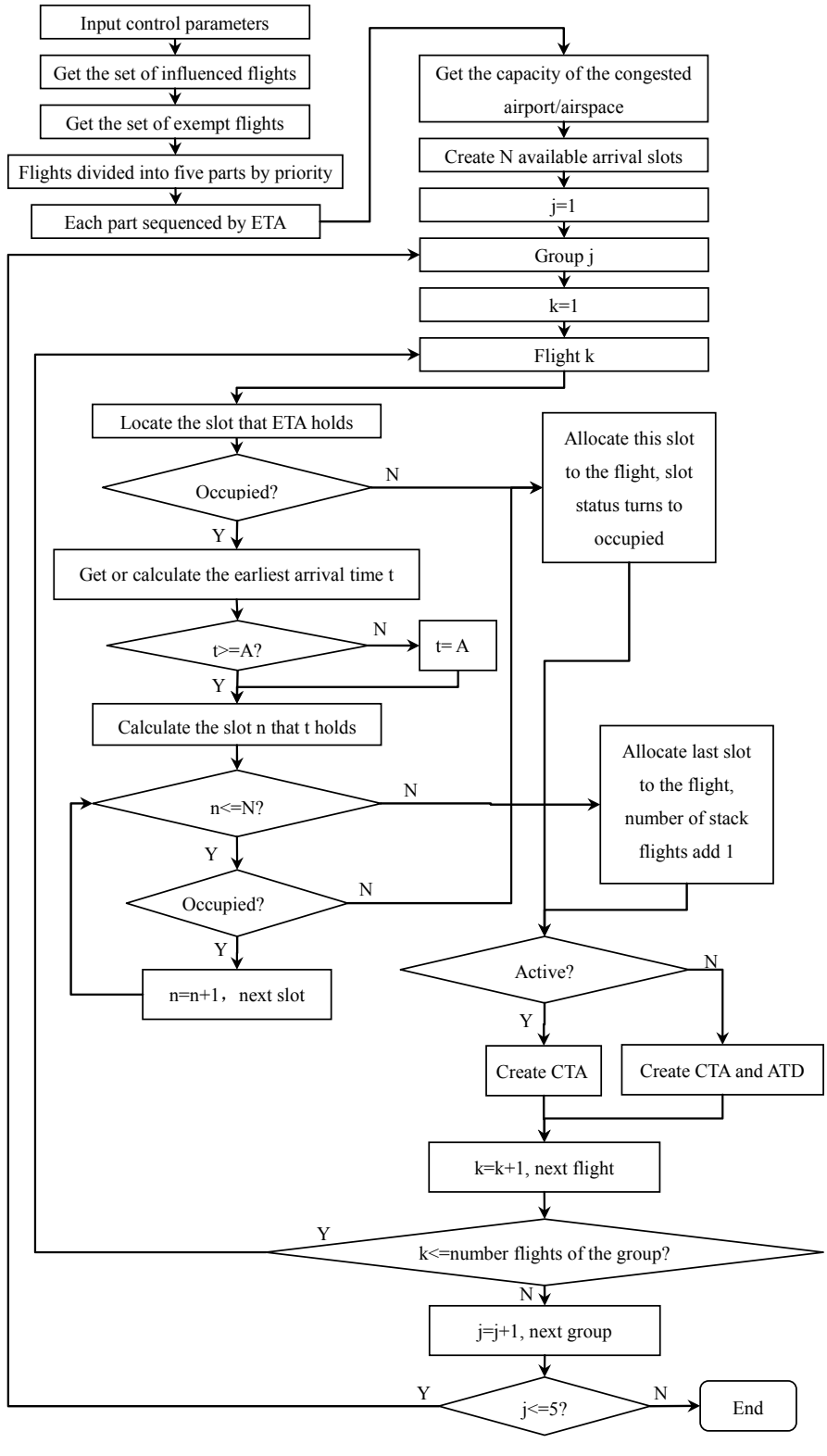


Figure 1. Flow chart of the slot allocation method.

Table 1. Distribution rules of each 15-minute capacity when hour capacity is m.

Capacity value is m(/60min)	First 15 min capacity value	Second 15 min capacity value	Third 15 min capacity value	Fourth 15 min capacity value
Can be divisible by 4	m/4	m/4	m/4	m/4
Divided by 4 remainder is 1	$\lfloor m/4 \rfloor$	$\lfloor m/4 \rfloor$	$\lfloor m/4 \rfloor$	$\lfloor m/4 \rfloor+1$
Divided by 4 remainder is 2	$\lfloor m/4 \rfloor$	$\lfloor m/4 \rfloor+1$	$\lfloor m/4 \rfloor$	$\lfloor m/4 \rfloor+1$
Divided by 4 remainder is 3	$\lfloor m/4 \rfloor$	$\lfloor m/4 \rfloor+1$	$\lfloor m/4 \rfloor+1$	$\lfloor m/4 \rfloor+1$

Table 2. Generating rules of slot value sequence under different capacity values.

Capacity value (/15min)	Number of slots	Slot value sequence	Example
1	1	15	10:00-10:14(15 min)
2	2	7-8	10:00-10:06(7 min), 10:07-10:14(8 min)
3	3	5-5-5	10:00-10:04, 10:05-10:09, 10:10-10:14
4	4	3-4-4-4	
5	5	3-3-3-3-3	
6	6	2-3-2-3-2-3	
7	7	2-2-2-2-2-2-3	
8	8	1-2-2-2-2-2-2-2	
9	9	1-2-1-2-1-2-2-2-2	
10	10	1-2-1-2-1-2-1-2-1-2	
11	11	1-2-1-1-2-1-1-2-1-1-2	
12	12	1-2-1-1-1-1-2-1-1-1-1-2	
13	13	1-1-1-1-1-2-1-1-1-1-1-1-2	
14	14	1-1-1-1-1-1-1-1-1-1-1-1-1-1	
15	15	1-1-1-1-1-1-1-1-1-1-1-1-1-1-1	

Firstly, determine whether the estimated arrival time slot is occupied or not, if not occupied, allocate the time slot to the flight, and the time slot status becomes occupied, so slot allocation for this flight has ended. If occupied, obtain or calculate the earliest arrival time of the flight. If the airline provides the earliest arrival time, then access this time, if the airline did not give the earliest arrival time, the traffic management specialist will assign a first arrival time to the flight. The earliest time of arrival can be obtained by estimating time of arrival and subtract 20 minutes. If the earliest arrival time is less than the control beginning time A, then the earliest time of arrival equals A. Calculate the slot n that the earliest arrival time located in, if

n<=N determine whether the time slot n is occupied or not. If not occupied, allocate the time slot to the flight, and if the time slot status becomes occupied, slot allocation for this flight has ended. If occupied, navigate to the next time slot n+1, then judge and recycle. If n>N, it means there is no right time slot that could assign to the current flight, then the last time slot is forced to be allocated to the flight. The beginning of the last time slot has the intention of the flight arrival time, at the same time the cumulative number of stack flights adds one, and slot allocation for this flight has ended.

The intended arrival time of the flight is equal to the beginning value of slot times, for example, for the 14:24-14:27 time slot, the intended arrival time is 14:24.

Active state refers to cruising or the flight is about to take-off time shortly after the measure is issued. If a flight is active, only intentional arrival time for the flight is generated. If the flight is not active, the intentional arrival time for the flight is changed to the beginning time of the slot allocated. The intentional take-off time for the flight can be calculated by Eq. 6. Flying time is estimated by the latest calculated 4D Track.

Experiment Results

Use the July 10, 2009, Shanghai Hongqiao Airport arrival flight flow as study data, the beginning time is 9:00, the end time is 12:00, results for various slot values were shown in Table 3. From Table 3 we can find that the fewer stack flights, the more efficient the algorithm we used, and the algorithm is optimal when the stack flights are few or none.

Comparison of experiment results between GSHF (Group Slot Hold First) and ASHF (All Slot Hold First) were shown in Table 4. From Table 4 we can find that GSHF algorithm optimized much better than the ASHF algorithm in total delay and maximum delay when there are lots of stack flights.

Table 3. Results for various slot values.

value Slot / Index	S=10	S=6	S=5
900-1000(min/each)	60	60	60
1000-1100(min/each)	30	20	12
1100-1200(min/each)	no limit	30	30
Number of flights without slot	22	10	4
Total delay of controlled flights(min)	440	1077	485
Average delay of controlled flights(min)	24.4	35.9	13.5
Maximum delay of controlled flights (min)	250	104	65

Table 4. Comparisons of experiment results.

Algorithms	Capacity value	Stack	Total delay	Average delay	Maximum delay
GSHF	5	25	2477	165	180
	7	19	1946	93	152
	9	13	1388	51	115
	10	10	1079	36	115
	12	4	498	14	115
	14	0	141	4	25
ASHF	5	25	2480	165	180
	7	19	1957	93	175
	9	13	1391	52	125
	10	10	1087	36	115
	12	4	498	14	115
	14	0	141	4	26

Conclusions

The new algorithm we designed has the following features:

(1) Improved the priority partition method. RBS algorithm for the United States considered the issue of priority of the flight, but the flights are only divided into three groups: exempt flight, controlled flights and other flights. Through the analysis we find that there are a large number of successive domestic flights, which have their particularities. Flight delays due to successive flights occur from time to time, and it will expand caused by the successive flights delay. Because there are lots of successive flights in China, we added two new classes in the division of flights called the successive flight, one class with delays and the other without delays.

If the successive flights without delays were given delays, they will have to obtain greater losses, so the successive flights without delays have the priority higher than normal flights. However, when allocating the take-off slots for the successive flights with delays, the aircraft can not arrive in time due to the delay, therefore, the delayed successive flight must be given the lowest priority. We do not delay the successive flight without delay by raise the priority, in order to encourage airlines to reduce successive flight delays.

(2) Rules for slot allocation algorithm. As can be seen through the flow chart, the algorithm we have designed firstly takes the time slot in the priority group into account, followed by the distribution of the order in accordance with the time slot, also with the order of distribution of such flights within each group. When flights are not able to use the time slot occupied, we consider the possibility of using the slot earliest arrival time located for the current flight, followed in turn by the search method, which not only takes into account the priority of flights, but also takes into account making full use of the time slots and improves the efficiency of slot

utilization.

(3) Distribution rules of each 15-minute capacity when hour capacity is m. Hour capacities are transferred into 15-minute capacities unified. The rules give a good solution to the challenge that the capacity of the hour is not able to be allocated to each 15 minutes averagely.

(4) Generating rules of the slot value sequence under different capacity values
The rules give a good solution to the challenge that the capacity is not divisible in the course of the slot value generated by the capacity.

Acknowledgments

This work was supported by Hi-tech Research and Development Program of China (863 Program 2006AA12A105) and Key Project of the Tianjin Science & Technology Pillar Program (08ZCKFGX04100).

References

Ball, M. O., Hoffman, Chen, C. Y., and Vossen, T. (2000). "Collaborative decision making in air traffic management." *The National Center of Excellent Aviation Operation Research* (*NEXTOR*) T.R.

Dong, Y. (2007). "Air traffic flow management algorithm for equitable distribution of resources." *Nanjing University of Aeronautics and Astronautics*.

Traffic Management Project Study Group. (1999). "Study on air traffic flow management model and algorithm." *Proceeding of Air Traffic Flow Management Symposium.*

Vranas, P. B., Dimitris, J. B., Odoni, A. R. (1992). "The multi-airport ground-holding problem in air traffic contro." *OR* 263-92, 1992.

Zhang, T., Zhan, J., and Zhao, Y. (2004). "Japanese air traffic flow management system inspection report." *Tianjin: Civil Aviation University of China*.

GNSS Evolutionary Architecture for Civil Aviation Utilization

Qian Sun[1], Jun Zhang[2] and Yanbo Zhu[3]
[1] Doctoral Student, [2] Prof., [3] Ph.D, School of Electronics and Information Engineering, Beihang University, Beijing 100191, China, Email: sunqian@ee.buaa.edu.cn

Abstract

The continuing growth of aviation places increased demands on airspace capacity and emphasizes the need for the optimum utilization of the available airspace, which demands the navigation system being especially reliable after an aircraft approach has commenced. Navigation technology development in GPS from USA, Compass in China, and Galileo in Europe provides the foundation for aviation's global evolution. Area navigation (RNAV) is a method of navigation that permits aircraft operation on any desired flight path and Required Navigation Performance (RNP) is a statement of the navigation performance accuracy necessary for operation within a defined airspace. All elements of communication, navigation, surveillance (CNS), and air traffic management (ATM) are combined into a CNS/ATM-based system in which RNP/RNAV provides the path for this transition. On the other hand, GNSS evolutionary architecture (GEA) and updated requirements need to be studied. In this paper, analysis and detailed differences between RNAV/RNP toward navigation specifications are first carried out, then key technologies in GEA study panel (GEAS) including three candidates such as GIC, RRAIM, and ARAIM are studied; at last corresponding algorithm and basic architecture are proposed to test the capability of ARAIM implementation with frequency diversity. The idea discussed in this paper is suitable for the development of satellite navigation utilization in civil aviation and helps determine future air traffic management policy.
Key Words: air navigation service (ANS); RNP/RNAV; GNSS architecture; integrity monitoring

Introduction

The global positioning system was developed by the U.S. Department of Defense (DOD) to provide precise estimates of position, velocity, and time to users worldwide (Parkinson and Spilker 1996; Kaplan and Hegarty 2006). The DOD

approved the basic architecture of the system in 1973, and the system was declared fully operational in 1995. The primary application of GPS for civil aviation is as a navigation sensor in instrument meteorological conditions for all phases of flight: departure, en-route, non-precision approach, and precision approach. Today, GPS capability serves more than 200 million users with a variety of applications, of which a modest population of approximately 200,000 GPS users require aviation integrity (Walter et al. 2008; Hegarty and Chatre 2008). However, this aviation number grows steadily since every new Airbus or Boeing aircraft is outfitted with the GPS receiver or multi-mode GNSS receiver, and most of new general aviation aircraft also carry GNSS (ARINC 2005). The United States has planned to modernize GPS including diversity of frequency civil signals available on future satellites system (U.S. Air Force 2005; Hegarty and Chatre 2008). In addition, Europe has plans to implement the new GALILEO system, which will also have multiple-frequency civil signals. Due to the leverage of frequency diversity and over-specified solutions brought by redundant enough satellite numbers, it may be possible to extend global navigation satellite system (GNSS) to be sole navigation mean during aviation phases by achieving instrument approaches with vertical guidance, which requires significantly smaller alert limits.

Rapid changes in technology in the area of navigation performance, including the change from source-referenced area navigation systems, provide the foundation for aviation's global evolution to satisfy the continuing growth of aviation demands on airspace capacity and the need for the optimum utilization of the available airspace. Navigation systems must be especially reliable after an aircraft approach has commenced and must be available greater than 99% of the time regardless of the weather and specific scenario. The development will be marked by combining all elements of communication, navigation, surveillance (CNS), and air traffic management (ATM) into tomorrow's CNS/ATM based systems (Whelan 2001). Concepts in CNS/ATM such as RNP provide the path for this transition (ICAO, 2008). In order to promote satellite civil utilization, the specifications of LNAV(lateral navigation) and VNAV (vertical navigation) have been studied and identified by related technique special committee and organization (Cabler and DeCleene 2002) and the GNSS evolutionary architecture study panel was formed by the FAA in October 2006 with the objective of evaluating future GNSS-based architectures to provide robust worldwide instrument approach guidance known as LPV-200 in the 2025 to 2030 time frame (GNSS Evolutionary Architecture Study Phase I 2008). GEAS includes three candidates such as GNSS integrity channel (GIC), relative receiver autonomous integrity monitoring (RRAIM) (van Graas and Soloviev 2007), and absolute receiver autonomous integrity monitoring (ARAIM).
Though many countries have developed their own navigation satellites such as GLONASS of Russia, Galileo from Europe, and Compass in China, GPS is the sole open satellite navigation system for civil use worldwide. In some mountainous areas or plateau regions in China, specific route aircrafts have equipped RNAV procedures to take advantage of GPS to improve capacity of transportation and safety due to the limitation of ground based navigation facilities such as VOR and DME. Some

researchers are studying similar areas (Chen et al. 2001; Guo et al. 2007; Xue et al. 2009) in China to ensure utilization integrity in civil aviation.

In this paper, the concept of PBN proposed by ICAO (international civil aviation organization) toward performance based on navigation is introduced to specify the improved operational efficiency derived from the application of area navigation (RNAV) techniques referring to the terms of accuracy, integrity, availability, continuity, and functionality. Navigation specifications that require on-board performance monitoring and alerting are termed RNP specifications. Navigation specifications are used by states as the basis for certification and operational approval. Therefore, in the context of a particular airspace, the performance requirements are identified in navigation specifications supported by the appropriate navigation infrastructure such as navigation sensors and augmentation equipments called Navaid. Detailed analysis and differences between two modes are carried out. On the other hand, detailed description of requirements and analysis on three candidates of GEAS are carried out. In addition, the feasibility is primarily proven through basic algorithms and procedures. At last various validated and refined algorithms toward three candidates of the GEAS is depicted and analyzed. Several key numerical parameters are evaluated in order to support the GEAS activity. Corresponding simulation and basic architecture are carried out to test the effect of implementation.

In all, the idea discussed in this paper is suitable for the development of satellite navigation utilization in civil aviation and helps determine future air traffic management policy.

RNP/RNAV for Civil Aviation

Area navigation (RNAV) capability is an existing aircraft capability that supports airspace development in the form of improved design of procedures and more efficient operations. Required Navigation Performance (RNP) is an emerging tool for the development of more efficient airspace and operations. An aircraft area navigation system developed for RNP operations provides reliable, repeatable, and predictable performance through specific RNP RNAV capabilities and features. RNP RNAV provides a means to meet the requirements of RNP airspace and operations, along with Vertical Navigation (VNAV) enabling further improvements in RNP airspace design and operations. Requirements for navigation applications on specific routes or within a specific airspace must be defined in a clear and concise manner. This is to ensure that the flight crew and the air traffic controllers (ATCs) are aware of the on-board RNAV system capabilities in order to determine if the performance of the RNAV system is appropriate for the specific airspace requirements. Here, emphasis on GNSS in RNP RNAV will be discussed and analyzed.

RNP/RNAV

Concept. RNAV is defined as "a method of navigation that permits aircraft operations on any desired course within the coverage of station referenced navigation

signals or within the limits of a self-contained system capability or combination of these (ICAO 2008)."

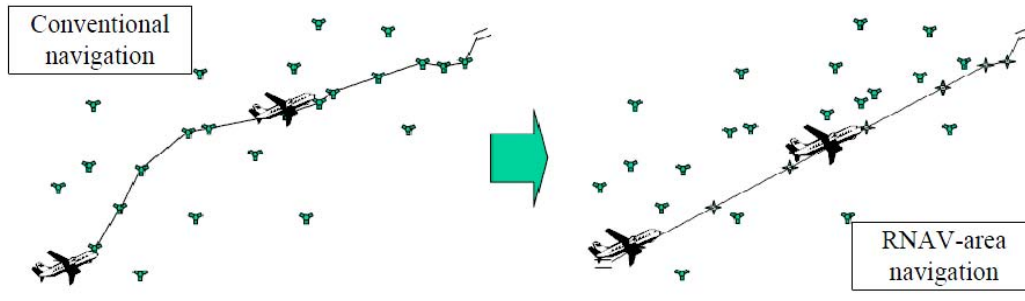


Figure 1. RNAV concept.

RNAV systems are recognized for their horizontal 2D capability to utilize one or more navigation sensor sources to determine the aircraft position, compute flight paths referenced to navigation aids or points defined by latitude and longitude, and provide guidance cues or tracking of the flight path. In addition to this capability, many RNAV systems include a 3D capability to define vertical path profiles based upon altimetry and a built-in model of the aircraft and engine performance, and provide guidance cues or tracking of this vertical path. An added feature to RNAV systems has been 4D, in the form of determining and managing time of arrival to a specific point along the flight profile (Ober et al. 1998).

It may determine position using any navigation aids including GNSS. The RNAV system uses an on-board navigation database or information, which prestored information and data such as for airports, navigation aids, departure/arrival procedures, routes/airways, altitudes, speed, frequencies, elevation, distance, and magnetic variation, to facilitate flight crew definition of the flight path for the operation. The RNAV systems provide information such as path deviation (e.g., CDI) and speed commands, data interfaces for guidance cues such as flight director commands, and graphical map displays to enable the flight crew or system to follow and track the RNAV flight path.

Required Navigation Performance (RNP) (Meyer and Bradley 2001) is a statement of the navigation performance accuracy necessary for operation within a defined airspace (ICAO Document 9613, Manual on Required Navigation Performance defines the RNP concept). RNP can include both performance and functional requirements and is indicated by the RNP type, which was developed to allow the specification of airspace and operation requirements without the constraints of the slow process for specifying equipment and systems. By making RNP non-specific for equipment, the airspace designer was freed from equipment mandate concerns to apply RNP in specifying improved airspace or procedures that all users and their aircraft equipment would meet. RNP could then be applied as one consideration in requirements such as route widths and traffic separation. RNP is defined as a 95% containment value and its relationship to navigation performance accuracy left the

other aspects of integrity, availability, coverage, etc. to other technical bodies to specify.

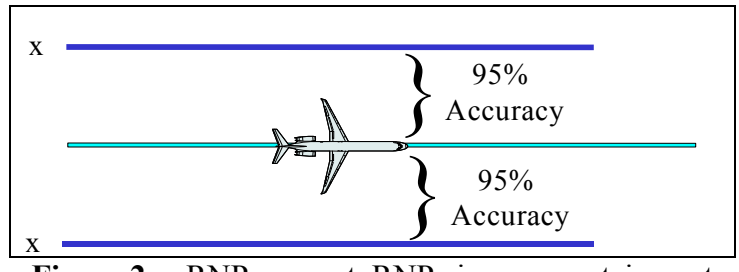


Figure 2. RNP concept. RNP airspace containment

These descriptions address various aircraft characteristics that were causing variations in flight trajectories, leading to more repeatable, reliable, and predictable flight tracking, as well as smaller obstacle assessment areas.

RNP/RNAV

Specifications. A navigation specification is either an RNP specification or an RNAV specification and the basic differences between RNP and RNAV are that an RNP specification includes a requirement for on-board self-contained performance monitoring and alerting, while an RNAV specification does not and instead, accuracy is ensured operationally by qualifying specific sensors or through ATM. Many RNAV systems, while offering very high accuracy and possessing many of the functions provided by RNP systems, are not able to provide assurance of their performance. Recognizing this, and to avoid operators incurring unnecessary expense, where the airspace requirement does not necessitate the use of an RNP system, many new as well as existing navigation requirements will continue to specify RNAV rather than RNP systems. It is, therefore, expected that RNAV and RNP operations will co-exist for many years. The required performance is an accuracy value met by using an appropriate RNAV system, which is that the total system error (TSE) does not exceed the specified RNP value ('x') for 95% of the flight time in either cross-track or along track. The containment region (see Figure 2) quantifies the navigation performance where the probability of an annunciated deviation greater than 2 x RNP is less than 1×10^{-5}.
The basis for user confidence is integrity, and RNP containment limit represents a 99.999% or better probability per flight hour level of assurance that misleading RNP status has not occurred. Continuity is the capability of the total system to satisfy the containment integrity requirement without unscheduled interruption during the intended operation, whose limit is a 99.99% or better probability that the navigation performance is satisfied without unscheduled interruptions. Continuity helps prevent unnecessary interruptions and false warnings of a loss of navigation capability (RTCA DO-236B, 2003).

Different RNP/RNAV types (shown in Table 1) are defined for specific routes during aviation and aircrafts with the procedures specified by FAA and operated worldwide

(FAA AC 90-96A, 2005 and FAA AC 90-100A, 2007). Given certain types of RNP/RNAV equipment, pilot or flight crew utilize default types dependent upon phase of flight. In the United States, allowing Wide Area Augmentation System

Table 1. Standard RNP/RNAV types.

RNP RNAV Type	Applicability/ Typical Operation	Normal Performance 95% Accuracy	Airborne Containment Region	Hazard Class (Misleading Information/Loss of Navigation)2, 3
RNP – 2 RNAV	En-route	2 NM	+/- 4 NM	Major / Minor
RNP – 1 RNAV1	Terminal Area	1 NM	+/- 2 NM	Major / Minor
RNP - 0.3 RNAV	Approach	0.3 NM	+/- 0.6 NM	Major / Minor

(WAAS) to satisfy the RNP/RNAV requirements without having to impose unique operational issues on the equipment and users of that equipment stand-alone navigation units (TSO-C146) can support operations for standard RNP RNAV types of 2, 1, and 0.3.

Specification for GNSS on RNP/RNAV. Many RNP/RNAV procedures have been developed or are planned worldwide with GNSS as one enabling technology. GNSS offers many benefits over traditional navigation aids including the ability to fly arbitrary routes rather than being constrained by the location of ground navigation facilities. Provisions of improved navigation services in areas are not presently covered by ground navigation aids, therefore, alleviate some of the expense of maintaining expansive networks of ground navigation facilities. Performance requirements for navigation sensors generally fall within four categories: accuracy, integrity, continuity, and availability. Accuracy is the degree of conformance between true aircraft position and that position estimate provided by the navigation sensor. Integrity is the ability of a navigation system to provide timely warnings when the system cannot be safely used for navigation. Continuity is the capability of a navigation system to perform its function without unscheduled interruptions during the intended operation. Availability is the fraction of time during which a system is usable to perform an intended operation.

Integrity Monitoring. Integrity is the most important affairs related to safety-critical during flight which are commonly specified using three parameters: 1) an alert limit: An alert limit is the maximum allowable navigation system position error before safety would be unacceptably compromised if the user is not promptly notified; 2) time-to-alert: The time-to-alert is the maximum allowable period from the onset of an out-of-tolerance condition until an alert is provided or probability of hazardously misleading information; and 3) integrity level: The integrity level or probability of

hazardously misleading information is the maximum acceptable probability of occurrence of an out-of-tolerance condition without a timely alert.

Table 2. GNSS signal-in-space performance requirements.

Operation	accuracy (95%) Horizontal	vertical	Integrity level	continuity	Availability
En-route	3.7 km	N/A	$1-1\times10^{-7}$ / h	$1-1\times10^{-4}$ - $1-1\times10^{-8}$/h	0.99 - 0.99999
Terminal	0.74 km	N/A	$1-1\times10^{-7}$ / h	$1-1\times10^{-4}$ - $1-1\times10^{-8}$/h	0.99 - 0.99999
NPA	220 m	N/A	$1-1\times10^{-7}$ / h	$1-1\times10^{-4}$ - $1-1\times10^{-8}$/h	0.99 - 0.99999
APV I	16 m	20 m	$1-2\times10^{-7}$ / approach	$1-8\times10^{-6}$ / 15s	0.99 - 0.99999
APV II	16 m	8 m	$1-2\times10^{-7}$ / approach	$1-8\times10^{-6}$ / 15s	0.99 - 0.99999
CAT I	16 m	6~4 m	$1-2\times10^{-7}$ / approach	$1-8\times10^{-6}$ / 15s	0.99 - 0.99999

Table 3. GNSS integrity requirements.

	Alert Limit		Time to Alert
	Horizontal	Vertical	
En-route (Oceanic)	7.4 km	N/A	5 min
En-route (Local)	3.7 km	N/A	5 min
Terminal	1.85 km	N/A	15 s
NPA	556 m	N/A	10 s
APV I	40 m	50 m	10 s
APV II	40 m	20 m	6 s
CAT I	40 m	15 ~ 10 m	6 s

GEAS for LPV-200

The GNSS Evolutionary Architecture Study (GEAS) group explores the different possibilities for aircraft navigation utilizing the satellite signals available in the mid-term future with GPS. Therefore, it is assumed that civil dual frequency signals will be available. This eliminates the current dominant ranging error source: the ionospheric delay. The objectives of the GEAS are focused on potential system implementations providing worldwide coverage, initially looking to satisfy LPV-200 operations (shown in Figure 3.). In order realize the lower decision altitude, the navigation system must ensure integrity, in the other words, it must guard vigilantly against the possibility of hazardously misleading information (HMI) and be able to detect any threatening faults or rare-normal events within a few seconds (TTA). The risk of an integrity failure during the approach must be less than 10^{-7} to 10^{-9} per approach (depending on the minimum decision altitude).

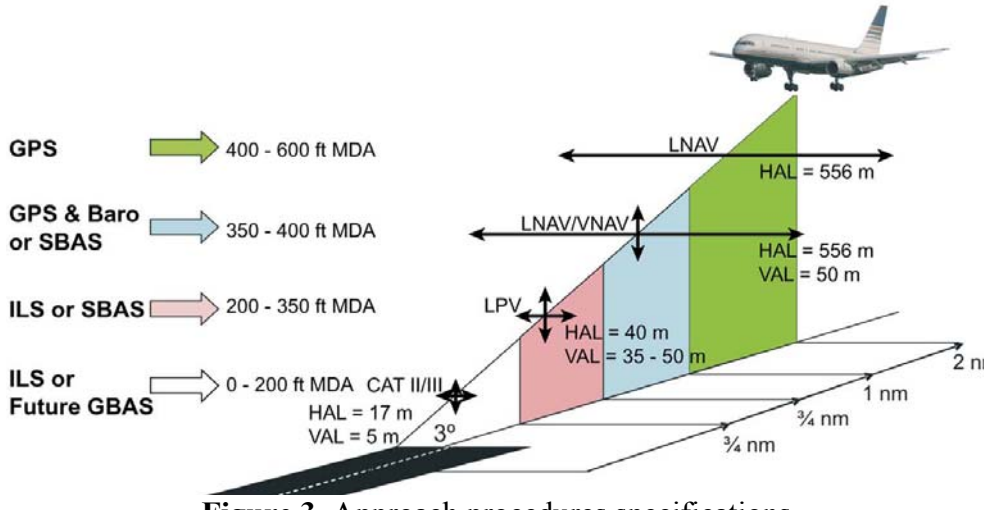


Figure 3. Approach procedures specifications.

Candidates Algorithms and Parameter Analysis. The panel identified three candidate architectures that employ a ground monitoring network that observes the satellite signals to identify and exclude faults (Walter et al. 2008). However, they place very different fractions of the integrity burden on the aircraft versus the aircraft-external monitors. In detail, they are: 1) Fast GNSS Integrity Channel (GIC) which places all of the responsibility for monitoring the signal in space outside of the aircraft; they are similar to SBAS in this respect; 2) Absolute Receiver Autonomous Integrity Monitoring (ARAIM), which places most of the responsibility on the avionics; they are similar to RAIM; and 3) Relative Receiver Autonomous Integrity Monitoring (RRAIM), which shares the integrity burden between the aircraft and the external monitors. Algorithms of three classes are being studied by many researchers now to make sure the transition was successfully achieved with the modernized GNSS (Lee and McLaughlin 2007).

The integrity requirement states that the probability of HMI must be kept below 2×10^{-7}per approach that historically was evenly split between horizontal (localizer) and vertical (glideslope); $1 \times 10^{-7}$for each of two independent systems. Since horizontal and vertical guidance are not independently derived for GNSS, we have been using 1×10^{-7}per approach as the target requirement. HMI is any time that the actual vertical position error (VPE) is greater than the dynamically calculated upper bound known as the vertical protection level (VPL) without a timely alert to the pilot. Further, the VPL must be below a static vertical alert limit (VAL) to ensure that the aircraft be kept safely away from obstacles. For LPV-200, the VAL is 35 m. The requirement that VPL be below the VAL is the dominant limitation to performance.

Accuracy requirement is known as the effective monitor threshold (EMT). The requirement is that a fault must be detected at least 50% of the time when an error is present that creates a vertical positioning error equal to the probability. The EMT is evaluated as 15 m.

Continuity requires that the above requirements be met continuously for the duration of the approach. Given that the above requirements are met at the initiation of the approach, the probability that one of them will be exceeded must be at or below 8×10^{-6} per 15 second interval (Walter 2008).

Availability is the fraction of time that all above requirements are met. For the system to be useful, it must be available at least 99% of the time at any location where LPV-200 service is authorized. For scheduled service, it may need to be available for even greater percentages of time (between 99.9% and 99.999%).

Candidates algorithm. Today's GPS has only one frequency that can be used in civil aircraft. Starting around 2009, new GPS satellites will begin to broadcast the signals in a third band with central frequency equal to 1176.45 MHZ called L5 (Van Dierendonck and Hegarty 2000). The integrity machines of the future will also draw great benefit from the frequency diversity provided by the new signals because it will obviate most of the ionospheric effects that trouble aviation use of GPS today. The L1 and L5 signals are combined in a way to eliminate the first order ionospheric delay, on which the algorithms of candidates are studied. Algorithms related to GIC, RRAIM, and ARAIM architectures all based on the frequency diversity are being studied by some labs and corporations. Conduct the VPL check since the accuracy models have not yet been fully verified, there are many further researches to do.

Considering the feasibility of the GEAS candidates, the following focuses the study about a RAIM algorithm taking advantage of frequency diversity based on ARAIM using HPL temporarily instead of VPL.

ARAIM Based on Dual Frequency Sources

Primary study has been carried out on RAIM method (Sun and Zhang 2009) related single fault with single frequency, on the basis of which a new algorithm is proposed here under dual frequency navigation sources.

The linearized GNSS measurements equation is given by:

$$y = H \cdot x + e \tag{1}$$

where y is the $n \times 1$vector of pseudorange measurement, H is the $n \times 4$ pseudorange observation matrix, and x is the 4×1 unknown state vector representing the true position including clock bias and ephemeris error, and e is the $n \times 1$ random noise. Since y is contaminated by an unknown random error, so Eq. 1 should be treated as a stochastic equation, and x can be determined using parameter estimation techniques. The least square estimate is used to obtain the unknown parameters. The LS solution of x is written as:

$$\hat{x} = (\mathrm{H}^T \mathrm{H})^{-1} \mathrm{H}^T \cdot y = A \cdot y \tag{2}$$

If another frequency is available, we can use them to get an equation and solve ionospheric delay. First, the delay can be written as:

$$40.29\mathrm{TEC}^i / f^2 = f^{-2} k^i \tag{3}$$

where TEC^i is the total electron content in the transmission path of ith SV, f is carrier frequency. The combined equation is written as:

$$\begin{cases} \mathbf{y}_{L1} = \mathbf{H}\mathbf{x} + f_{L1}^{-2}\mathbf{k} + \mathbf{e}_{L1} \\ \mathbf{y}_{L2} = \mathbf{H}\mathbf{x} + f_{L2}^{-2}\mathbf{k} + \mathbf{e}_{L2} \end{cases} \tag{4}$$

In order to obtain a general algorithm, here L1, L2 represent the two available civil GPS signals, which in the specific case L1 may be 1277.6 MHZ, and L2 equals to 1176.45 MHZ. $\mathrm{k}=[\mathrm{k}^1 \cdots \mathrm{k}^2 \cdots \mathrm{k}^n]$, k^i is defined above. $\mathbf{e}_{L1}$and $\mathbf{e}_{L2}$ is the noise in two frequency.

Because there is a parity matrix Q of H, we can get:

$$\mathbf{Q}\mathbf{H} = \mathbf{0} \ , \ \mathbf{Q}\mathbf{Q}^T = \mathbf{I}_{n-4} \tag{5}$$

Define parity vector $\mathbf{p}$,

$$\mathbf{p} = \mathbf{Q}\mathbf{y} = \mathbf{Q}\left(\mathbf{H}\mathbf{x} + \mathbf{e}\right) = \mathbf{Q}\mathbf{e} \tag{6}$$

We can get the statistic $\eta = \mathbf{p}^T\mathbf{p}$ which is the same as SSE (Sum of the Squares) derived from LSR (Least Square of Residuals), and it has the same feature in statistical distribution. The solution of P can be obtain as:

$$\mathbf{P}_{dual} = \sqrt{\sigma_{L1}^{-2} f_{L1}^{-4} + \sigma_{L2}^{-2} f_{L2}^{-4}}^{-1} \left| \sigma_{L2}^{-1} f_{L2}^{-2}\mathbf{P} \ -\sigma_{L1}^{-1} f_{L1}^{-2}\mathbf{P} \right| \tag{7}$$

Combined with the weighted idea to reduce the impact of elevation angle, the function can be normalized by its independent covariance matrix $\sigma_{L1}^2\mathbf{I}$, $\sigma_{L2}^2\mathbf{I}$:

$$\begin{cases} \sigma_{L1}^{-1}\mathbf{y}_{L1} = \sigma_{L1}^{-1}\mathbf{H}\mathbf{x} + \sigma_{L1}^{-1}\mathbf{e}_{L1} \\ \sigma_{L2}^{-1}\mathbf{y}_{L2} = \sigma_{L2}^{-1}\mathbf{H}\mathbf{x} + \sigma_{L2}^{-1}\mathbf{e}_{L2} \end{cases} \tag{8}$$

So the new linear measurement equation can be written as:

$$\mathbf{y}_{dual} = \begin{bmatrix} \sigma_{L1}^{-1}\mathbf{y}_{L1} \\ \sigma_{L2}^{-1}\mathbf{y}_{L2} \end{bmatrix} = \begin{bmatrix} \sigma_{L1}^{-1}\mathbf{H} \\ \sigma_{L2}^{-1}\mathbf{H} \end{bmatrix} \mathbf{x} + \begin{bmatrix} \sigma_{L1}^{-1}\mathbf{e}_{L1} \\ \sigma_{L2}^{-1}\mathbf{e}_{L2} \end{bmatrix} = \mathbf{H}_{dual}\mathbf{x} + \mathbf{e}_{dual} \tag{9}$$

The LS solution of x is written as:

$$\hat{\mathbf{x}} = \left(\mathbf{H}_{dual}^T\mathbf{H}_{dual}\right)^{-1}\mathbf{H}_{dual}^T\mathbf{y}_{dual} \tag{10}$$

Then,

$$\begin{aligned}\left(\mathbf{H}_{dual}^{T}\mathbf{H}_{dual}\right)^{-1} &= \left(\begin{bmatrix}\sigma_{L1}^{-1}\mathbf{H}^{T} & \sigma_{L2}^{-1}\mathbf{H}^{T}\end{bmatrix}\begin{bmatrix}\sigma_{L1}^{-1}\mathbf{H}\\ \sigma_{L2}^{-1}\mathbf{H}\end{bmatrix}\right)^{-1} \\ &= \left(\sigma_{L1}^{-2}+\sigma_{L2}^{-2}\right)^{-1}\left(\mathbf{H}^{T}\mathbf{H}\right)^{-1}\end{aligned} \tag{11}$$

So the expression can be as followed:

$$\hat{\mathbf{x}} = \left(\sigma_{L1}^{-2}+\sigma_{L2}^{-2}\right)^{-1}\left(\mathbf{H}^{T}\mathbf{H}\right)^{-1}\mathbf{H}^{T}\begin{bmatrix}\sigma_{L1}^{-1}\mathbf{I} & \sigma_{L2}^{-1}\mathbf{I}\end{bmatrix}\mathbf{y}_{dual} \tag{12}$$

Given the parity space matrix $\mathbf{Q}_{dual} = \begin{bmatrix}\mathbf{Q}_{L1} & \mathbf{Q}_{L2}\end{bmatrix}$ of H_{dual}, we can get:

$$\mathbf{Q}_{dual}\mathbf{H}_{dual} = \mathbf{0} \ , \ \mathbf{Q}_{dual}{\mathbf{Q}_{dual}}^{T} = \mathbf{I}_{n-4} \tag{13}$$

Due to

$$\sigma_{L1}^{-1}\mathbf{Q}_{L1}\mathbf{H} + \sigma_{L2}^{-1}\mathbf{Q}_{L2}\mathbf{H} = \mathbf{0} \ , \ \mathbf{Q}_{L1}\mathbf{Q}_{L1}^{T} + \mathbf{Q}_{L2}\mathbf{Q}_{L2}^{T} = \mathbf{I}_{n-4} \tag{14}$$

We can easily understand:

$$\begin{aligned}&\mathbf{Q}_{L2} = -\sigma_{L1}^{-1}\sigma_{L2}\mathbf{Q}_{L1} \\ &\left(1+\sigma_{L1}^{-2}\sigma_{L2}^{2}\right)\mathbf{Q}_{L1}\mathbf{Q}_{L1}^{T} = \mathbf{I}_{n-4}\end{aligned} \tag{15}$$

So,

$$\mathbf{Q}_{L1} = \frac{\sigma_{L2}^{-1}}{\sqrt{\sigma_{L1}^{-2}+\sigma_{L2}^{-2}}}\mathbf{Q}, \ \mathbf{Q}_{L2} = \frac{-\sigma_{L1}^{-1}}{\sqrt{\sigma_{L1}^{-2}+\sigma_{L2}^{-2}}}\mathbf{Q} \tag{16}$$

At last the matrix with weighted dual method can be written as:

$$\mathbf{p}_{dual} = \mathbf{Q}_{dual}\mathbf{y}_{dual} = \begin{bmatrix}\mathbf{Q}_{L1} & \mathbf{Q}_{L2}\end{bmatrix}\begin{bmatrix}\sigma_{L1}^{-1}\mathbf{y}_{L1}\\ \sigma_{L2}^{-1}\mathbf{y}_{L2}\end{bmatrix} = \frac{\sigma_{L1}^{-1}\sigma_{L2}^{-1}\mathbf{Q}}{\sqrt{\sigma_{L1}^{-2}+\sigma_{L2}^{-2}}}\left(\mathbf{y}_{L1}-\mathbf{y}_{L2}\right) \tag{17}$$

In order to verify the effect, a universal integrity parameter in RAIM should be solved as:

$$\begin{aligned}&R = y - H\cdot\hat{x} \\ &= [I - H(H^{T}H)^{-1}H^{T}]\cdot v = S\cdot v\end{aligned} \tag{18}$$

Then according to the HPL

$$K_{SLOPE,i} = \sqrt{(A_{1i}^{2}+A_{2i}^{2})(n-4)/S_{ii}} \tag{19}$$

$K_{SLOPE,i}$ is the slope of horizontal positioning error and statistics only related constellation geometry strength of the ith SV. $A^2_{1i,}$ A^2_{2i} is the A(1,i) and A(2,i) unit in matrix A. S_{ii} is the ith factor in diagonal line of matrix S defined above.

Horizontal Protect Level of every sample moment can be calculated as:

$$HPL = K_{SLOPE,\max} \cdot \eta_{\min} \tag{20}$$

Compare this value with the constant HPL specified in historically authoritative documents, if it is greater than the constant value, the system would alert the pilot, meanwhile indicating that it is not safe to use GNSS as the sole navigation mean.

Simulations

Geometries are evaluated every minute for a full 24-hour period (1440 epochs) with 27 SVs in a specific location. Coverage is calculated as the fraction of the users that meet a 99.5% availability goal. The 99.5% availability goal was chosen as a trade between simulation time and expected fidelity of the models. Accurately determining higher availabilities often require modeling additional effects beyond geometry and require longer simulation runs. In general, the clock/ephemeris and variance values will be functions of the aircraft based on integrity monitoring and algorithms. For this analysis, a simpler estimate of performance is obtained by using constant values that are close to the expected values for well-observed regions. For this analysis, we will assume the following values:

$$\sigma_{clk-eph1,2} = 1.5m, \sigma_{L1,L2} = 15m$$

Positioning error is shown in Figure 4. The scale is bigger in single frequency than using dual frequency. It indicates the new frequency can improve the positioning performance and may be an expected improvement in integrity monitoring. Figure 5 shows the HPL in 24 hours, even though there is one moment or two that the value exceed the allowed limit the performance ensure the integrity requirements.

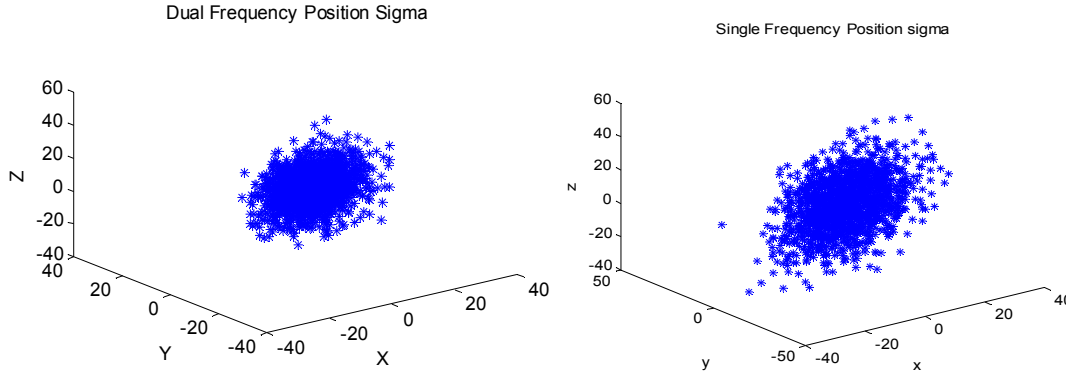


Figure 4. Position error in single and dual frequency.

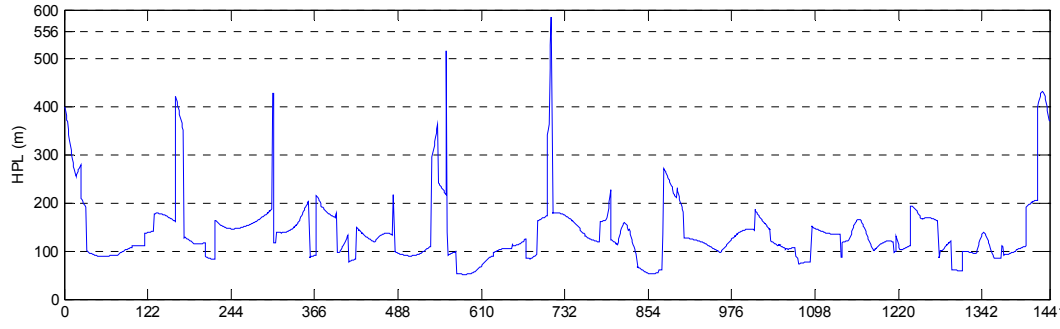


Figure 5. HPL in 24 hrs.

Conclusions

Detailed analysis and differences between RNAV/RNP toward navigation specifications are carried out and key technologies in GEA study panel (GEAS) including three candidates such as GNSS integrity channel (GIC), relative receiver autonomous integrity monitoring (RRAIM), and absolute receiver autonomous integrity monitoring (ARAIM) are studied. At last corresponding algorithm and basic architecture are proposed to test the capability of ARAIM implementation with frequency diversity. The idea discussed in this paper is suitable for the development of satellite navigation utilization in civil aviation and helps determine future air traffic management policy. Besides, the study on ARAIM including vertical navigation provides direction for further research.

Acknowledgments

The author would like to thank Dr. Yanbo Zhu, Dr. Miaoyan Zhang, and Prof. Jun Zhang at Behang University for their encouragement and support. The research is supported by the National Basic Research Program of China (973 Program, No. 2010CB731800), the National Natural Science Foundation of China (No. 60879012) and The National High Technology Research and Development Program of China (863 Program: No.2006AA12A101 and No.2006AA12Z312). The author wishes to acknowledge long time support from MST, NBRP, and NNSF.

References

ARINC (2005). "Multi-mode receiver (MMR)-Digital." *ARINC Characteristic* 755-3.

Cabler, H., and DeCleene, B. (2002). "LPV: new, improved WAAS instrument approach." *Proceedings of ION GPS 2002,* 1013-1021.

Chen, J. P., Xu, Q. F. and Liu, G. J. (2001). "Analysis of different algorithms for GPS RAIM HPL." *Journal of the Pla Institute of Surveying and Mapping*, 1-4.

FAA AC 90-96A (2005). "Approval of U.S. operators and aircraft to operate under instrument flight rules (IFR) in European airspace dDbasic area navigation (B-RNAV) and precision area navigation (P-RNAV)[S]." *U.S.A: Department of Transportation Federal Aviation Administration.*

FAA AC 90-100A (2007). "U.S. terminal and en-route area navigation (RNAV) operations[S]." *U.S.A: Department of Transportation Federal Aviation Administration.*

GNSS Evolutionary Architecture Study Phase I. (2008). "Panel report."

Guo, Y. T., Zhu, Y. B., and Huang, Z. H. G. (2007). "Study on key trajiectory prediction techniques of civil aviation aircraft." *Journal of Civil Aviation University of China,* 20-24.

Hegarty, C. J., and Chatre, E. (2008). "Evolution of the global navigation satellite system (GNSS)." *Proc. IEEE*, 1902-1917.

ICAO. (1996). "Annex 10 to the Convention of International Civil Aviation *Aeronautical Telecommunications*", *vol. I, Radio Navigation Aids* .

ICAO. (2007). "Annex 10 to the Convention of International Civil Aviation", *vol. I, Radio Navigation Aids, Amendment 82.*

ICAO (2008). "Performance-based navigation manual, 3rd ed." *ICAO DOC 9613.*

Kaplan, E., and Hegarty, C. (2006). "Understanding GPS: principles and applications, 2nd ed." *MA: Artech House.*

Lee, Y. C., and McLaughlin, M. P. (2007). "Feasibility analysis of RAIM to provide LPV-200 approaches with future GPS." *Proc. ION GNSS-2007,*2898-2910.

Meyer, T. and Bradley, J. (2001). "The evolution from area navigation (RNAV), required navigation performance (RNP), to RNP RNAV." *Global Navigatino Satellite System Panel Meeting.*

Nakamura, D. (2001). "General information on the functional & technical aspects of required navigation performance (RNP) area navigation (RNAV) and applications." *Boeing Airplane Documents.*

Ober, P. B., Theunissen, E., Meijer, R. C. and van Willigen, D. (1998). "RNP requirements for 4-D navigation." *ION GPS1998.*

Parkinson, B., and Spilker, J. J. (1996). "Global positioning system: theory and applications." *AIAA.*

RTCA DO-236B (2003). "Minimum aviation system performance standards (MASPS): required navigation performance for area navigation (RNP RNAV) contains system requirements for operations in an RNP RNAV environment." *RTCA SC181.*

Sun, Q., Zhang, J. (2009). " RAIM method for improvement on GNSS reliability and integrity" *Digital Avionics Systems Conference, DASC '09. IEEE/AIAA 28th*.

U.S. Air Force, GPS Wing, Los Angeles Air Force Base (2005). "Navstar GPS space segment/user navigation user segment L5 interfaces." *IS-GPS-705*.

Van Dierendonck, A. J. Van and Hegarty, C. (2000) "The new L5 civil GPS signal" *GPS World.*

Van Graas, F., and Soloviev, A. (2007). "Coasting with relative carrier phase RAIM." *in Briefing to GEAS Panel.*

Walter, T., Enge, P., Blanch, J., and Pervan, B. (2008). "Future architecture to provide aviation integrity." *IEEE Proceedings.*

Walter, T., Enge, P., Blanch, J., and Pervan, B. (2008). "Worldwide vertical guidance of aircraft based on modernized GPS and new integrity augmentations." *Proceedings of the IEEE*, 1918-1935.

Whelan, C. (2001). "Evolution of air navigation services during this decade." *The Journal of Navigation*, 447-462.

Xue, R., Zhang, J., and Zhu, Y. B. (2009). "Cascade dual frequency smoothing for local area augmentation system." *Chinese Journal of Aeronautics*, 49-55.

Application of Jute Geo-Textile in the Moorum-Sand Mixed Layer as Granular Subbase of Flexible Pavement

Tapash Kumar Roy[1]

[1] Sr. Lecturer, Department of Civil Engineering, Bengal Engineering & Science University, Shibpur. Howrah-711103, West Bengal. India;
Email: tapash@civil.becs.ac.in

Abstract

In present scenario, non-availability of the conventional materials for construction of subbase of flexible pavement in many places in India has become one of the greatest problems for constructors due to variety of reasons. This problem may be overcome by decreasing the thickness of the sub-base layer by improving their strength, which ultimately reduces the cost of construction and also elongates the life span of pavement. In view of this, an experimental program was undertaken to study the possible uses of jute-geotextile in the construction of Granular Sub base layer of flexible pavement, with moorum and sand to evaluate their usefulness as a better sub base material and the result of these tests are reported in this paper.

Introduction

A good road network constitutes the basic infrastructure that forces the development process through connectivity. In the present scenario, India is the second largest in the world having a road network of 3.3 million kilometers after U.S.A (Indoria, 2009). **Pradhan Mantri Gram Sadak Yojana** as well as **Golden Quadrilateral** scheme are some important road network launched by the Government of India for providing the connectivity to the unconnected rural areas and improving connectivity among the metropolitan cities. As a result huge quantity of conventional road materials such as moorum, brickbats, crushed concrete, crushed stone, natural sand etc. are required for constructing the road network. In present situations, availability of such materials is not plenty because natural materials being exhaustible in nature, its quantity is gradually declining gradually. On the other hand, cost of extracting good quality of natural material is increasing. So need has been arisen to reduce the use of these conventional materials in the construction of road. This need may be

fulfilled by decreasing the thickness of the sub-base layer by improving their strength, which ultimately reduces the thickness of the base course and surfacing materials. Thus reduces the cost of construction to result cost effective road making and also elongates the life span of the constructed roads.

Literature Review

Investigative programs had been undertaken by different investigator by studying behaviors of the different combinations of various materials for improving the engineering properties of different layers of pavement. Jain et al. (1997) conducted a study on sand subgrade material for stabilization by using the natural fiber like coir and jute.

Another important area of utilization of fly ash has been long recognized in the construction of embankment as fill materials in reclamation works or in road construction for sub-base course (Reddy et al. 1999) or base courses (Tike et al. 1999). Satyanarayan et al. (2004), Kate (2005), Sridevi et al. (2005), Dhawan et al. (1988), and Behera et al. (2005) made a study on the effect of fly ash and lime when added to the expensive soil used for construction of road base, subbase and embankments. Experiments were also carried out by using waste materials like RHA and lime sludge which are being produced by various industries to stabilize the clayey soil (Chandra et al. 2005). Ravi Shankar et al. (2008) made a study to modify the characteristics of lateritic soil with addition of pond ash and cement.

Now-a-days different types of geo-synthetics have been widely used as reinforcing materials for construction granular sub-base and base layer for obtaining better strength. Geosynthetic Materials Association (2000) made a study about the incorporation of Geosynthetic as reinforcements into flexible permanent either as base or subbase and also used as subgrade restraint for construction of flexible or rigid roadways over weak subgrade conditions to aid in support of equipment loads on the unpaved base, or subbase, course during construction. Axel Nernheim (2005) discussed about the interaction behaviour between soil and geosynthetic of geosynthetic reinforced structures. But most these materials are not biodegradable and often lead to disposal crisis and environmental pollution.

Thus probability of using Jute Geo-textile which is eco-friendly along with cheap material Moorum available in major part of the country in different engineering construction for successful and cost effective application seems to be encouraging. In view of this, an exhaustive programme was undertaken by the author to study possible uses of jute geo-textile in the construction of Granular Sub base layer of flexible pavement, with moorum and sand to evaluate their usefulness as a better sub base material and the result of these tests are reported in this paper.

Materials Used

Four types of materials namely Soil, Sand, Moorum and Jute Geo-textile have been

used in the study to make an attempt to use them in the cost effective mix in the construction of Granular Sub-base layer of flexible road pavement. The properties of materials used in the investigation are described below:

Soil: Soil has been collected from the shallow depth of the nearby field in the campus of Bengal Engineering and Science University, Shibpur at Howrah, W.B., India and the evaluated properties of soil are given in the following table-1

Sand: Sand used in this investigation collected from Bankura, a district of West Bengal, India. Evaluated properties for the same are shown in the following table1.

Moorum: Moorum used in this investigation was obtained from Purba Medinipur, a district of West Bengal, India and the physical properties of the said materials are reported in the following Table 1.

Jute Geo-textile: Jute Geo-textile was collected from Indian Jute Research Association (IJIRA), Kolkata a district of West Bengal, India. Based on the standard specification of the fiber of the geo jute geo-textile given by IJIRA, the grey

Experimental Methods and Discussions on Test Results

For determining the quantitative information about the conventional mixed materials used for Granular subbase (GSB) layer, moorum has been mixed with sand in the ratio of 80:20. The physical properties of this mixed material have been evaluated in the laboratory as per the standard procedure laid down in the relevant BIS codes and the experimental results are given below:-

Table 1. Properties of the soil, sand and RHA used:

Sl. No.	Properties	Soil	Sand	Moorum
1	Specific Gravity	2.61	2.97	2.75
2	Grain Size Distribution:			
	Clay size particles	80%	0%	23.0%
	Silt size particles	14.66%	7.29%	9.11%
	Fine sand size particles	5.44%	42.63%	10.77%
	Medium sand size particles	0%	43.83%	15.0%
	Coarse sand size particles	0%	6.25%	26.24%
	Gravel size particles	0%	0%	15.88%
3	Heavy Compaction:			
	Proctor's Maximum Dry Density (gm/cc)	1.866	2.127	2.425
	Optimum Moisture Content (%)	13.4	8.47	13.0
4	CBR:			
	Soaked (%)	3.5	15.86	9.0
	Un Soaked (%)	5.0	17.85	14.3
5	Permeability (mm/sec)	5.23×10^{-6}	.037	0.000013

Table 2. Specifications for Jute Geo-textile as per IJIRA

Properties	Gray (Untreated)
Weight (g/m^2)	760
Threads/dm (MD × CD)	102 × 39
Thickness (mm)	2
Width (cm)	76
Strength (kN/m) (MD × CD)	20 × 20
Elongation at break (%) (MD × CD)	10 × 10
Pore size(O_{90})	300
Water permeability at 10 cm water head ($l/m^2/s$)	50
Puncture resistance (N/cm^2)	380

(untreated) woven jute geo-textile has been chosen for this study. The properties of the said materials are reported in the following Table 2.

From the grain size distribution analysis, it was found that the soil gradation consists of 2.27.0% Clay size particle, 7.95.0% Silt size particle, 14.74%, Fine sand size particle, 22.78% Medium sand size particle, 18.18% Coarse sand size particle and 34.08% Gravel size particle.

Heavy compaction results showed that maximum Proctor's density and optimum moisture content are 2.3 g/cc and 13.0% respectively and Specific gravity is 2.6.
CBR test values at OMC were 18.0% and 12.0% in unsoaked and soaked condition respectively.

From the results of experiment on moorum mixed sand, it is observed that the grain size distribution of the same has satisfied the specifications for the road and bridge works (Fourth Revisions, 2001) laid by Indian Roads Congress for granular subbase materials.

In the earlier part of investigation, the CBR values were determined by following laboratory-testing procedure. But for getting the spatial effect, the soil used for subgrade was prepared in a large container of diameter 36 cm. and height 60 cm. and energy was applied for compacting in three layers by using four hundred and sixty nine blows in each layer, so that energy applied per unit volume of each layer remain same as that used in laboratory CBR mould for the case of three-layer compaction. The placement moisture content was kept same in each case as evaluated in optimum moisture content of soil.

Next on top of the compacted subgrade, the granular subbase materials (moorum mixed sand) have been compacted in two layers of equal depth as the pavement is constructed in the practical field. For determining the effect of reducing the thickness of layer, three different compacted layers of total thickness of 150mm, 200mm and 250 mm have been used in this investigation, in where the compacted thickness of each layer is varying as 75 mm, 100 mm and 125 mm respectively. For achieving the compacted thickness of 75mm, 100 mm and 125 mm, loose materials

have been used as 112.5 mm, 150 mm and 187.5 mm respectively and the in-situ CBR tests have been performed on the same.

Further for knowing the effect of jute geo-textile on this granular subbase material, the jute-geotextile has been inserted at different positions of the compacted GSB layers of 150mm, 200mm and 250mm in two stages. In 1st stage, one layer of jute geo-textile has placed in-between the top of compacted subgrade and the bottom of the subbase. In 2nd stage, the said material has been placed in between two compacted layers of subbase inspite of earlier position and the in-situ CBR values have been determined for all the cases. The experimental values are given in the Table 3 and shown graphically in Figure 1.

Table 3. CBR Values of differentlayers of granular subbase

Description of Layer	Thickness of layer		
	150 mm	200 mm	250 mm
	In-situ CBR value (%)		
Subbase without jute geo-textile	8.80	9.50	10.2
Subbase with one layer of jute geo-textile above subgrade and below subbase	26.50	38.00	44.24
Subbase with one layer of jute geo-textile in between two compacted layers of subbase.	28.00	41.31	47.65

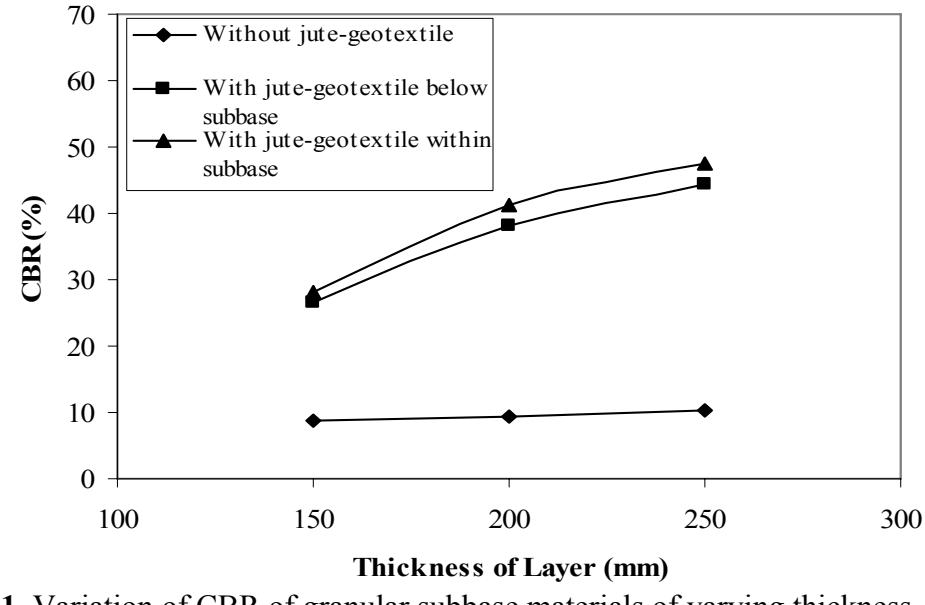


Figure.1 Variation of CBR of granular subbase materials of varying thickness with and without jute geo-textile

From the Table-3 and Figure-1, it is observed that In-situ CBR on compacted moorum mixed sand layers are 8.80, 9.50 and 10.2 for total layer thickness of 150 mm, 20 0mm and 250mm respectively of moorum mixed sand of proportion 80:20. However with application of jute geo-textile in the compacted moorum mixed sand layers in different positions, these values increases significantly due to development

of better friction between the fiber of jute geo-textile and moorum mixed sand materials. With the insertion of one layer of jute geo-textile in-between the compacted layer of subgrade and subbase, the said values increases to 26.50, 38.00 and 44.24 for total layer thickness of 150 mm, 200 mm and 250 mm respectively. Further the said layer has been replaced in the mid-position of the compacted layers of GSB materials instead of earlier position of jute geo-textile and the tested in-situ CBR values increases to 28.00, 41.31 and 47.65 for total layer thickness of 150 mm, 200 mm and 250 mm respectively due to development of better friction between the of jute geo-textile and moorum mixed sand materials from both sides of jute geo-textile layer. So the CBR value for insertion of one layer of jute geo-textile in the mid-position compacted layers of for compacted subbase layer of total thickness 200 mm is nearly 4.5 times the corresponding value of the compacted moorum mixed sand layers without any jute geo-textile layer.

Conclusion

From the above studies the following conclusions can be made:
(1) Crisis in procurement of conventional materials for construction of granular subbase course can be reduced by lowering the layer thickness of the same with improving strength. (2) Better improvement has been observed in CBR due to application of jute-geotextile with the compacted granular subbase course compare to that of the Subbase without jute geo-textile. (3) Significant improvement has been observed for insertion of one layer of jute geo-textile in between two compacted layers of subbase for compacted total layer thickness of 200 mm.

References

Axel Nernheim, (2005). "Design and test methods for geosynthetic reinforced structures." *Electronic Journal of GTE*, EJGE, USA.

Behera S.N., Sabat A.K., Prasad S.C., (2005). "Utilization of fly ash to reduce environmental pollution – a case study.", *Proceedings of the National Conference on Geotechnics in Environmental Protection,* Allahabad, VII, 28-30.

Chandra, S., Kumar, S. and Anand, K. R., (2005). "Soil stabilization with rice husk ash and lime sludge." *Journal of Indian Highway, Indian Road Congress*, (May) 87-98.

Dhawan P.K., Goswami K., Bhasin N.K., (1988). "Utilization of fly ash in road construction.", *National Workshop on Utilization of Fly Ash.* Roorkee, B,73-76.

Geosynthetic Materials Association, (2000). "Geosynthetic reinforcement of the aggregate base/subbase courses of pavement structures." *GMA White Paper II,* 0627001 GMA, Ryan R. Berg & Associates, Inc.

Indoria, R.P., (2009). "Use of locally available materials in road construction." *Journal of Indian Highway, Indian Road Congress*, (May).

IS 2720. Method of Test for Soil (Part ∇III), (1980). Laboratory Determination of Moisture Content & Dry Density.

IS 2720. Method of Test for Soil (Part X∇I), (1985). Laboratory Determination of CBR.

IS 2720. Method of Test for Soil (Part ∇), (1985). Laboratory Determination of Liquid & Plastic Limits.

Kate J.M., (2005) "Effect of fly ash and lime additives on swelling of expensive soils.", *Proceedings of Indian Geotechnical Conference,* Ahmedabad, 291-294.

Reddy R.M., Kumar N.D., Prasad K.N. and Rao K.M., (1999). "Beneficial use of fly ash in sub base courses." *Proceedings of the national seminar Fly ash characterization and its Geotechnical applications, Fly ash Mission,* TIFAC, Bangalore. 203-208.

Satyanarayana P.V.V., Rama Rao R., Krishna Rao C.V., (2004). "Utilization of lime fly ash stabilized expensive soil in roads and embankments.", *Proceedings of Indian Geotechnical Conference,* Warangal (India). 465-467.

Sridevi, G., Rama Rao, M., J. Padmini, B. Priyanka, and M. Swapna., (2005). "Effect of lime stabilization fly ash layer on cbr values of black cotton soils.", *Proceedings of the Indian Geotechnical Conference.* Ahmedabad. 1, 305-307.

Tike G.K., Kumar S., Seehra S.S. and Sikdar P.K. (1999). "Utilization of fly ash in construction of base courses." *Proceedings of the National Seminar on Fly ash Characterization and its Geotechnical Applications, Fly ash Mission,* TIFAC, Bangalore, 222-227.

Recent Development of Pavement Engineering Research in Hong Kong

L. W. Cheung [1], W. G. Wong[2] and Gordon L. M. Leung[3]

[1]Laboratory Manager, [2]Associate Professor, [3]Research Associate (PhD Candidate)
Department of Civil and Structural Engineering, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China; E-mail: gordonlm.leung@polyu.edu.hk

Abstract

The paper discusses the successful adoption of four major types of pavement materials in Hong Kong, namely porous asphalt, recycled asphalt pavement, asphalt rubber and stone mastic asphalt, based on the research works carried out in recent years by the Hong Kong Road Research Laboratory. Regarding the porous asphalt (PA), the study results indicated that the design with polymer modified binders exhibit smaller temperature susceptibility and higher resistances against ageing, rutting and fatigue than the one with conventional penetration 60/70 binder. For the study of new pavement incorporated with 15% of recycled asphalt pavement (RAP), it was found that the RAP materials performed as good as the control mixture of virgin material in terms of stiffness, water sensitivity, rutting and fatigue resistances. The site and laboratory testing results of asphalt rubber (AR) showed that the performance of the asphalt rubber wearing course was better than that of the conventional wearing course in providing good rideability, good fictional property, high water damage resistance and quiet surface. The properly designed stone mastic asphalt (SMA) mixture with modifier and admixture were able to enhance the rutting and fatigue resistances of a conventional SMA material. It is anticipated that the future direction of pavement engineering research in Hong Kong will be a more prominent utilization of accelerated pavement testing technique, and the first step taken would be the development of an accelerated pavement testing (APT) facility which will also be highlighted towards the end of the paper.

Key Words: porous asphalt; recycled asphalt pavement; asphalt rubber; stone mastic asphalt, accelerated pavement testing

Introduction

Within the last few years, the development of pavement materials in Hong Kong has been largely assisted by the various research works conducted by the Hong Kong Road Research Laboratory (HKRRL) of the Hong Kong Polytechnic University, and the achievements were well recognized by the government and the local highway industry.

The key successes are such as the introduction of porous asphalt (PA) as the surface course for most of the high speed roads in 2003, the adoption of Recycled Asphalt Pavement (RAP) (of 15%) in all bituminous road layers in 2004, the review of road drainage design and re-drafting of Road Note No. 6 in year 2006, road trials involving the use of Asphalt Rubber (AR) modified material at in 2007, as well as the use of modified Stone Mastic Asphalt (SMA) to withstand heavy traffic on highways in recent time. Other studies like road marking technology, pavement instrumentations, accelerated pavement testing technique, resilient behaviour of Hong Kong subgrade soil materials and the utilization of pavement management system (PMS) are currently underway. Most of the works can generate long term benefits to the highway industry in terms of "energy, cost and time savings".

Porous Asphalt

The use of porous asphalt has begun in places around the world since 1950s; and unlike dense graded or SMA mixture, porous asphalt comprises no fine material. Porous Asphalt is commonly used as a high performance, skid resistant wearing course to reduce traffic noise and water splashes in the wet days. This material was introduced into Hong Kong in the late 1990's. Soon after the introduction, it was noted that quite a number of porous asphalts suffered from premature failure. Developments of pot holes and severe fretting were observed, within the first three years of the service life. Details of the procedures of material production, laying, compaction and quality control were scrutinized by engineers from the government. It was reported that everything had gone according to the best practice at the time. In fact, the issue of early failure of porous asphalt was solved sweetly. Credit went to the collaboration effort of the government engineers and the HKRRL researchers. The two parties worked closely and set up scopes of the research on the porous asphalt for Hong Kong pavement together.

The research on the performance of porous asphalt in Hong Kong comprised three types of binders, which included a conventional penetration 60/70 binder, a styrene-butadiene-rubber polymer (SBR) modified binder and a pre-blended styrene-butadiene-styrene (SBS) polymer modified binder, and their respective porous asphalts. Properties of binder play an important role in the performance and durability of mixtures. To evaluate the effect of the binders on the performance of porous asphalt, penetration test, softening point test, viscosity test and dynamic shear rheometer (DSR) test were conducted on three conditions, including the original stage, short term ageing and long term ageing, of each of the three binders. Figure 1

shows the test procedures on the binders, of which the Rolling Thin Film Oven (RTFO) ageing is for short term ageing simulating material condition before trafficking, and Pressure Ageing Vessel (PAV) ageing is for material condition with a few years of age. The binder test results indicated that the two polymer (SBR and SBS) modified binders had the smaller temperature susceptibility and exhibited better performance in resisting ageing, rutting and fatigue than the conventional penetration 60/70 binder. They were more suitable to be used than the conventional 60/70 in the hot climate areas in Hong Kong.

There were two mixing methods for producing polymer modified asphalt mixtures in the Hong Kong asphalt industry at the time of the study, namely, wet process and dry process. In the wet process, polymer was mixed with binder before mixing with the aggregates at a specified temperature. In the dry process, polymer, bitumen and aggregate were mixed together. That is to say, for the later, the polymer is to modify the asphalt mixture rather than the binder. In this study, the dry mixing method was applied to produce the polymer (SBR) modified porous asphalt which was to simulate the material production process being used in the local batching plant. For the other binders, the wet process was employed.

The performance of the three types of porous asphalts was judged by the value of stiffness modulus and their ability to resist moisture damage, permanent deformation. Results of the study showed that despite of the use of the modified binder, the polymer (SBR) modified porous asphalt had the least ability to resist water damage, suggesting that the deficiency was caused by a less effective mixing method, dry mix process, which was adopted by the local asphalt industry for producing this mix.

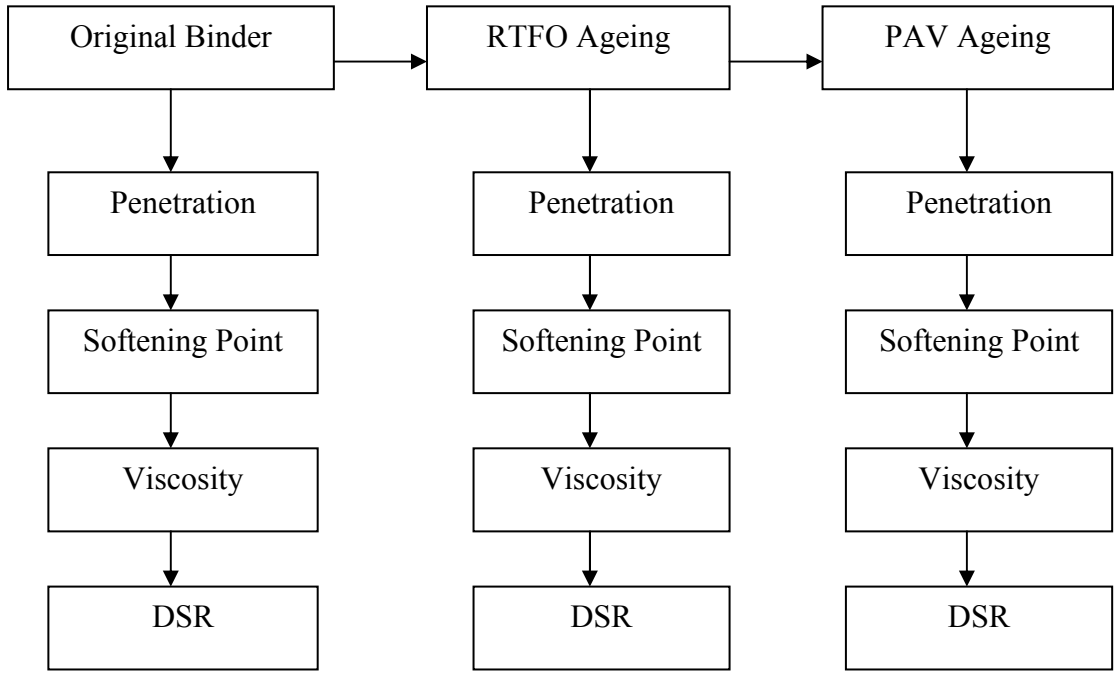


Figure 1. Test procedures on binders.

Recycled Asphalt Pavement

The incorporation of recycled asphalt pavement material reduces the demand for both new aggregate and asphalt binder in the mixture being produced. Reuse of RAP for pavement construction has been considered as a good practice in many developed countries.

In Hong Kong, about 0.2 million tonnes of RAP were generated each year and about 1 million tonnes of new bituminous mixtures were produced each year in the last decade. Therefore, the ratio between the RAP produced and the new materials was around 0.2 (20%). Of these RAP materials, a total of about 0.1 million tones were delivered to land fill sites as construction and demolition waste, and the remaining were used as temporary road filling materials in private projects. As the sites of reclamation and capacity of public fill are reducing rapidly, the disposal problem of RAP in Hong Kong is eminent. In fact, RAP can be a valuable material, when designed properly, for road maintenance and construction. Furthermore, the use of RAP for road pavement in Hong Kong could eliminate disposal problem, release space of dumping site, conserve natural materials (less bitumen and aggregate needed) and reduce cost of construction and maintenance.

Nevertheless, the use of RAP in hot mix asphalt can be a tricky business, involving additional processes, such as control of the RAP quality and homogeneity, assessment of RAP binder content and its consistency property, and dealing with the obvious challenge of day-to-day variation of RAP binder and RAP aggregate gradation. This is not just a matter of lack of skill in handling RAP material, but also lack of experience (i.e. asphalt production can go wrong anytime and has its moments particular in the case of insufficient skill and experience).

The study on asphalt mixtures incorporating with RAP was the pioneer of introducing recycled materials for the structural layers of asphalt pavement in Hong Kong. The objective of the study was to verify the suitability of incorporating 15% RAP into bituminous materials based on the performance of the asphalt mixture with RAP without the addition of rejuvenating agent or one grade softer binder.

The performance tests conducted included determination of stiffness, water sensitivity, rutting resistance and fatigue resistance. It was found that the RAP materials performed as good as the control mixture of virgin material. As the results of the study, bituminous paving materials incorporating 15% RAP, without changing any design specification, are now being used in Hong Kong. There is no reason why the incorporation of RAP should stop at only 15% RAP, particularly for large projects involving full reconstruction of pavement which generates huge volume of RAP. With the encouraging results of the current study, further research has been carried out to incorporate higher RAP percentage to release the full benefit of reusing RAP in Hong Kong and to eliminate the waste problem arising from materials of pavement rehabilitation.

Asphalt Rubber

The study on the use of crumb rubber was another research effort on reusing waste materials. It has been recognised that the benefits of incorporating crumb rubber into asphalt mixtures as paving materials are two folds – reduction of the waste problem created by scrap tyres and improvement of the performance of asphalt materials.
To assess the potential of using crumbed rubber tyre for wearing course material in Hong Kong, wearing course incorporating with crumb rubber (ARWC), were constructed at the Hong Kong EcoPark. Two ARWC pavement sections, each 33 meters in length, were selected for detailed research study. In addition, a control section of conventional wearing course material was constructed on the site for performance comparison. Figure 2, shows the location plan of the trials, of which Section A is the control. The major tasks of the study were 1) to study the effects of the crumb rubber modifier on the performance of bituminous mixture in Hong Kong; 2) to assess the procedures of material handling, storing and transportation; and 3) to assess the methods of material mixing, laying and compaction.

A series of laboratory tests and on-site tests were designed. The laboratory tests targeted at studying the performance (including permanent deformation resistance, stiffness, water damage resistance and fatigue resistance) of cored samples taken from site, in comparison with the traditional wearing course material. The on-site performances of these trial sections were monitored at three month intervals for a year. The on-site tests included determination of the structure performance using Falling Weight Deflectometer FWD (Photo 1), measurement of the change of skid resistance by pendulum skid tester (Photo 2). Two modes of noise measurement were also taken. The first mode made use of a conventional pass-by "Control Coast-By" method (Photo 3) with stationary measurement equipment set up on the road side to record noise level when vehicles passed by. The second mode of noise measurement made use of the technology of Close-Proximity equipment (Photo 4), which was designed to measure the noise at source. Other site measurements included determination of surface texture, permeability and surface roughness.

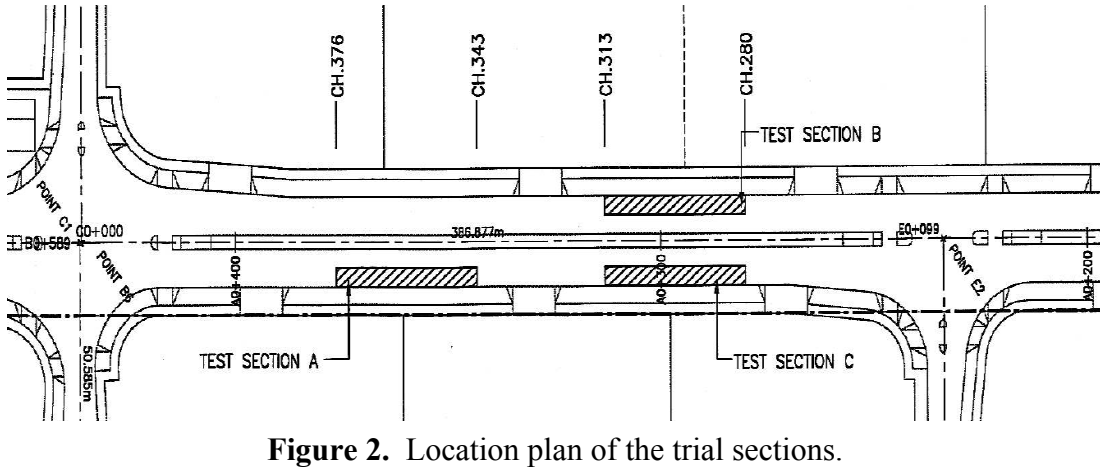


Figure 2. Location plan of the trial sections.



Photo 1: Structural Capacity Determination by FWD Testing

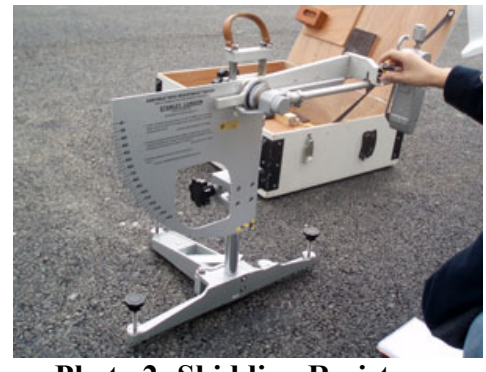


Photo 2: Skidding Resistance Determination by Pendulum Skid Tester

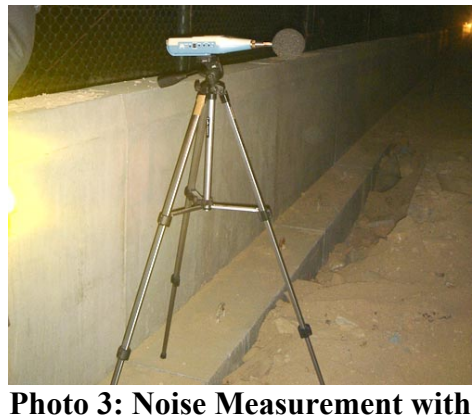


Photo 3: Noise Measurement with Control Coast-By Method



Photo4: Noise Measurement by Close-Proximity Equipment

The results obtained from the site and laboratory tests indicated that the performance of the ARWC was better than that of the conventional wearing course. It was able to provide good rideability, good fictional characteristics, high resistance to water damage and quiet surface. The experience gained from the on-site trial had provided vital information for the further development of bituminous mixture incorporating with crumb rubber in Hong Kong.

Stone Mastic Asphalt

Stone Mastic Asphalt (SMA) was originally developed in Germany in the late 1960s as an effort to resist detrimental effects of tyres equipped with winter driving studs. Despite the use of studded tyres, which was later banned, the interlocked skeleton of coarse aggregates within SMA continued to be proved as very durable and rut resistant material.

To fulfill the increasing demand for high quality, more durable, low noise, high skid-resistant surfacing materials for carrying the high traffic volume in Hong Kong, a

research study was conducted at HKRRL to investigate the potential of using different types of SMA. To gain the benefit of making use of the latest pavement material advancement, apart from the conventional penetration grade bitumen, the uses of SBS binder, AR binder and binder with cement modifier in SMA were studied.

Table 1 presents the SMA materials tested in the study. They included five types of SMA. The materials of SMA10, cem-SMA10, SMA10-SBS and SMA10-AR were designed to meet the latest EU grading envelop of 10mm nominal maximum aggregate size and air void requirements between 3.5% and 6.5%. SMA8-Germany was to meet the grading requirements of the Germany specification for SMA 0/8 LA that provided high air void content of 9%-10%.

The performances of these surfacing materials were assessed under the controlled condition in laboratory. It was found that the proper uses of modifier and admixture were able to enhance the performance of the SMA materials. For instance, the rutting resistance of a properly designed SMA mixture after incorporating modified binder could be double. It has been suggested to conduct site trials at locations with severe rutting problem to evaluate the potential benefit of applying SMA in the upper layers (including the wearing course layer and the base course layer). It was also noted that the fatigue lives of the SMA's with modifiers were much higher than that of the conventional SMA material.

Table 1. Types of SMA Tested.

Materials		SMA10	Cem SMA10	SMA10 SBS	SMA10 AR	SMA8 Germany
Sieve Size (% Passing)	14mm	100	100	100	100	100
	10mm	97	97	97	98	96
	6.3mm	-	-	-	-	50
	5mm	33	33	33	44	20
	2.36mm	26	26	26	35	16
	1.18mm	22	22	22	27	11
	0.6mm	20	20	20	22	8
	0.3mm	17	17	17	18	7
	0.15mm	14	14	14	13	6
	0.075mm	10	10	10	9	5
Ordinary Binder or Modified Binder Contents, % by Weight of Total Mixture		6.5	6.5	6.75	7.9	7.4
Portland Cement Content, % by Weight of Dry Aggregate		-	3	-	-	-
Air Voids in Mix, %		3.9	4.0	4.0	6.1	9.6

Note:Weight of aggregate includes the weight of hydrated lime and cement.Cem-SMA10: SMA with 3% cement, SMA10-SBS: SMA with SBS binder; SMA10-AR: SMA with AR binder.

The requirements of using modified bitumen, additives and high binder content generally push up the unit production cost of SMA. Life Cycle Cost Analyses (LCCA) conducted in the current study suggested that the whole life cost, taking into account the reduced maintenance and the longer service life, the costs of applying modified SMA materials are very likely lower than that of the conventional material without modified bitumen.

Accelerated Pavement Tester (APT)

In parallel to the materials development, the Research Laboratory has been regularly upgraded with advanced test equipments, aiming to pursue better knowledge in pavement technology for Hong Kong. With the use of more advanced equipments, it is no doubt that better understanding on the behaviour of pavement materials can be obtained. However, certain shortcomings for laboratory techniques are still inevitable such as tests involved are far from being near the real scale situation and tremendous efforts have to be made to compile and interpret the results in order to make estimation on how the whole pavement system would perform.

In response to these needs involved for our research works and the growing trend of real scale road testing technique worldwide, several years ago, the Hong Kong Road Research Laboratory had decided to develop a full-scale accelerated pavement testing facility (APT).

The basic design of the APT involves dual axle loaded wheels running on a 22m diameter circular wheel track. The design speed of the loaded wheels is in a range of 1 – 70 km/hr and the maximum load carrying capacity is to be 70 kN per half axle. The structural arms are designed to be foldable to allow for more convenient construction and re-construction of the wheel track from time to time. The general layout of the APT facility design and the conceptual design of the folded arm are shown in Figures 3 and 4 respectively. The overall construction works of the facility has commenced in mid 2009, which involved the formation of a 2m high subgrade soil platform (consisted of completely decomposed granitic soil) for the supports of the wheel track and a reinforced concrete shallow footing for the fixture of the APT's mechanical components. The detailed design and fabrication of the APT are underway at the time of writing this paper, it is anticipated that the whole facility will start operating in mid 2010.

Pavement instrumentations are located at various layers of the wheel track to monitor the responses of pavement such as stress, strain, moisture and temperature. Under the laboratory condition, the study of aspects such as calibration, installation, data acquisition and verification of dynamic performance of the various instrumentations were carried out.

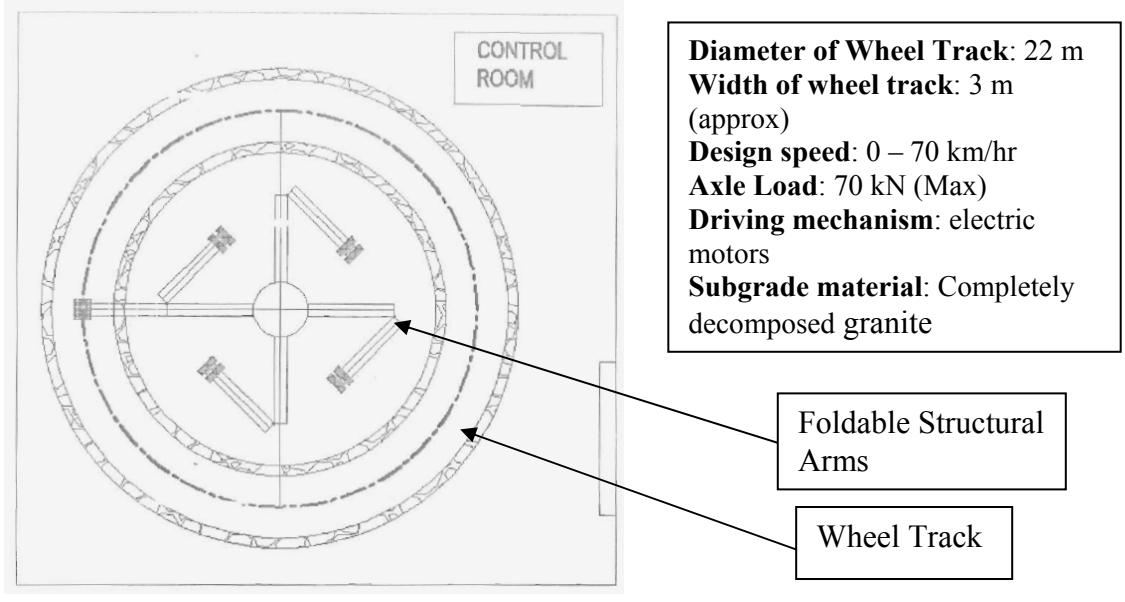


Figure 3. General Layout of the APT Design.

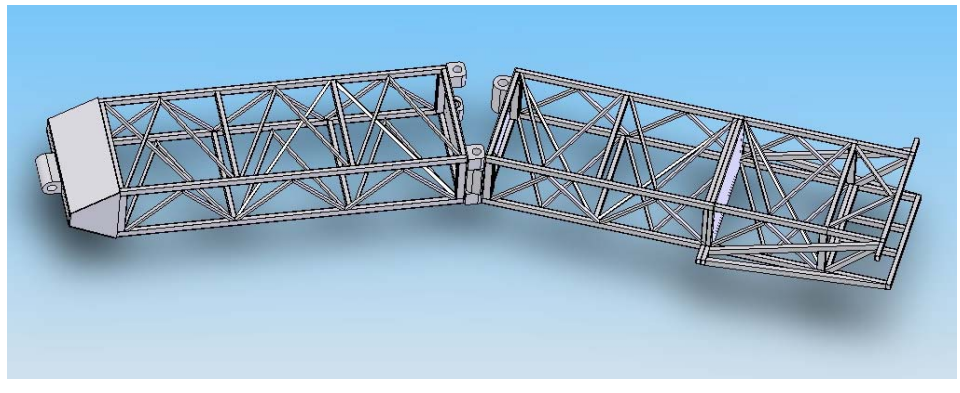


Figure 4. Foldable Design of Structural Arms.

Conclusions

Within the past several years, the development of pavement materials in Hong Kong as mentioned in this paper was substantially assisted by laboratory testing. Apart from this and in a certain extent, those materials have already borne the challenges on real roads through the daily public usage of traffic, but obviously with very limited amount of scientific verifications. Hence the future direction would tend to be a more prominent utilization of the full scale accelerated testing and real time instrumentation techniques to study the material behaviour in a quality and systematic manner.

References

Arizona Department of Transportation, (2008). "Asphaltic concrete friction course (Asphalt-Rubber)." *Standard Specifications for Road and Bridge Construction*.

Huber, G, (2000). "Performance survey on open-graded friction course mixes." *NCHRP Synthesis* 284, Transportation Research Board, National Research Council.

Jorge, B. Sousa, (2006) Proceedings of the Asphalt Rubber 2006 Conference, Palm Springs, USA.

Malpass, G. A, (2003) "The use of reclaimed asphalt pavement in new superpave asphalt concrete mixtures." *Ph.D. Thesis*, North Carolina State University.

Manual of Contract Documents for Highway Works, (2004) "Road pavement – bituminous bound materials." *Volume 1, Specification for Highway Works, UK*.

Nicholls, J. C, (1997). "Review of UK porous asphalt trials." *TRL Report* 264. Transport Research Laboratory, Crowthorne, UK.

State of California Department of Transportation, (2005) Use of Scrap Tire Rubber, State of the Technology and Best Practices.

Transportation Research Board, (2004). "Significant findings from full-scale accelerated pavement testing." *NCHRP Synthesis* 325, Washington D.C.

West, Randy C. and Moore, (2006). "Evaluation Georgia's compaction requirements for stone matrix asphalt mixtures." *The USA National Center for Asphalt Technology and the Georgia Department of Transportation,* 2006. Section 414.

Wong W.G, (2004). "Testing on polymer modified binders and friction course (Porous Asphalt)." *Report Submitted to The Highways Department of Hong Kong.*

Wong W.G, (2006)."Testing of bituminous materials incorporating with reclaimed asphalt pavement (RAP)." *Report Submitted to The Highways Department of Hong Kong*.

Wong W.G, (2009). "Monitoring on-site performance of asphalt-rubber wearing course." *Report Submitted to The Highways Department of Hong Kong*.

Wong W.G, (2009)."Research on the performance of gap-graded asphalt surfacing materials." *Report Submitted to The Highways Department of Hong Kong*.

Wong C.C and Wong W.G, (2007)."Effect of crumb rubber modifiers on high temperature susceptibility of wearing course mixtures." *Construction and Building Materials* 21 (2007) 1741–1745.

A Nonlinear Viscoelastic Constitutive Model of Aged Asphalt Mixture

Xiaoge Tian[1]

[1] School of Traffic & Transportation, Changsha University of Science & Technology, Changsha, Hunan, P.R.China, 410076, Email: tianxiaoge@126.com

Abstract

Asphalt mixture is a complicated thermal rheological material, and its deformation is related to some factors, such as the stress state, stress history, temperature, and loading rate. So it has a very complicated nonlinear viscoelastic property. To find a nonlinear viscoelastic constitutive model of aged asphalt mixture, stress relaxation tests were carried out on asphalt mixtures (AC-13C), having five different aged degrees, under different temperatures and strain levels. Then, a fractional exponential model was founded to characterize the stress relaxation performance of aged asphalt mixture, and the regression relationships of the values of the model parameters, varying with the aging degree, strain and temperature, were established applying the non-linear regression method according to the test results. Finally, a nonlinear viscoelastic model of aged asphalt mixture was established.
Key Words: asphalt mixture; nonlinear viscoelastic; constitutive model; stress relaxation test; aging

Introduction

Asphalt mixture is a typical viscoelastic material, it behaves linearly only when its strain is less than 0.1% (Monismith et al. 1965) and usually behaves non-linear viscoelastic state under the actually working conditions. Many researchers have studied the nonlinear viscoelastic character of asphalt mixture (Lockett 1972; Findley et al. 1976; Schapery 1969；Schapery 1982; Rabotnov 1980; Zhang and Zhang 2003; Li 2003; Zhang 1999), but there is not a standard constitutive model currently.

Stress relaxation tests were conducted on different aged asphalt mixtures, under different temperatures and strain levels. Then the fractional exponential model was founded to characterize the stress relaxation performance of aged asphalt mixture, and the regression relationships of the values of the model parameters, varying with

the aging degree, strain, and temperature, were established applying the non-linear regression method from the test results. Finally, a nonlinear viscoelastic model of aged asphalt mixture was established.

Test Scheme

The object of this paper is to establish a nonlinear viscoelastic model from the stress relaxation tests of aged asphalt mixtures under different temperatures and strain levels. A Shell asphalt, AH-70, was used, and the mixture, AC-13C, was formulated according to the Technical Specifications for Construction of Highway Asphalt Pavement (JTG F40-2004). The mixed loose mixtures were aged in the oven for 0h, 4h, 8h, 12h, or 24h to obtain different aged degree asphalt mixture, then the aged mixtures were fabricated into beams, whose size is 50mm*50mm*250mm. All the stress relaxation tests were conducted on a MTS810, and the test temperature was 10C or 15C, the strain level was 0.0012με, 0.0014με, 0.0016με, 0.0018με, or 0.0020με.

Asphalt

The asphalt used in this paper was an AH-70. Table 1 shows the test results.

Mixture Gradation

The asphalt mixture in the tests was a dense gradation mixture, AC-13C, formulated according to the Technical Specifications for Construction of Highway Asphalt Pavement (JTG F40-2004). And the gradation was shown in Table 2. The optimal asphalt content was determined as 5.3% using the Marshall tests.

Table 1. Test results of AH-70 asphalt.

Items	Specifications	Results	Test Method
Penetration (25 °C,100 g,5s) (0.1 mm)	60~80	73	T 0604-2000
Penetration index	-1.5~+1.0	-0.196	T 0604-2000
Ductility (5cm/min, 10 °C) (cm)	≥15	39	T 0605-1993
Ductility (5cm/min, 15 °C) (cm)	≥100	>100	T 0605-1993
Soft temperature (Ring & Ball) (°C)	≥46	48.5	T 0606-2000
Wax content (%)	<2.2	0.6	T 0615-2000
Flash point (°C)	≥260	332	T 0611-1993
Solubility (%)	≥99.5	99.77	T 0607-1993
Density (15 °C)	Measured	1.041	T 0603-1993
Dynamic viscosity @60 °C (Pa.s)	≥160	199	T 0620-2000
Rolling Thin Film Oven (RTFO) Test @163 °C 85 min			
Mass loss (%)	≤±0.8	0.3	T 0609-1993
Residual penetration ratio (%)	≥61	62	T 0604-2000
Residual ductility (10 °C) (cm)	≥6	9	T 0605-1993
Residual ductility (15 °C) (cm)	≥15	66	T 0605-1993

Table 2. Gradation of AC-13C.

Gradation	Size of sieves (mm) and percent of passing (%)										
	19	16	13.2	9.5	4.75	2.36	1.18	0.6	0.3	0.15	0.075
Upper limit	100	100	90	68	38	24	15	10	7	5	4
Lower limit	100	100	100	85	68	50	38	28	20	15	8
Formulation	100	100	95.2	71.8	51.6	32.9	23.6	18.3	12.1	9.3	5.8

Sample

To obtain aged asphalt mixture having different aging degree, the newly mixed asphalt mixture was placed loosely into a square plate, depth was 21–22 kg/m^2, then the plates were placed into an oven, 135 ^{o}C±1 ^{o}C, and taken out to turn over once an hour. Once the mixtures were aged in the oven for 0h, 4h, 8h, 12h or 24h, respectively, the mixtures were taken out and fabricated into a beam. The sizes were 50 mm*50 mm*250 mm, for the stress relaxation tests.

Test Results

The stress relaxation tests of different aged asphalt mixtures were conducted on a MTS-810. The test procedure is omitted here because it is an usual test. The initial strain had five levels, i.e., 0.0012 με, 0.0014 με, 0.0016 με, 0.0018 με, or 0.0020 με and the test temperature was set to 10^{o}C or 15^{o}C. The relaxation modulus was calculated from the tests results.The relaxation modulus of different aged asphalt mixtures at different initial strain levels (10 ^{o}C) were shown in Figure 1.

The relaxation modulus of different aged asphalt mixture at different initial strain levels (15 ^{o}C) were shown in Figure 2.

Nonlinear Viscoelastic Model of Aged Asphalt Mixture

A fracture exponential model was founded to formulate the nonlinear viscoelastic behavior, based on the weaknesses of the widely used models, standard linear viscoelastic solid model, and the Rabotnov model. The fracture exponential model was shown in Eq. (1):

$$E(t) = E_\infty + (E_0 - E_\infty)\exp\left(-\beta\left[(\gamma+\alpha)t\right]^{1-\alpha}\right) \tag{1}$$

where, E_0 is the initial (instantaneous) relaxation modulus, E_∞ is a equilibrium relaxation modulus，α, β and γ are the material parameters, ($0<\alpha<1, \beta>0, \gamma>0$), and the values of them can be determined by fitting the test results.

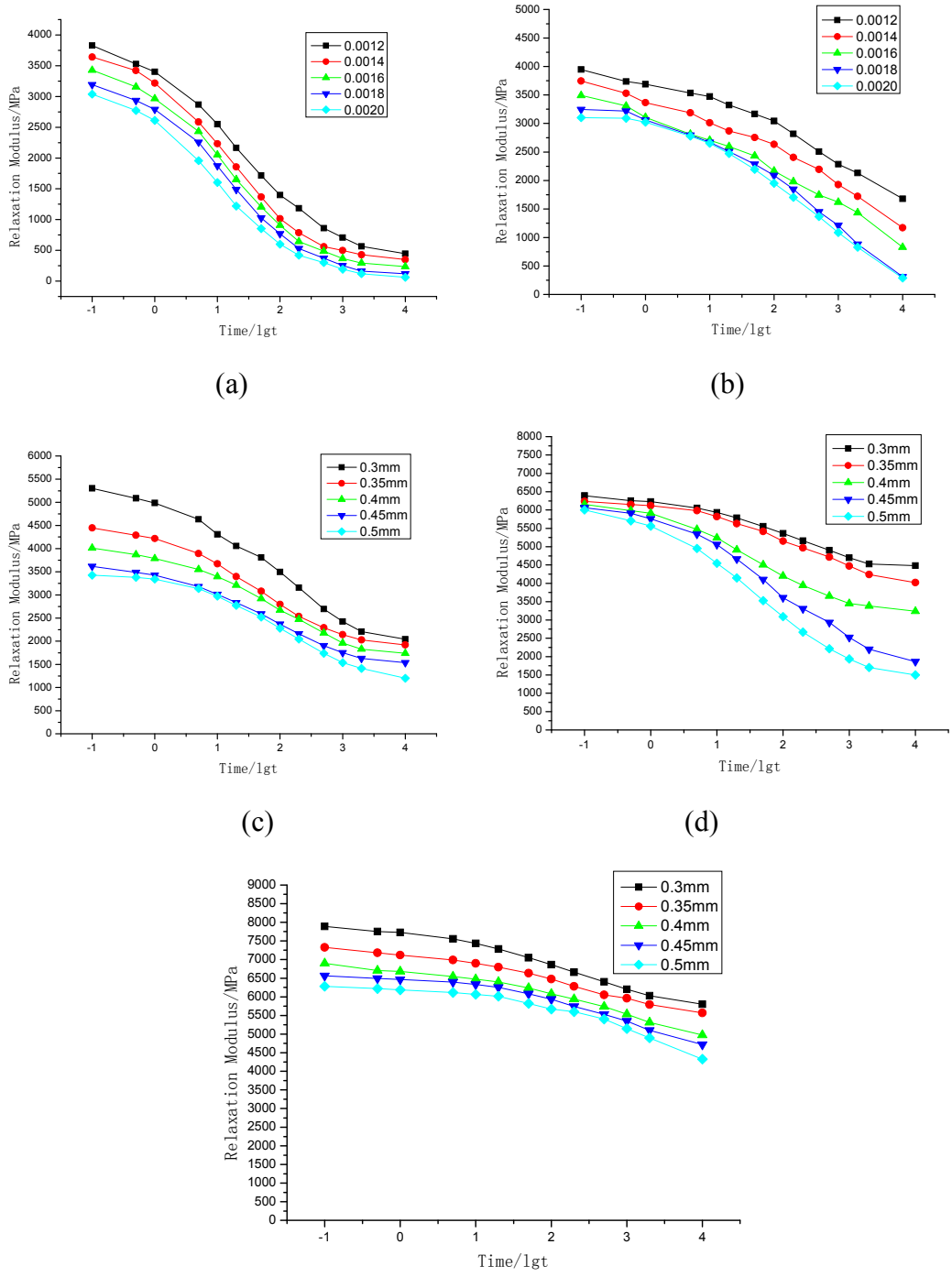


Figure 1. The relaxation modulus of different aged asphalt mixtures at different initial strain levels (10 °C) (a. 0 h, b. 4 h, c. 8 h, d. 12 h, e. 24 h).

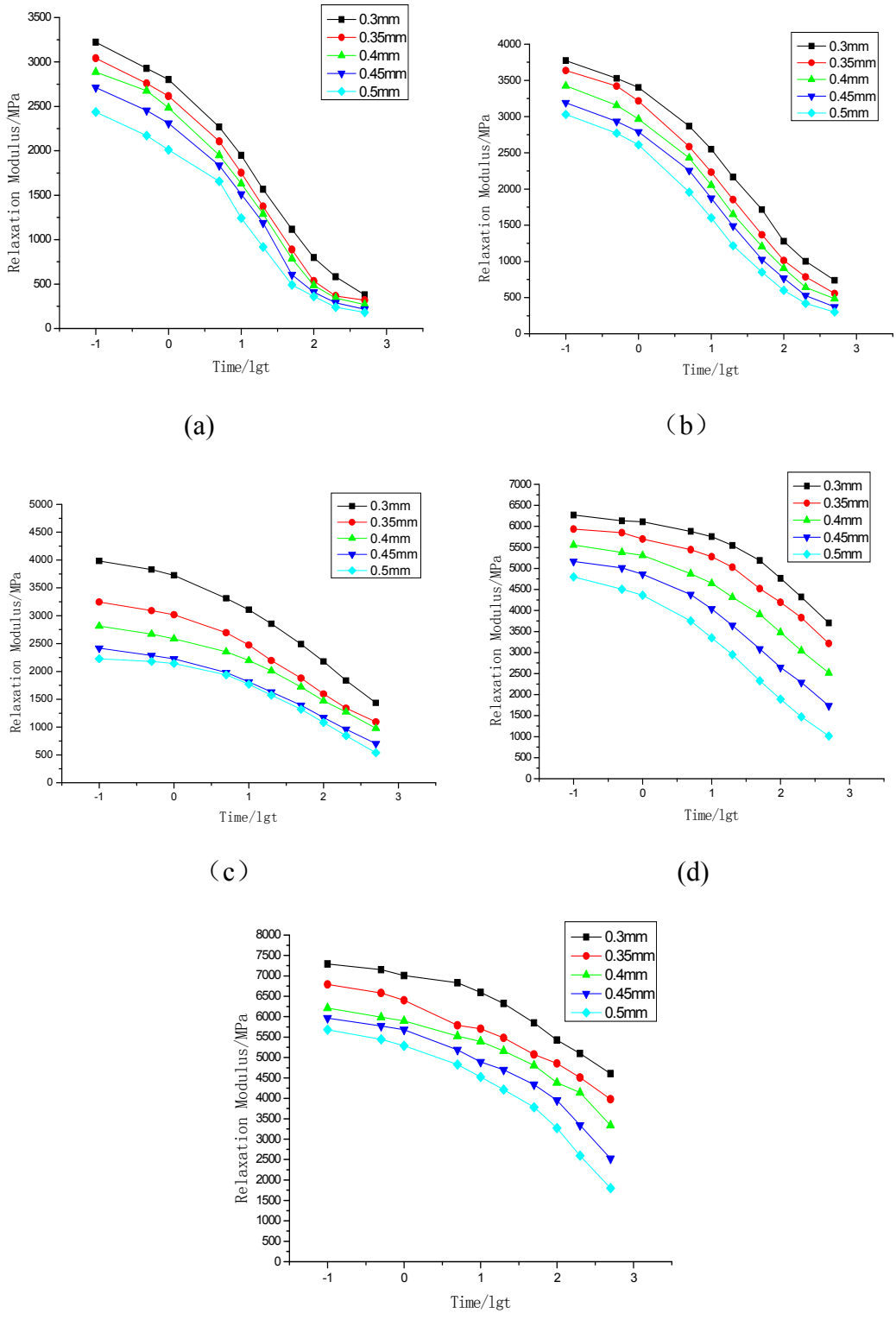


Figure 2. The relaxation modulus of different aged asphalt mixture at different initial strain levels (15 ^{o}C) **(a.0h, b.4h,c.8h,d.12h,e.24h)**.

As the values of each parameter in the fracture exponential model would vary with temperature, aged content and strain level, it was assumed that the values of these parameters were quadratic polynomial relative to these factors. i.e.,

$$\mathrm{E}_0 = a_1 \cdot \varepsilon^2 + b_1 \cdot \mathrm{h}^2 + c_1 \cdot T^2 + d_1 \cdot \varepsilon + e_1 \cdot h + f_1 \cdot T + g_1 \quad (2)$$

$$E_\infty = a_2 \cdot \varepsilon^2 + b_2 \cdot h^2 + c_2 \cdot T^2 + d_2 \cdot \varepsilon + e_2 \cdot h + f_2 \cdot T + g_2 \quad (3)$$

$$\alpha = a_3 \cdot \varepsilon^2 + b_3 \cdot h^2 + c_3 \cdot T^2 + d_3 \cdot \varepsilon + e_3 \cdot h + f_3 \cdot T + g_3 \quad (4)$$

$$\beta = a_4 \cdot \varepsilon^2 + b_4 \cdot h^2 + c_4 \cdot T^2 + d_4 \cdot \varepsilon + e_4 \cdot h + f_4 \cdot T + g_4 \quad (5)$$

$$\gamma = a_5 \cdot \varepsilon^2 + b_5 \cdot h^2 + c_5 \cdot T^2 + d_5 \cdot \varepsilon + e_5 \cdot h + f_5 \cdot T + g_5 \quad (6)$$

where, ε is the strain level in the stress relaxation test, h is the aged hours of the asphalt mixture in the oven (Hr), and the T is the test temperature (^{o}C). A_i, b_i, c_i, f_i, and d_i were the fitted constants (i=1–5).

The relaxation modulus of different aged degree asphalt mixtures under different temperatures and strain levels were fitted nonlinearly to the equations (Eq.1 to Eq.6) using the software ORIGIN. Then, the relations of relaxation modulus to loading time, aged degree, strain level, and temperature were obtained, that was the nonlinear viscoelastic model of aged asphalt mixture, as shown in Eq. (7).

$$E(t) = E_\infty + (E_0 - E_\infty)\exp\left(-\beta[(\gamma+\alpha)t]^{1-\alpha}\right) \quad (7)$$

where,

$$\mathrm{E}_0 = 6.76\times10^8\varepsilon^2 - 0.5\mathrm{h}^2 + 0.56T^2 + 187.91h - 3.64\times10^6\varepsilon - 1142.8T + 825807 \quad (\mathrm{R}^2 = 0.867)$$

$$E_\infty = 1.5h^2 - 1.5\times10^7\varepsilon^2 - 29.51T^2 + 172.85h - 1.6\times10^6\varepsilon + 671.45T - 882.91 \quad (\mathrm{R}^2 = 0.944)$$

$$\alpha = 0.035\mathrm{T}^2 - 0.9\mathrm{T} - 1.39\times10^{-5}h^2 + 0.01h - 1.83\times10^5\varepsilon^2 + 558.84\varepsilon + 5.65 \quad (\mathrm{R}^2 = 0.660)$$

$$\beta = 0.037\mathrm{T}^2 - 0.908\mathrm{T} + 0.01h^2 - 0.018h - 2.7\times10^5\varepsilon^2 + 954.57\varepsilon + 4.89 \quad (\mathrm{R}^2 = 0.698)$$

$$\gamma = 0.037\mathrm{T}^2 - 0.9\mathrm{T} + 6.61\times10^{-5}h^2 - 0.08h + 5.9\times10^4\varepsilon^2 - 212.96\varepsilon + 5.18 \quad (\mathrm{R}^2 = 0.515)$$

Conclusion

The stress relaxation tests of different degree of aged asphalt mixtures (0h, 4h, 8h, 12h, and 24h) were conducted on the MTS810 at different temperature and strain levels. And the variation patterns of relaxation modulus to aged degree, strain level, loading time, and temperature were observed. Based on defects analysis of the often used nonlinear viscoelastic models, a fracture exponential model was put forwarded. Then, the values of parameters in the model were regressed according to all the test results, and the relations of every parameter's values to the aged degree, strain level, temperature and loading time were regressed. Finally, a fracture exponential model was founded to characterize the nonlinear viscoelastic effects of an aged asphalt mixture.

Acknowledgment

The study was supported by the National Natural Science Foundation of China (No. 50878032).

References

Zhang C., and Zhang, W. (2003). "Research progresses on nonlinear viscoelasticity and its applications," *Natural Science Journal of Xiangtan University*. 28-32.

William N. Findley, James S. Lai, Kasif Onaran. (1976). "Creep and relaxation of nonlinear viscoelastic materials", *North-Holland Pub. Co*.
Lockett, F. J. (1972). "Nonlinear viscoelastic solids," *London*: *Academic Press.*

Monismith, C. L., Secor, G. A., Secor, K. E. (1965). "Temperature-induced stress and deformations in asphalt concrete," *Proceedings of AAPT*, 34.

Rabotnov, Yu. N. (1980). "Elements of hereditary solid mechanics," *Mir Publ.,* Moscow.

Schapery, R. A. (1969). "On the xharacterization of nonlinear viscoelastic materials," *Polym. Eng. Sci*., 295-310.

Schapery, R. A. (1982). "Models for damage growth and fracture in nonlinear viscoelastic particulate composites," *Proceeding Ninth U.S. National Congress of Applied Mechanics,* 237-245.

Zhang, W. (1999). "Practical expressions of relaxation modulus and creep compliance," *Natural Science Journal of Xiangtan University*, 26-28.

Li, Y. (2003). "Research on the theory and the solution approach of nonlinear viscoelastic problems," *Xiangtan University*, 2003.

Influence Factors and Anti-Rutting Agent of High-Temperature Stability of Asphalt Mixture

Ping Li[1], Xiuli Wang[1], Huishan Li[1] and Yu Qian[2]

[1] College of Civil Engineering, Lanzhou Univ. of Tech., No. 287 Langongping Street, Lanzhou; Email: liping@lut.cn
[2] Tianhong Communications Tech. Co., Ltd., No.14 Chang'an Street, Xi'an; Email: Qianyu@126.com

Abstract

Rutting in asphalt pavement can be influenced by many factors of high-temperature stability of mixture. Taking three continuous-graded and two gap-graded asphalt mixtures as investigation objects, Dynamic Stability (DS) and Rutting Depth (RD) were measured in cases of different temperatures, wheel rolling times and contents of a new anti-rutting agent (ARK-1) by Wheel Tracking Test. The results of testing showed that resistance to rutting of the fine aggregate mixture would be superior to that of the coarse aggregate mixture. No obvious differences between the continuous-graded and gap-graded mixtures could be observed in resistance to rutting as temperature was equal or lower than 60^{o}C. High-temperature stability of the gap-graded or coarse aggregate mixture was better than that of the continuous-graded or fine aggregate mixture as temperature was beyond 60^{o}C. ARK-1 could dramatically improve resistance to rutting of the asphalt mixture. A project example on Baotian freeway between stations 35+460 and 36+460 proves that the costs of pavement made with ARK-1 will be more economical than pavements with SBS-modified binder.
Key Words: asphalt mixture; high-temperature stability; anti-rutting agent; dynamic stability (DS); rutting depth (RD)

Introduction

Rutting is a major distress in asphalt pavement in cases where heavy traffic occurs during high-temperature season. Destruction of a pavement structure in early periods or many traffic accidents may be caused by rutting of pavement if rutting area and depth are beyond the limited value. A great deal of pavement maintenance must be

employed to improve pavement performance after one or two hot summer seasons. How to effectively and economically improve high-temperature stability of asphalt mixture is still an important topic of research in recent years.

In recent research work, a full-scale pavement-testing program was used to evaluate the rutting resistance of Superpave mixtures with and without polymer modification using a Heavy Vehicle Simulator (HVS) (Okan, 2006). Pavement sections with two 5 cm lifts of SBS-modified mixtures clearly outperformed those sections with two 5 cm lifts of un-modified mixture. The effects of different crumb rubber modifiers (CRMs) on the high temperature susceptibility were investigated by Soon-Jae et al. (2007), Cheuk Ching and Wing-gun (2006) and Huang et al. (2002). Permanent deformation tests that cylindrical samples produced with Marshall compaction were compared with the wheel-tracking tests conducted using three different elastomeric polymer modifiers as shown by Halit et al. (2007). Similar research could be found in Xiao, Serji and Hsein (2007), which reported that crumb rubber in the HMA could effectively improve the rut resistance of mixes with two rubber types, four rubber contents and three crumb rubber sizes. Ali and Burak (2004) and Gandhi and Gallaway (1970) analyzed effects on fine aggregate angularity (FAA) in relation with the resistance to rutting of hot-mix asphalt using rutting tests. The rutting test result showed that higher FAA values increased resistance to rutting of hot-mix asphalt. Performance of hot mix asphalt (HMA) with anti-rutting agent PR PLASTS in case of different contents were studied by Xiao et al. (2006) and Wu and Xu (2005).

Although currently available methods and modified binders for resistance to rutting of asphalt mixture have been studied in a laboratory, a number of shortcomings must be noted, specifically with respect to considerations of asphalt gradations, temperature, compacting degree and interacting micro-mechanism. In addition, various factors affecting high-temperature stability of an asphalt mixture are so complicated that some researchers have found it either difficult to apply onsite or inappropriate to apply during high temperature days. In this paper, rutting resistance in cases of different gradations, temperatures, wheel rolling times and contents of a new anti-rutting agent (ARK-1) were analyzed in a laboratory; and comparison studies between ARK-1 and styrene–butadiene–styrene (SBS) were conducted in Baotian freeway, which was constructed in May 2007.

Material and Experimental Method

Asphalt Binder/Asphalt Mixture Tests. The asphalt binder used in this study is AH-90 produced by Lanlian Oil Refinery. Details of the specification tests are available in the Standard Test Methods of Bitumen and Bituminous Mixtures for Highway Engineering ("PRC Industry Standard", 2000). All aggregates are collected from a stone factory in Wushan County where a large amount of basalt has enough strength to meet requirement of Baotian freeway construction between stations 34+808 and 94+560. Coarse, fine aggregate tests are given in Test Methods of Aggregate for Highway Engineering ("PRC Industry Standard", 2005).

Table 1. Aggregate gradations of mixture.

Items	Mesh passing (%)										
Sieve size (mm)	19	16	13.2	9.5	4.75	2.36	1.18	0.6	0.3	0.15	0.075
No. 1		100	95	76	53	37	26	19	15	10	6
No. 2		100	92	80	62	48	36	26	18	14	8
No. 3	100	90	76	60	34	20	13	9	7	5	4
No. 4	100	94	82	62	32	21	18	18	15	13	11
No. 5	100	88	70	48	23	15	14	14	12.4	11	9

Five aggregate gradations of mixture are shown in Table 1. In these gradations, No. 1, No. 2 and No. 3 are continuous-graded mixtures, and No. 4 and No. 5 are gap-graded mixtures. The Marshall test plays an important role on asphalt mixture design because this test can control volumetric parameters and optimum asphalt-aggregate ratio of asphalt mixture. However, the optimum asphalt-aggregate ratio by the Marshall test is often higher than that ratio by wheel tracking test. In this study before the wheel tracking tests are conducted, samples of five gradations of asphalt mixture are made in case of optimum asphalt-aggregate ratio by the Marshall test.

Wheel Tracking Test. The specimens are mixed in the optimum asphalt-aggregate ratio and compacted by the Plate Compactor according to the Marshall test. Demension of the specimen is 300×300×50 mm (three specimens per parallel control test) with different ARK-1 contents with ranges of 0% ~0.4 % (weight ratio of ARK-1 to aggregate). Slab of the compactor may pass on 24，36 and 48 times back and forth and 9 kN load pressures during compaction. In this study, the wheel tracking instrument was made by Changsha Yaxing Digital-control Co., Ltd. All wheel tracking tests are conducted at 55，60，65 and 70^{o}C (0.5^{o}C accuracy) and 0.7 MPa (0.05 MPa accuracy) to measure Dynamic Stability (DS) and Rutting Depth (RD).

Micro-mechanism Analysis of ARK-1

ARK-1 is a kind of black solid particle that contains more than 95% fiber polymers and less than 5% fillings. Its density is about 0.91 ~ 0.98 g/cm^3, and its size is about 2~4 mm by using sieving test. Its melting index (MI) is more than 0.1 g/10 mins to ensure enough melt flow rate at 190^{o}C and under 50kN load pressures. The softening point of ARK-1 is between 150^{o}C and 170^{o}C.

The interacting micro-mechanism, which ARK-1 can improve anti-deformation performance of asphalt mixture under high temperature conditions, may explain that: (1) as mixing temperature is higher than melting point of ARK-1, the cohesive flowing ARK-1 particles have a strong plasticity so that various shapes filling voids among mineral aggregates can be formed using an action of interlocked-denseness force. Once temperature is lower than the glass transition temperature of ARK-1, relative displacements among mineral aggregates will be limited; (2) if there are

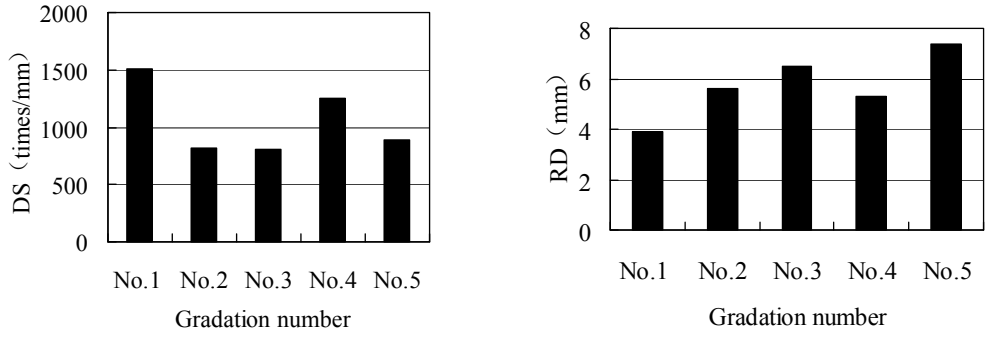
enough ARK-1 particles in the hot mix asphalt, the mixture will be interconnected to grow a three-dimensional mesh structure. Aggregates in this mixture can be firmly restricted within the mesh structure so that the whole strength of HMA will be improved; (3) if heated up to the melting point of ARK-1 or higher temperature, mineral aggregates will be entirely enveloped by the flowed, infiltrated ARK-1 in mixture; if cooled, stiffness and strength of mixture will be gradually increased on the interface of aggregates and ARK-1. Consequently, ARK-1 can also be regarded as a hot melt adhesive in HMA; (4) as a new polymer material, ARK-1 can absorb light oil to increase the viscosity of asphalt binder and the stiffness modulus of mixture so that the resistance to rutting of asphalt mixture will be greatly improved.

Experiment Results and Discussion

Influence of Gradations on Rutting. Different gradations of asphalt mixture may largely influence pavement performance at high temperatures. Results of the wheel tracking test to five gradations of mixture without ARK-1 are presented in Figure 1. Compared to the continuous-graded and gap-graded mixtures, resistance to rutting of two gradations has no obvious differences.

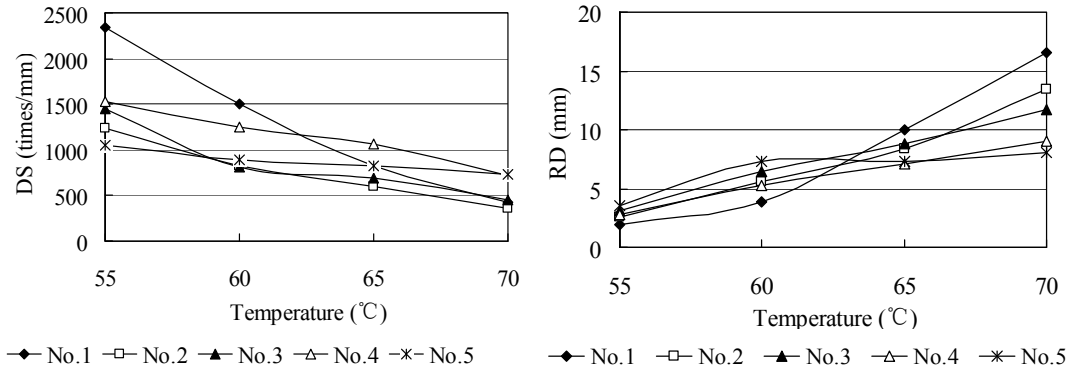
In earlier research on resistance to rutting of asphalt mixture, some test results show that DS of the fine and coarse aggregate mixtures has no obvious differences (Liu and Dai, 2004; Aho et al., 2001), or DS of the coarse aggregate mixture is higher than that of the fine aggregate (Peng et al., 2005). In Figure 1, it is deduced that the fine aggregate mixture may be superior to the coarse aggregate mixture in rutting resistance.

Influence of Temperature on Rutting. Asphalt binder has temperature-sensitive behavior across most of the in-service temperature range. DS and RD of five gradations at 55℃, 60℃, 65℃ and 70℃ are shown in Figure 2. Decreasing rate of DS and increasing rate of RD with the increasing temperature to the gap-graded mixtures are obviously slower that those of DS and RD to the continuous-graded mixtures. It is indicated that resistance to rutting of the gap-graded mixtures may be superior to that of the continuous-graded mixtures as temperature is higher than 60°C.



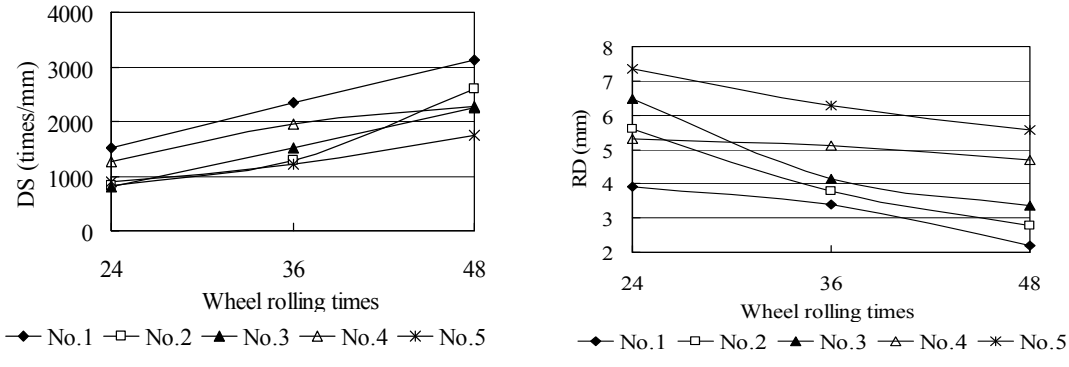
(a) (b)

Figure 1. Influence of different gradations on DS and RD.



(a) (b)

Figure 2. DS and RD of five gradations versus temperature for five gradations.



(a) (b)

Figure 3. DS and RD of five gradations versus wheel rolling times.

With the temperature increase, asphalt binding force for mineral aggregate particles will be gradually reduced due to viscosity declining significantly. Resistance to deformation of mixtures may be mainly determined by mineral aggregate particles. Consequently, resistance to rutting of the fine aggregate mixture may be worse than that of the coarse aggregate mixture if temperature is beyond 60°C.

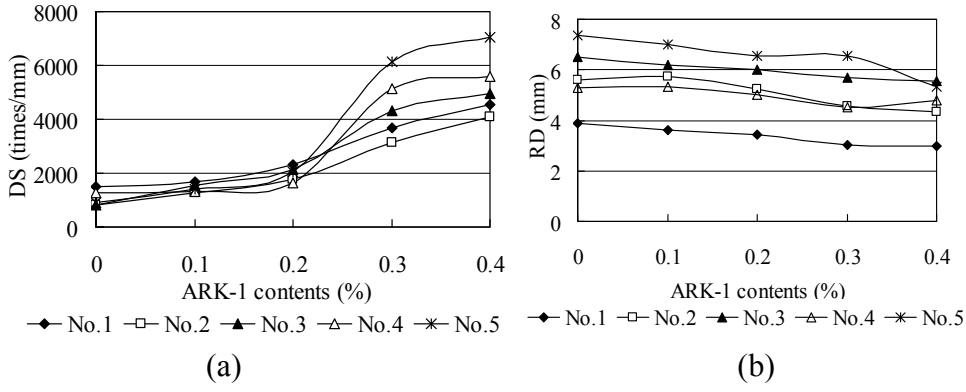
Influence of Wheel Rolling Time on Rutting. According to the Standard Test Methods of Bitumen and Bituminous Mixtures for Highway Engineering ("PRC Industry Standard" T0703), the density of a sample slab that was rolled over 24 times back and forth will be approximately equal to a sample obtained by Marshall compaction. However, due to the two different methods for mixture samples, the density of the wheel rolling sample passed over 24 times back and forth is often less than that of the Marshall sample. DS and RD of five gradations varying with wheel rolling times are shown in Figure 3.

With an increase of wheel rolling times, DS of the continuous-graded mixture will increase more obviously than that of the gap-graded mixture, and RD of the continuous-graded mixture will decrease more quickly than that of the gap-graded

mixture. Therefore, it indicates that the gap-graded mixture will be affected slightly by wheel rolling times.

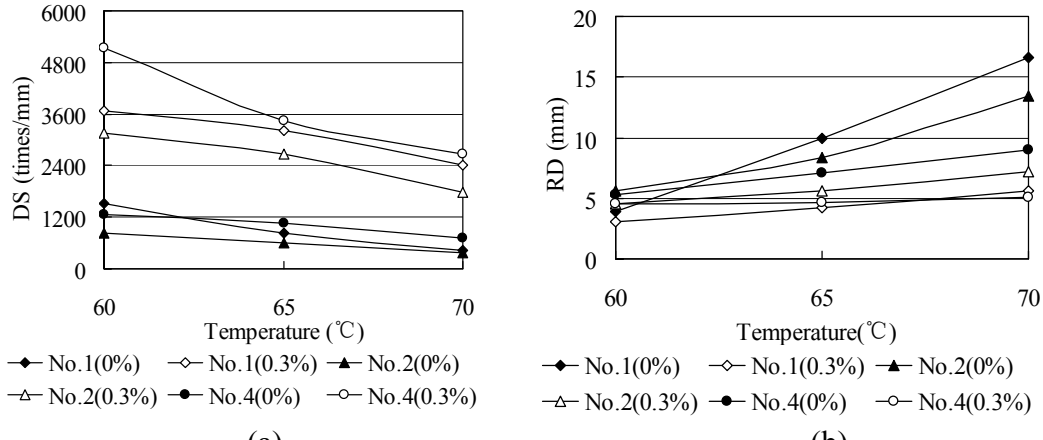
Influence of Anti-Rutting Agent on Rutting. For ensuring more effective asphalt binder absorbed by mineral aggregates, asphalt aggregate ratio of mixture with ARK-1 may be added 0.2%~0.4% than the optimum asphalt-aggregate ratio given by the Marshall test. DS and RD varying with ARK-1 contents (weight ratio of ARK-1 to aggregate) in cases where of repeated 24 times of wheel rolling at 60°C are shown in Figure 4. When ARK-1 contents are beyond 0.2%, the DS of the gap-graded and coarse aggregate mixture is larger than the DS of the continuous-graded and fine aggregate mixture. Generally, it is not economical to improve high-temperature stability of a mixture, only to increase ARK-1 contents. The optimum ARK-1 contents are about 0.3%~0.5% for improving high-temperature performance of asphalt mixture.

In order to verify ARK-1 effect on resistance to rutting of asphalt mixture beyond 60°C, taken No. 1, No. 2 and No. 4 as analysis objects, DS and RD of mixtures with 0% and 0.3% ARK-1 at 60°C, 65°C and 70°C are shown in Figure 5. Since



(a) (b)

Figure 4. DS and RD of five gradations versus ARK-1 contents.



(a) (b)

Figure 5. DS and RD with 0% and 0.3% ARK-1 contents versus temperature.

interlocked-denseness force and mesh structure in mixture with ARK-1 formed may not be easily changed under load pressure as temperature is lower than mixing temperature, resistance to rutting of mixtures with ARK-1 can be improved dramatically, especially that of the gap-graded mixtures.

Field Tests and Application

The Baotian freeway located in the Gansu section is one of 10 important projects established by State Department of China in Development of the West Regions. This route is 91.114 km long from the Niubei village, where the juncture of Gansu and Shanxi province to the Ganquan town in Maiji area of Tianshui country. In BTLM2 contract between stations 34+808 and 94+560, materials and thickness of pavement structure are shown in Table 2. In the upper layer, SBS (I-C) and ARK-1 are provided by Gansu Yilong asphalt and Tianhong Communication Tech. Co. Ltd., respectively. SBS (I-C) and ARK-1 tests as shown in Table 3 are conducted by the Highway Engineering Detection Centre of Chang'an University. The Marshall Immersion Test, Indirect Tension Test under freeze-thaw (AASHTO T283) and Wheel Tracking Test are carried out to evaluate water stability and high-temperature stability of a mixture in the laboratory of Lanzhou University of Technology, as shown in Table 4.

High-temperature and water stability of asphalt mixtures with 6% SBS and 0.3% ARK-1 are very approximate, which makes it difficult for pavement engineers to decide which of the two materials is the best. As cost analysis of the two materials in the upper layer is considered, asphalt mixture with 6% SBS between stations 35+460 and 35+960 and 0.3% ARK-1 between stations 35+960 and 36+460 will cost about ￥334,550 and ￥308,200, respectively. Then, mixtures with 0.3% ARK-1 were applied in the upper layer of the pavement to the Baotian freeway in May 2007.
ARK-1 is applied more frequently on highway construction due to its convenient storage and transportation, simple construction technology as well as lower costs than SBS-modified asphalt binder. For example, asphalt pavement with 0.4% ARK-1 in the upper layer had been used in a steep slope section of a road from Lanzhou to Lintao city. Many national roads located in the Qinghai province, such as the 109, 214, 215 and 227 routes, had utilized ARK-1 to improve the high-temperature stability of asphalt pavement.

Table 2. Scheme of pavement structure.

Structure layer	Material	Thickness (cm)
Upper layer	AC-16	4
Middle layer	AC-20	6
Bottom layer	AC-25	8
Upper base	Cement stabilized gravel (5.5% cement)	25
Lower base	Cement stabilized gravel (4% cement)	20

Table 3. Test results of SBS (I-C) modified binder and ARK-1.

Binder types	Experiment items		Unit	Test result	Test method
SBS (I-C)	Penetration	15°C, 100 g, 5 s	0.1 mm	36	ASTM D5
		25°C, 100 g, 5 s		79	
		30°C, 100 g, 5 s		124	
	Penetration index (PI)		-	0.791	
	Ductility (5°C, 5 cm/min)		cm	42	ASTM D113
	Softening point ($T_{R\&B}$)		°C	63.0	ASTM D36
	Kinematic viscosity (135°C)		Pa·s	1.23	ASTM D4402
	Flash point		°C	246	ASTM D92
	Solubility		%	99.8	ASTM D2042
	Elastic recovery		%	86	ASTM D6084
	Segregated softening point difference		°C	2.1	ASTM D5976
	Density (15°C)		g/cm^3	1.028	ASTM D70
	RTFOT (163°C, 75min)	Mass loss	%	–0.11	ASTM D2872
		Penetration (25°C)	%	65.6	ASTM D2872 D5
		Ductility (5°C)	cm	21	ASTM D2872 D113
ARK-1	Density		g/cm^3	0.94	ASTM D70
	Melting index (MI) 190°C, 50 kN		g/10mins	9.2	ASTM D1238
	Softening point		°C	164	ASTM D36
	Particle size		mm	2–4	-

Table 4. Results of water stability and wheel tracking test.

Upper layer	Admixture contents (%)	Residual stability (%)	Test strength ratios (%)	DS (times/mm)
AC-16	6% SBS	89.3	90.4	6780
	0.3% ARK-1	89.8	90.5	6650

Conclusions

Different gradations of mixture, temperature, wheel rolling times and contents of ARK-1 were measured by the Wheel Tracking Test. In order to verify the effectiveness of ARK-1, a comparison between SBS and ARK-1 was conducted on the Baotian freeway. The following conclusions may be drawn:

(1) The resistance to rutting of the fine aggregate mixture may be higher than that of the coarse aggregate mixture in cases where the asphalt binder and temperature were the same.

(2) As temperature is equal or lower than 60°C, no obvious differences between the continuous-graded and gap-graded mixtures can be observed in the Wheel Tracking Test. Since temperature is higher than 60°C, resistance to rutting of the gap-graded mixture is superior to that of the continuous-graded mixture.

(3) The resisting rut performance of the continuous-graded mixture is more sensitive to wheel rolling times than that of the gap-graded mixture. It is indicated that controlled compaction degree of continuous-graded mixture is a primary factor in reducing rut phenomenon of pavement.

(4) ARK-1 can dramatically improve resisting rut capability of mixture, especially gap-graded mixture. Comparing ARK-1 with SBS in pavement performance, the two materials have a similar effect on high-temperature and water stability.

Acknowledgment

This research is supported by the Fund of Natural Science of Gansu Province (project account: 0803RJZA011and 0916RJZA024) and by the Fund of Natural Science of Zhejiang Province (project account: y1090039). The authors thank BTLM2 Project Department of Baotian freeway for supplying asphalt binder, aggregate materials and assistance with collection of samples, Guo Rui and Wang Genqiang with their help with testing.

References

Aho, B. D., Vavrik, W. R., and Carpenter, S. H. (2001). "Effect of flat and elongated coarse aggregated on field compaction of hot-mix asphalt." *Transportation Research Board.*

Ali, T., and Burak, S. (2004). "Determination of fine aggregate angularity in relation with the resistance to rutting of hot-mix asphalt." *Construction and Building MATERIALS*, 155-163.

Cheuk Ching, W., and Win-gun, W. (2006). "Effect of crumb rubber modifiers on high temperature susceptibility of wearing course mixtures." *Construction and Building Materials*, 1741-1745.

Gandhi, P. M., and Gallaway, B. M. (1970). "The effect of aggregate characteristics on the behavior of asphalt mixtures under static and dynamic loading conditions." *American Society for Testing and Materials.*

Halit, O., Atakan, A., Sureyya, T., and Fazil, C. (2007). "Laboratory performance comparison of the elastomer-modified asphalt mixtures." *Building and Environment*, 1-8.

Huang, B., Mohammad, L. N., Graves, P. S., and Abadie, C. (2002). "Louisiana experience with crumb rubber-modified hot-mix asphalt pavement." *Transportation Research Record. 1798*, Transportation Research Board, Washington, D.C., 1-13.
Liu, H. Y., and Dai, J. L. (2004). "Effect of different gradations on asphalt mixture resistance to rut." *J. of Chang'an Univ. (NSE)*, 24(5), 11-15.

Okan, S., Hong-Joong, K., Mang, T., and Bouzid, C. (2006). "Comparison of rutting resistance of unmodified and SBS-modified Superpave mixtures by accelerated pavement testing." *Construction and Building Materials*, 1-9.
Peng, Y. H., Sha, A. M., and Liao, Z. G. (2005). "Experimental analysis on influence factors in Wheel Tracking Test." *Highway of Chinese and Foreign*, 25(1), 18-22.

Soon-Jae, L., Serji, N. A., Bradley, J. P., and Kwang, W. K. (2007). "Laboratory study of the effects of compaction on the volumetric and rutting properties of CRM asphalt mixtures." *J. Materials in Civil Engineering*, 19(12), 1079-1089.

Wu, S. S., and Xu, X. J. (2005). "A study on performances of asphalt mixtures intermingled with anti-rut agent PR PLASTS." *Highway*, 15(1), 156-159.

Xiao, F. P., Serji, A. M., and Hsein, J. C. M. (2007). "Rutting resistance of rubberized asphalt concrete pavements containing reclaimed asphalt pavement mixtures." *J. of Materials in Civil Engineering*, 475-483.

Xiao, Q. Y., Rui, S. Q., Wang, H., and Zheng, T. (2006). "Research on the effect of the PR PLASTS contents on asphalt." *J. of Wuhan Univ. of Tech.*, 36-39.

Effect of Activators on Mechanical Properties of Steam-Cured Fly Ash Concrete

Zhimin He[1,2] and Junzhe Liu[1]

[1] Faculty of Civil Engineering, Ningbo University, Ningbo 315211, China, Email: hezhimin@nbu.edu.cn.
[2] School of Civil Engineering and Architecture, Central South University, Changsha 410075, China

Abstract

In precast concrete elements manufacturing,steam-cured concrete incorporating 30% fly ash encountered the problem of a too low demoulding compressive strength. To resolve it, this paper developed a new steam-cured concrete (AFSC) incorporating fly ash and a chemical activator. Experiments were conducted to investigate the mechanical properties of AFSC. The corresponding mechanism was also discussed by testing the microstructure of concrete. Results indicate that the demoulding compressive strength of AFSC can meet production requirements, and compressive and flexure strength of AFSC at later ages increase well. Compared with that of ordinary steam-cured concrete, AFSC has a higher tensile strength, and the capability of AFSC to resist cracks is enhanced remarkably. At an early age, addition of the chemical activator can distinctly accelerate the extent of hydration of the fly ash cement systems, and thus the microstructure of concrete becomes denser.
Key Words: activator; steam curing; fly ash; mechanical properties; microstructure

Introduction

Steam-cured concrete is widely used in producing railway prestressed precast elements in China. The ordinary steam-cured concrete is usually designed with a 50MPa～60MPa compressive strength and a high-volume cement, which typically is 480 ~ 500 kg/m^3. Due to high cement content, precast concrete elements have many problems, such as higher hydration heat, bigger pore size and porosity, more interface micro-cracks and vulnerable to the impact of alkali aggregate reaction. Some researchers (Ho and Chua 2003; Liu and Xie 2001) have reported that the concrete with fly ash and ground granulated blast furnace slag or chemical activators has a high early strength and better later compressive strength-gaining rate than the

control concrete. However, pozzolanic reaction is a slow process, and its contribution to strength occurs only at later ages, so the early strength of concrete will be significantly reduced if a large amount of fly ash is used. In precast concrete elements manufacturing, steam-cured concrete incorporating 30% mineral admixtures encountered the problem of a low demoulding compressive strength. A number of studies had been carried out on the activation of fly ash using chemical activators (Fernandez and Palomo 2005; Ma et al. 1995). These studies involved using different activating methods including alkali activation and sulfate activation.
This paper firstly examines whether significant improvements in the reactivity of fly ash can be realized through the use of some types of chemical activator. Secondly, mechanical properties and microstructure of fly ash concrete with chemical activators has been investigated under steam curing. The results of this study may be useful to investigate the utilization of high-volume of fly ash in railway prefabricated concrete.

Experimental

Materials. The cement used in experiments was ordinary Portland cement with grade P.O 42.5, and its strength properties are listed in Table 1. FA I came from Yaomeng Power Plant, and its physical properties are listed in Table 2. The sand came from Ruzhou, and its fineness modulus was 2.82. The coarse aggregates were crushed pebble. Diameter of coarse aggregates were all confined in 5 mm～20 mm and crushed index was 5.1%. A superplasticizer of sulfated naphthalene formaldehyde base made in the Shenxiang Chemical Plant was used in the mix, whose commercial name was NF-2. Mixing water was clean water. The design grade of concrete strength after demoulding is 50 MPa, and 60 MPa at 28 days. Experimental proportion and related parameters are shown in Table 3. The total amount of cementitious material is 450 kg/m^3.

Table 1. Strength of cement.

Index	Flexural strength /mpa		Compressive strength/mpa	
	3d	28d	3d	28d
Value	4.85	9.78	22.4	49.4

Table 2. Physical properties of cement.

Index	Fineness (80μm)/%	Surface area/m^2/kg	Density /g/cm^3	Setting time /(h：min)	
				Initial setting	Final setting
Value	2.4	382	3.1	2:45	3:50

Table 3. Physical properties of FA.

Index	Fineness (45μm)/%	Water content /%	Water demand ratio /%	Ignition loss/%	Density /g/cm^3	Surface area/m^2/kg
Value	1.9	0.25	92	2.2	2.22	597

Table 4. Experimental proportion and related parameters.

NO	*cementitious material components/%*			*SP*	*NF-2*	*W/B*
	C	FA	Activator			
Z_1	70	26	A/4*	0.33	0.07	0.27
Z_2	70	26	B/4	0.33	0.07	0.27
Z_3	70	26	C/4	0.33	0.07	0.27
Z_4	70	26	D/4	0.33	0.07	0.27
Z_5	70	26	E/4	0.33	0.07	0.27
Z_6	70	26	F/4	0.33	0.07	0.27
Z_7	100	-	-	0.36	0.10	0.295
Z_8	70	26	B/4	0.33	0.10	0.27
Z_9	70	30	-	0.33	0.15	0.27

Note: A is type of activator, 4 is the proportion of activator to total weight of cementitious material.

Specimen Preparation and Test Methods. The concretes were prepared in the laboratory using a mixer and then vibrated for 3 min. The treatment temperature and its duration, together, have a profound effect on the evolution of hydration reaction and the products. In this study, the heat-treatment cycle has a total duration of 13 h, the preheating duration and heating duration is 2 h, the treatment duration is 8 h, and the cooling duration is 1 h. The treatment temperature chosen is 60±5℃. After steam curing, the specimens were demoulded, and some were used to measure the compressive strength (fcuz) at once. The others were put in the standard curing room. Its 28-day compressive strength is named as fcuz, 28. Specimen for cubic compressive strength has a measurement of 100×100 ×100 mm. Specimen for prism and elastic modulus have a size of 100×100×300 mm. Specimen for flexural strength have a size of 100×100×400 mm. Mechanical properties of concretes are determined according to the Chinese National standard GBJ 81-1985.

Discussion of Results

Type of Chemical Activator Optimization. Figure 1 illustrates the effect of a variety of chemical activators on the steam-cured fly ash concrete compressive strength development. Results indicate that the addition of activator A has a certain increase effect on 28-day compressive strength, but has no significant effect on demoulding compressive strength, and 90-day compressive strength decrease?. Activator C has a certain improvement effect on 28-day and 90-day compressive strengths, but this improvement is not significant. Activator D has no significant effect on both early and later term compressive strength. The large increase in demoulding strengths is particularly significant for E-type activators, but has no significant effect on later term compressive strength. For activator F, there is a clearly significant improvement of strength after 28 days, however, the improvement of demoulding and 90-day strength is not significant. Compared with the other five kinds of activators, it can be found that activator B has the more pronounced improvement on steam-cured fly ash

concrete compressive strength development. Through multiple studies of comparative test, activator B is the most efficient activator.

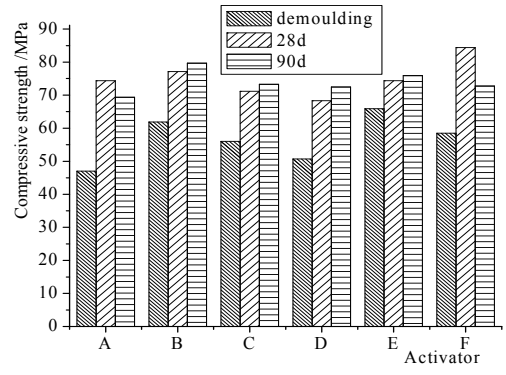


Figure 1. Effect of activator on compressive strength of concrete.

Mechanical Properties. Figure 2 illustrates that the results of compressive strength of concrete varies with age, and results indicate that steam-cured concrete (Z_8) together with FA and chemical activators has as high of a demoulding compressive strength as the the blank concrete (Z_7). Its 28-day compressive strength is more than 80 MPa, and later compressive strength keeps continual growth, especially if its 90-day compressive strength already exceeds that of the blank concrete. However, the steam-cured concrete only with FA has the lowest demoulding and later compressive strength, showins that incorporating chemical activators has a significant improvement in the compressive strength of steam curing concrete with FA. From Figure 3, it can be seen that incorporating chemical activators has beneficial effects because it enhances drastically the flexural strength at all curing ages. Concretes (Z1) made with chemical activators and FA show higher early demoulding and later flexural strength. Its flexural strength at three ages are higher than concretes (Z2) made with FA but without chemical activators 7%、10%、and 2%.

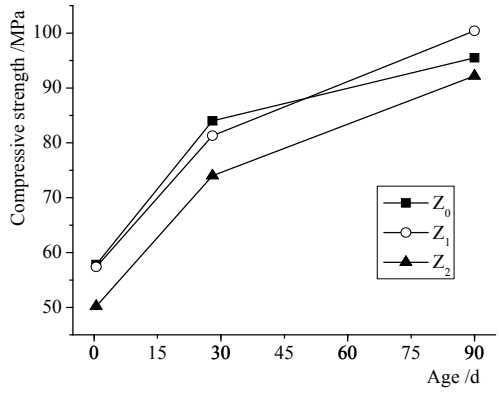


Figure 2. Compressive strength of of concretes varies with age.

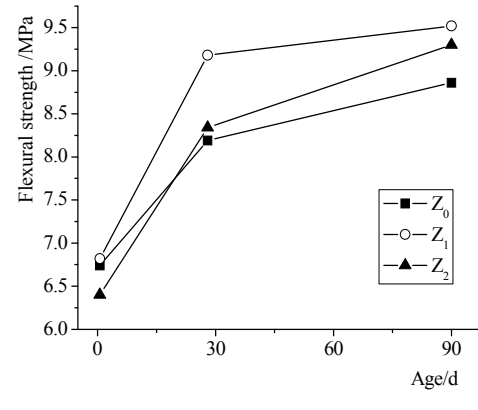


Figure 3. Flexural strength concretes varies with age.

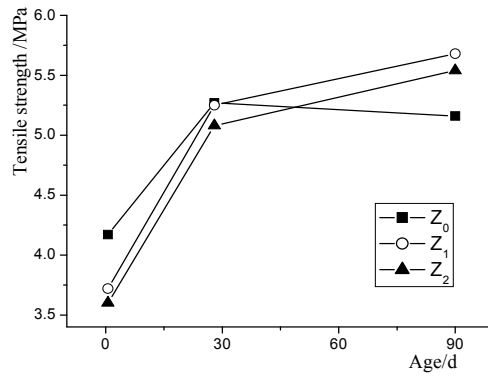


Figure 4. Tensile strength of concrete varies with age.

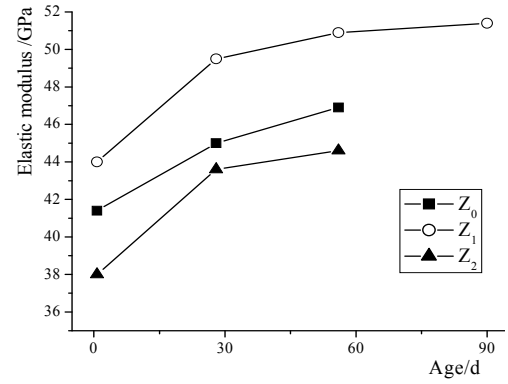


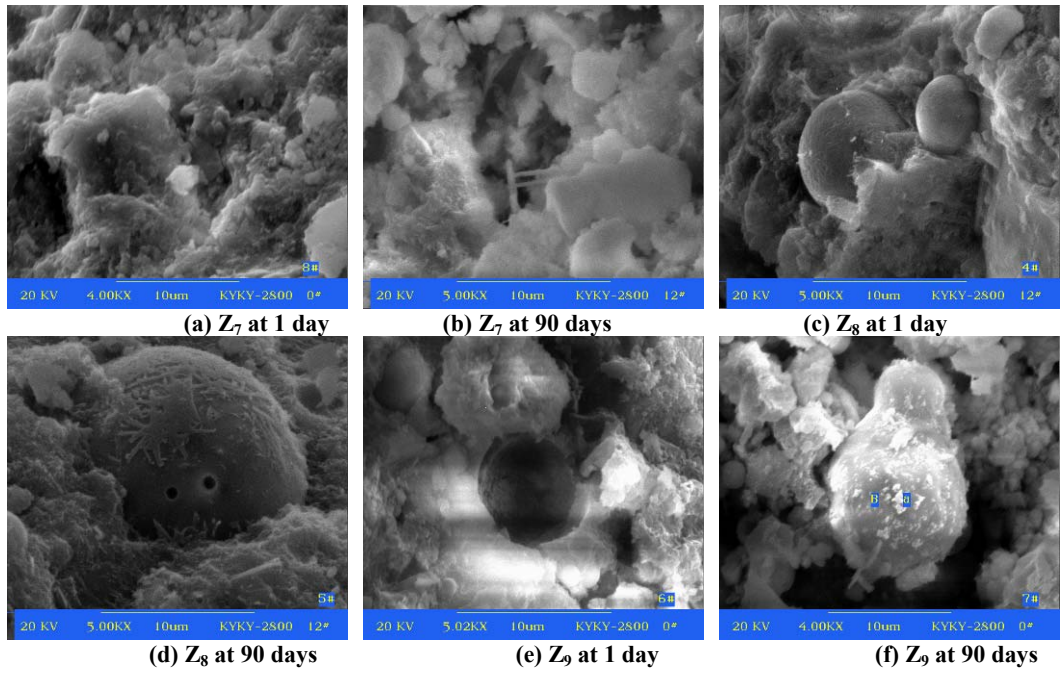
Figure 5. Elastic modulus of concrete varies with age.

Figure 4 shows the results of tensile strength of concrete that varies with age. The later strength of blank steam-cured concrete has a lower increment degree compared with steam-cured concrete incorporating mineral admixtures and activator, and its increment of tensile-splitting strength inverted to some extent.

The results discussed above show that the fly ash-cement system can improve the adaptability of steam curing. Steam-cured concrete containing FA has a high strength-gaining ratio and a high demoulding strength in the presence of chemical activators. The development of later compressive strength is the same as the moist-cured concrete.

Modulus of Elasticity. Elastic modulus of concrete is an important design parameter of concrete structure. Figure 5 illustrates the results of elastic modulus of concrete that varies with age. From Figure 5, it can be seen that elastic modulus of all concrete after demoulding are more than 38 GPa which can meet the production process and quality requirements of the sleeper, and increase with age. As to the increasing trend, the growth of elastic modulus of concrete is rapid before the age of 28 days, and gradually slows in the later term. The elastic modulus value for concrete with chemical activators is significantly higher than that of the other concretes, and this shows that chemical activators have remarkable improvements on the elastic modulus of steam-cured concrete.

Microstructure Scanning electron micrograph of steam-cured concrete is shown in Figure 6. Most hydration products are listed in Figure 6a, which has an integrated shape, but less fibrous, rod-like product. From Figure 6b, it can be seen that the pure cement steam-cured concrete has relatively large-sized of crystals at 90 days, and its internal structure is not very dense for which there are still some holes in existence. This is consistent with the literature (Peng and Huang 2008) conclusion that Portland cement concrete forms a coarse-grained structure after heat curing.



(a) Z_7 at 1 day (b) Z_7 at 90 days (c) Z_8 at 1 day

(d) Z_8 at 90 days (e) Z_9 at 1 day (f) Z_9 at 90 days

Figure 6. Micrograph of SEM in different ages.

Figures6c and 6d are micrographs of HPSC. From Figure 6c, it can be seen at 1 day that most of hydration products have been generated, and the surface of fly ash particles has some etching marks which are confirmed as hydration products by electron spectroscopy analysis. Therefore, the degree of hydration of fly ash must be increased in the presence of chemical activators because the surface is activated by the destruction of the structure of the glass. This is consistent with the results that mix Z8 has a higher early demoulding strength. From Figure 6e, the surface of fly ash particles is smooth, and the degree of hydration of fly ash is very low.

In Figure 6d, there exist a large amount of network-like structures in the internal structure of concrete, which form a compact microstructure. The distribution of hydration products is relatively uniform and etching of glassy materials is more clearly visible.

Conclusions

When compared with the other activators, the chemical activator B is more effective in increasing the compressive strength at both early and later ages. Thus, the use of activator B is advantageous. AFSC(Z8) incorporating 30% fly ash and chemical activator B has a high demoulding strength and later strength-gaining rate. The introduction of chemical activator B can accelerate the pozzolanic reaction of fly ash, therefore, increasing the early strength of the concrete containing fly ash, and remarkably improve microstructure and reduce interconnected capillary pores in concrete.

References

Fernandez, J. A., and Palomo, A. (2005). "Composition and microstructure of alkali activated fly ash binder: effect of the activator." *Cement and Concrete Research,* 35(10), 1984-1992.

Gonen, T., and Yazicioglu, S. (2007). "The influence of mineral admixtures on the short and long-term performance of concrete." *Building and Environment*, 42(8), 3080-3085.

Ho, D. W. S., and Chua, C. W. (2003). "Steam-cured concrete incorporating mineral admixture." *Cement and concrete research,* 33(4), 595-601.

Liu, B. J., and Xie, Y. J. (2001). "Some factors affecting early compressive strength of steam curing concrete with ultra-fine fly ash." *Cement and Concrete Research,* 31 (10), 1455-1458.

Ma, W. P. Liu, C. L., and Brown, P. W. (1995). "Pore structures of fly ash activated by $Ca(OH)_2$ and $CaSO_4.2H_2O$." *Cement and Concrete Research,* 25 (2), 417-425.

Peng, G. F., and Huang, Z. S. (2008). "Change in microstructure of hardened cement paste subjected to elevated temperatures." *Construction and Building Materials,* 22(2), 593-599.

Effect of Sisal Fiber on Anti-slide Performance for Asphalt Pavement Surface

Xiaowen Lei[1], Xiaoming He[1], Peng Guan[1] and Pengfei Deng[2]

[1] Institute of Traffic, Wuhan Polytechnic University, P. R. China, 430023; Email: guanpeng1210@163.com
[2] School of Economics and Management, China University of Geosciences, P. R. China, 430074

Abstract

In order to improve the anti-slide performance for an asphalt pavement surface, the anti-sliding property of pavement was analyzed in this paper; the fiber additives in asphalt mixture were compared. Furthermore, certain groups of comparative experiments were conducted by adding the different length and different quantity of the sisal fiber toward the bituminous mixture and showing that the sisal fiber is the best asphalt mixture additive to improve the anti-slide performance for pavement surface. The best length and the best quantity of the sisal fiber in the bituminous mixture were determined.
Key Words: Sisal fiber; pavement surface; anti-slide performance; comparative experiment

Introduction

By the end of 2008, there was a total of 54,000 km of expressways in China, ranking the second in the world. Asphalt concrete is used as the paving material for 90% of the pavement. Meanwhile, with increasingly frequent traffic accidents, the safety of highway transportation has been a worldwide problem. The statistics from the World Health Organization shows that over 1.2 million people die from highway traffic accidents per year. On an average, two are being killed every minute, and the mortality in developing countries is higher than that in developed countries. As a developing country, China's road traffic safety situation is gloomy. There were 378,781 road traffic accidents in 2006, of which at least 89,455 people died and the direct economic loss hit 1.49 billion Yuan. The cause of traffic accidents are diverse, including human factors, vehicle factors, pavement factors, climate factors, objective and subjective reasons, etc. The insufficiency of anti-skidding of pavement is one of the important factors that influences the occurrence of traffic accidents.

Analysis of the Mechanism of Anti-Sliding of Pavement

Superior anti-skidding of pavement could supply enough frictional resistance between car tires and the road surface (Kulakowski 1990) and then ensure breaking at a safe distance. The magnitude of this force relates to tire characteristics, and it is mainly determined by the structure property of pavement, conditions of road surface and the driving speed.

The Structure Property of Pavement. Microtexture and macrostructure are included in structure characteristics of pavement texture (Chenye 2009). The tiny structure refers to the one that combines the horizontal surface of the road aggregate 0~0.5 mm with vertical direction 0~0.2 mm. The peak response of micro-structure necessary to protect the aggregate could penetrate water film of pavement in wet condition, and keep close contact between tire and road surface. According to the requirement of anti-slide performance, the best peak height is 0.01~0.1 mm. Frictional resistance of pavement is directly influenced by the density or distance of microtexture. The more of the rough density, the more touch point between tire and road surface, and the stronger adhesive force will be.

Drainability and gap among the aggregate particles of pavement are shown by their macrostructure. It is 0.5~50 mm in horizontally, and 0.2~10 mm in vertically. The drainability of asphalt anti-sliding surface is an important target to judge its macrostructure. Under the influence of high-speed driving, anti-sliding performance of wet pavement depends chiefly upon drainability of asphalt surface, to avoid the drift-slip phenomenon because of the continuous water film between tire and road surface. The greater the macrostructure is, the better drainability will be. Thus, macrostructure has an effect on anti-sliding of pavement under high-speed driving.

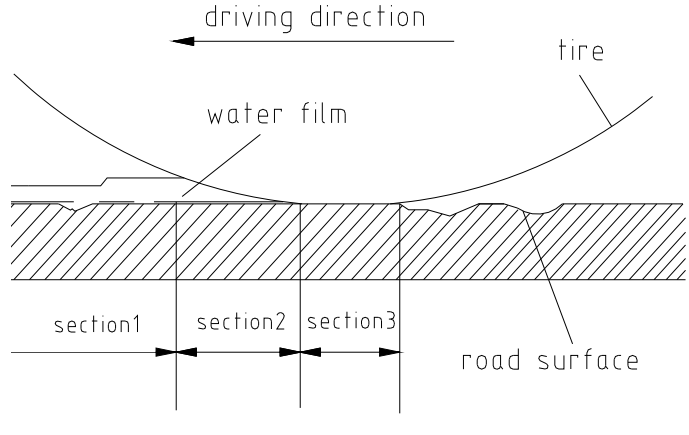


Figure 1. Interaction area.

Humidity of Road Surface. In general, it is enough to meet the requirements of safe driving about the anti-sliding of pavement when the road is dry. But when road conditions are wet, the anti-sliding of pavement will decline sharply (Lipengwei 2009). It could be explained in terms of interface state between tire and road surface. Contact interface is divided into three parts, as shown in Figure 1. Section 1 is the pavement in front of tire having continuous water film. Section 2 is the area with thin discontinuous water film under the stress from tire. In Section 3, water is squeezed away, and the structure between the tire and road surface is dry, then adequate anti-sliding will be provided.

Therefore, in order to guarantee that the road surface water could be removed immediately when vehicles are traveling at a high-speed, increase the depth of groove and the number of tires, on the other hand ensure there is enough drainability provided by rough road construction.

Speed of Driving. The anti-sliding performance of new pavement depends chiefly upon microtexture and macrostructure of pavement itself. But the "congenital" level of anti-slide is not consistent when the road is open to traffic, rather evolved attenuation trends by rolling effect of traffic load with different speed. The size of the decay rate is relating to speed, material properties, and composition of pavement anti-slide surface (as shown in Figure 2). When the attenuation reached a certain extent, anti-sliding would be reduced. Because the peak of the exposed aggregate surface of pavement was gradually worn by tires with the cumulative impact and friction under traffic load, aggregate surface is smooth grinding, then reduces the average shape coefficient of aggregate and adhesion between tire and road surface. The changes of surface micro-structure finally resulted in the decline of the level of anti-sliding of pavement.

A: both coarse and fine textures are rough; B: coarse texture is rough, fine one is smooth; C: coarse texture is smooth, fine one is rough; D: both coarse and fine textures are smooth.

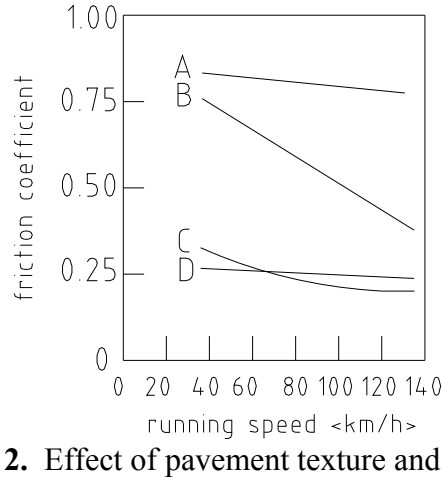


Figure 2. Effect of pavement texture and driving.

For the raw material, the quality of mineral aggregate, asphalt binder, mineral aggregate shape, and gradation will directly influence the anti-sliding of pavement. At the same time, as the additives, fiber has the effect on improvement of the anti-sliding.

Election of Fibers

It is an important research subject of fiber types and materials for engineering experts at home and abroad. According to its type, fiber can be generally classified into hard fibers and soft fibers. Hard fiber is produced by different processes include pulling, drawing, rolling and cutting; soft fiber is made from synthetic fibers (Luzhao 2005), which can be classified as glass fiber, Nylon fiber, polypropylene fiber, sisal fiber, etc. After the application research, tracking, and survey of different types of fibers, the summary is as follows:

Steel Fiber. Although Steel fiber has the pavement performance characteristics of high strength, thermostability, flexibility, high orientation, and so on, its metal corrosion is the root of its weakening function (Elsaigh et al. 2010 Van Dam 1995). In addition, it is bad for wear and tear on the wheels because the coefficient of friction of metal is less than concrete. Therefore, in recent years, the popularization and application of steel fiber in concrete pavement has decreased.

Synthetic. Synthetic steel fiber reinforced concrete is the second post-developed of soft-fiber reinforced concrete. It could completely overcome the problem the steel fiber concrete road surface of the "corrosion rust" and other phenomena of the road because of its inert material, less vulnerable to the effects of decay alkaline environment of concrete, and it does not absorb moisture. So, the synthetic fiber reinforced concrete developed very quickly. There are several different types of the synthetic materials as follows:

Glass Fiber. Glass fiber (GFRP) is curing molding at room temperature; its matrix is epoxy resin while the untwisted polymer sheet is a non-alkali glass fiber cloth. The mechanism is uses resin materials and glass fiber affixed to the surface of concrete structures to form composite materials, and it works to improve the components and structure by cooperation on the concrete (Darren et al. 2001), but the elastic modulus is relatively low, non-alkali-resistant, and long-term stress is susceptible to breakage. It is also easy to produce harmful particulates while using.

Nylon Fiber. Nylon fiber has a strong tensile, according to its material properties and processing techniques. It reaches only 200-300 Mpa normally so the enhancement of concrete is subjected to a great impact. Its density is relatively larger than the polypropylene fibers and sisal fibers, which are not conducive for the fiber to uniformly distribute the mixture. The limit of elongation is not as good as polypropylene fibers and sisal fibers.

Polypropylene Fiber. In recent years, the application of the preparation of concrete with polypropylene fibers is popular to improve the performance of concrete at home and abroad. Practice has shown that: the plastic shrinkage cracking of concrete can be inhibit and the concrete toughness can be reinforced, the concrete cracking and permeability can be improved by mix the polypropylene fiber into concrete. But its tensile strength is usually only 300–600 MPa, the same enhancement effect on the concrete will be greatly affected; its Young's modulus is relatively low (Heweiping, 2005).

Sisal Fiber.The sisal fiber (SF) is derived from the leaves of sisal plants, its chemical composition shown in Table 1, the fiber technology parameters shown in Table 2. In plant fiber, sisal fiber has the following advantages:

Table 1. Chemical composition of sisal fiber. unit:%

Fiber	Hemicellulose	Lignin	Pectin	Water soluble	Waxy lain	Moisture
65.8	12.0	9.9	0.8	1.2	0.3	10.0

Table 2. Performance comparison of fibers.

Types of fibers	Density (g/cm^3)	Tensile strength (MP)	Young modulus (GPa)	Elongation ultimate (%)
Steel fiber	7.80	400--1500	200--210	3.5--4.0
Glass fiber	2.60	400--1500	200--210	3.5--4.0
Nylon fiber	1.68	200--300	240--300	5.0--5.5
Polypropylene fiber	0.91	300--600	15--17	5.5--6.0
Sisal fiber	1.43	500--700	13--26	6.0

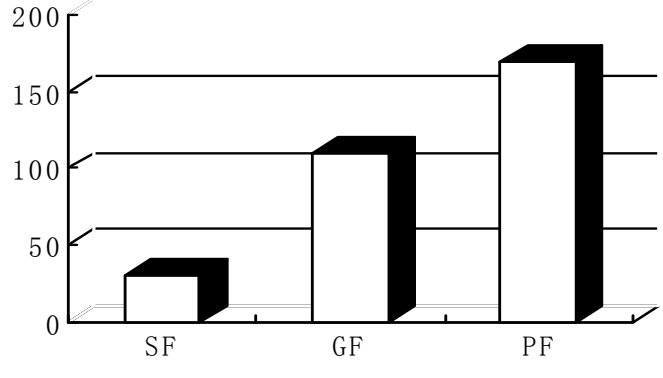


Figure 3. The relative price of different fibers.

- Sisal fiber requires a shorter growing period, the environment demand is not high, growing- process without pesticide and fertilizer, procedure of production without waste pollution, and the course of using without free harmful chemicals and particles.

- The sisal fiber consumes a small amount of energy in the period of growth, harvest, and processing. It has a high ability to absorb carbon dioxide that can alleviate the greenhouse effect.
- The sisal fiber absorbs energy well, impact resistance, and without brittle fracture (it is propitious for improving the security of vehicles).
- The sisal fiber is long, tough, and elastic. It is characterized by good tensile strength, anti-friction, acid-base resistance, marine corrosion resistance, microtherm, and so on.
- The sisal fiber is a renewable resource. It can be degraded naturally and doesn't hurt the

Environment. The cost is also lower for this type of fiber.(Chand, N 2008)(as shown in Figure 3).As shown in Table 2, both polypropylene fibers and sisal fibers have good properties. Figure 3 shows that the sisal fiber is more cost effective, too, compared to the polypropylene fiber. Thus it can be seen that, instead of polypropylene fiber, sisal fibers could be the late-model concrete additive.

Comparatively Experimental Analysis of Sisal Fiber's Impact on Anti-Sliding Performance of Asphalt Pavement

From the micro-structure of the sisal fiber, we can see that the sisal fiber can guarantee that the friction surface of material has a certain intensity. The filler materials which regulate the friction properties in the system are not easy to strip off from the matrix, so that not only reduces the wear and tear of materials, but also ensures material as a whole in the high-temperature friction conditions. It still has excellent friction properties.

Comparing steel fiber, glass fiber, nylon fiber, and polypropylene fiber reinforced friction materials, we can see that the sisal fiber reinforced friction material under temperatures. Its friction coefficient is moderate, and fluctuations are stable with temperature changes (see Figure 4).

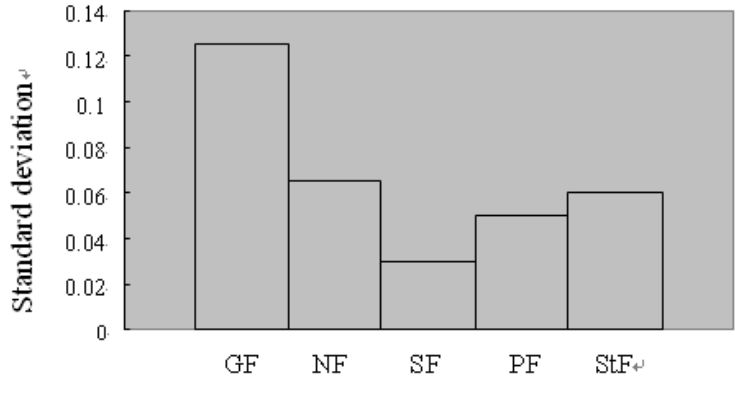


Figure 4. Standard deviation of coefficient of friction material which was reinforced by variety fibers.

Quantitatively add the sisal fiber to the asphalt mixture through the constant speed friction tester detected its friction performance in every period of time, come to surface friction coefficient attenuation law of sisal fiber-reinforced concrete, and analyze as follows by comparing with the surface friction coefficient attenuation law of ordinary concrete.

Analysis of the Attenuation Law on Friction Coefficient of Ordinary Concret Surface. The anti-slide performance data of different pavement materials all has undergone a process, which rapidly decreases from the initial value and gradually goes stable in the decay process. This common performance reacts to the inner regularity occurring in the usage process. The decay process can fit by appropriate mathematical models. The analytic method of relationship between pavement structure lateral force coefficient SFC of AC-13C and the standard axle laden time X is to firstly input lateral force coefficient data provided by Beijing-Zhuhai Expressway to SPSS(Highway research report 2006), then use SPSS to analyze and do curve fitting. Finally analyze the output result and curve comparison chart as follows.

We can see from the chart, the logarithmic model is distinctly superior to the exponential model. In this part of the output results: y (lateral force coefficient SFC) means the dependent variable; x (standard axle load times) means independent variables; R2 indicates the value of statistic quantity; d.f. indicates degree of freedom; F indicates test value; Sig indicates the actual significant level of test value F; also indicates the accompanied probability value p; Constant means the constant term; and b1 means the regression parameters. The specific model is:

$$y = -7.867 * \ln(x) + 91.001 \tag{1}$$

Table 3. Model summary and parameter estimation (AC-13C).

Model	Model summary					Parameter estimation	
	R Square	F	Df1	Df2	Sig.	Constant	b1
Logit	0.991	2370.302	1	22	0.000	91.001	-7.867
Index	0.926	277.053	1	22	0.000	4.036	-0.001

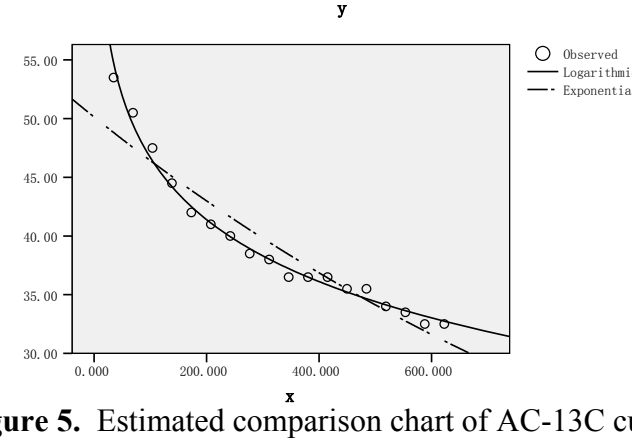


Figure 5. Estimated comparison chart of AC-13C curve.

The correlation coefficient is 0.991; lateral force attenuation coefficient is 7.867.

Analysis of The Attenuation Law on Friction Coefficient of Sisal Fiber Concrete. Laboratory testing can determine that the best dosage of sisal fiber in asphalt concrete is 0.2%, and the best length is 7 mm. Add the 0.2% sisal fiber to the AC-13 asphalt concrete, and the road performance has been significantly improved.
On September 6, 2007, road part K277 +200 ~ K277 +600 of the Beijing-Zhuhai Expressway paved 400 m long test section of sisal fiber-reinforced asphalt concrete (SFAC) and performed tracking and detection to its lateral force coefficient. On the test segment, set up four monitoring points, points 1 and 2 arranged on the wheel track of K277 +320 carriageway, points 3 and 4 arranged on the wheel track of K277 +460 carriageway.

Did statistics to the lateral force coefficient of the 4 point position used SPSS to do complete the analysis, do curve fitting. The output analysis result and curve comparison chart is shown in Table 4.

Average lateral force decay function of No. 1, 2, 3, and 4 monitoring points is as follows:

$$y = -5.03 * \ln(x) + 79.126 \tag{2}$$

Table 4. Model summary and parameter estimation (SFAC-13C).

model	Model summary					Parameter estimation	
	R^2	F	Df1	Df2	Sig.	Constant	b1
Logit	0.986	1753.137	1	24	0.000	79.126	-5.030

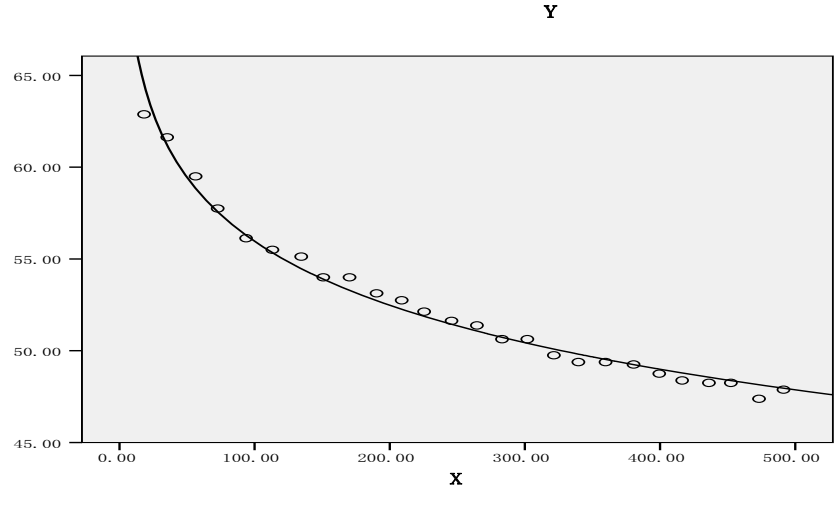


Figure 6. Curve fitting chart of the average lateral force coefficient of No. 1, 2, 3, and 4 monitoring points.

Its correlation coefficient is 0.986 and lateral force attenuation coefficient is 5.03. Through Eq. 1 and Eq. 2 we can see sisal fiber asphalt concrete (SFAC) compared with ordinary concrete. Its lateral force attenuation speed will be significantly slower. It can be seen that incorporation of sisal fibers effectively improves the anti-sliding performance of the pavement. In addition, make modification disposal to sisal fibers, sisal original fiber through the infiltration disposal of NaOH solution + heat + Borax - Formaldehyde-NaHSO3 solution can significantly improve the heat decomposition temperature of sisal original fibers , solve the key technical issues of the sisal fiber dramatically reducing the friction properties when temperature rise above 250℃(Megiatto et al 2007). The constant speed friction experiments show that the sisal fiber-reinforced friction material after the above disposal has excellent friction and anti-friction performance, which is expected to become a new generation of friction-braking fiber.

Conclusion

From the contrastively experimental analysis of the anti-sliding of pavement between the ordinary concrete and sisal concrete, and with the analysis of the performance and prices of variety fibers, it can be seen that sisal fibers have distinct advantages as the mixture additives.

With the modification of the sisal fiber, on the one hand, it could improve the content and crystallinity of the fiber, and then greatly enhance the mechanical property of fiber itself; on the other hand, to increase the surface area of fiber greatly means to increase the area of adhesion of fiber surface, and then enhance the boundary strength of sisal fiber, further enhance the mechanical property of composite material system. In the end, it is simple to come to the conclusion that the sisal fiber is the best choice as the mixture additive.

References

Chand, N., and Dwivedi, U.K. (2008). “Sliding wear and friction characteristics of sisal fiber reinforced polyester composites: effect of silane coupling agent and applied load,” *Journal of Polymer Composites*, 280-284.

Chen Ye. (2009). “Model analysis of anti-sliding property evaluation and mechanism of asphalt pavement,” *Journal of Shanxi Architecture,* 259-260.

Eddie, D , Shalaby, A , and Rizkalla, S. (2001). “Glass fiber-reinforced polymer dowels for concrete pavements,” *Journal of American Concrete Institute,* 201-206.

Elsaigh, W. A., Kearsley, E. P., and Robberts, J.M. (2010). “Steel fibre reinforced concrete for road pavement applications,” *24th Annual Southern African Transport Conference, SATC 2005:Transport Challenges for 2010*, 191-201.

He Weiping. (2005). "Polymer fiber in asphalt mixture," *Haichuan De Lan Nite AS of Reinforced Fiber Technical Literature Dedicated in Asphalt Road*.

Kulakowski, B.T. (1990). "Skid resistance manual report F 1WA-IP90-013," *U.S. Dept.of Transportation*.

Li Pengwei, X. (2009). "Anti-slip measures and the influence of anti-sliding performance of asphalt pavement," *Journal of Science and Technology promote Development,* 115-116.

Lu Zhao. (2005). "Application of soft fiber stiffeners in pavement," *Journal of Road Machinery and Construction Technology*, 18-19.

Megiatto Jr.,Jackson D., Oliveira,Franciéli B., Rosa,Derval S.; Gardrat,Christian, Castellan, Alain., Frollini,Elisabete. (2007). "Renewable resources as reinforcement of polymeric matrices:Composites based on phenolic thermosets and chemically modified," *Journal of. Macromolecular Bioscience*, 1121-1131.

Van Dam. T. (1995). "Use of steel-fiber reinforced concrete in pavement structures," *Proceedings of the Transportation Congress.Part 1 (of 2)*. 477-488.

Highway research institute of ministry of transport. (2006). "Western transportation construction projects," *General report on research project of pavement design and construction technology in loess region*.

Optimization of Steel Fiber Used for Induction Heating in Porous Asphalt Concrete

Quantao Liu[1], Erik Schlangen[1], Martin van de Ven[2] and Marco Poot[2]

[1] Micromechanics Laboratory (MICROLAB), Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands, E-mail: quantao.liu@tudelft.nl
[2] Road and Railway Engineering, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands

Abstract

An electrically conductive porous asphalt concrete used for induction heating and subsequently healing of cracks is prepared by adding conductive materials (steel fibers and steel wool) in this research. In this paper, the optimization of steel fiber used for induction heating is reported based on the electrical resistivity, induction heating speed and particle loss resistance of porous asphalt concrete. It is found that porous asphalt concrete containing steel fiber with smaller diameter or longer steel fiber is more electrically conductive and induction heatable than that containing the same content of steel fiber with bigger diameter or short steel fiber. It is also found that steel wool type 00 with length of 9.5 mm is more effective than short steel fiber type 1 and steel wool type 000 to improve the particle resistance of porous asphalt concrete. Finally, 8% (by volume of bitumen) of steel wool type 00 is considered as the best option used for induction heating in porous asphalt concrete.
Key Words: conductive porous asphalt concrete; electrical resistivity; induction heating; particle loss; water sensitivity

Introduction

Porous asphalt concrete with a minimum air void content of 20% is used extensively as a surface wearing course to improve the noise reduction and water drainage functions. In the Netherlands, at the moment about 90% of the highways are covered with porous asphalt concrete. One major disadvantage of porous asphalt is the poor performance in durability compared with dense graded asphalt pavement. The typical deterioration of porous asphalt pavement is ravelling. Ravelling is the loss of

aggregates from the surface of an asphalt pavement due to the friction of the tires of passing vehicles. In the Netherlands, ravelling is in about 76% of the cases the cause for maintenance or renewing of the top layer (Padmos, 2002). Due to ravelling, the average service life of porous asphalt surface course is only 10 years.

Asphalt concrete is known as a kind of self healing material (Little, 2007; Jo 2001). It is well known that the amount of healing can increase when the material is subjected to a higher temperature during the rest period (Little, 2007; Bonnaure 1982). However, it is difficult to stop traffic circulation on a road to allow enough self healing recovery at ambient temperature. A new conductive porous asphalt concrete used for induction heating and subsequently healing of cracks, has been developed so that the porous asphalt concrete can constantly repair itself and that the ravelling can be avoided or delayed (García, 2009; Liu 2009). The idea was to add steel fibers to porous asphalt concrete. When small cracks occur in the material, the temperature can be increased by induction heating via an external source to make the bitumen soft enough to start the self healing process to heal cracks and repair the bond between aggregates and binder. Steel fibers are added to porous asphalt concrete to enhance the conductivity of binder so that the micro-cracks can be healed by induction heating. Induction heating in the asphalt industry is a new technique. In this method, the power supply sends alternating current through the coil, generating an alternating electromagnetic field. When the conductive asphalt specimen is placed under the coil, this electromagnetic field induces currents flowing along the conductive loops formed by steel fibers. The induced current dissipates heat by Joule effect. It has been proved that the micro-cracks can be healed by the use of an external source to add heat via induction energy (Liu, 2010). Closing cracks on pavement via induction heating could be very useful because it can be repeated if damage returns. Another benefit of this method is that addition of fibers can improve particle loss resistance and water damage resistance of porous asphalt concrete in such a way that it may significantly improve the resistance to ravelling of porous asphalt concrete. The main purpose of this paper is to optimize steel fiber used for induction heating in porous asphalt concrete.

Experimental Methods

Materials. The stones used to make porous asphalt concrete specimens were quarry material (Bestone, Bremanger Quarry, Norway, size between 2.0 and 22.4 mm and density 2770 kg/m^3), crushed sand (size between 0.063 and 2 mm and density 2688 kg/m^3), and filler type Wigro 60 K (size<0.063 mm and density 2638 kg/m^3). The bitumen used was 70/100 pen, obtained from Kuwait Petroleum, with density 1032 kg/m^3, penetration 93 dmm and softening point 45.0 ºC. PA 0/16 was chosen as the mixture, for it was highly common used for surface wearing course in the Netherlands. The composition of the mixture based on the Dutch standard was shown in Table 1. Besides, three types of conductive steel fibers were mixed in the asphalt. The properties of the three steel fibers are shown in Table 1, where steel wool type 00 had three different lengths and steel wool type 000 was chopped by hand. Gyratory compactor was used to prepare specimens. The gyratory angle was 0.82º

and the vertical force was 600 kPa. About 60 gyrations were needed to compact all specimens to an air void content of 21%.

Electrical Resistivity Test. The electrical resistivity was used as a factor to select steel fibers for induction heating. The electrical resistivity measurements were done at room temperature around 20 ºC. The samples were cut from the gyratory compacted specimens; because flat surfaces were needed to assure good contact between sample and electrodes.The samples had a diameter of 100 mm and a thickness of 50 mm. Two electrodes made of rectangular copper plates in conjunction with wires connected with resistance tester were placed at both ends of the test sample to measure the electrical resistance. A small press was given to the copper plates to make good contact with the surface of the sample. The total contact resistance between the two electrodes was about 0.4 Ω, which was negligible with respect to the great resistances studied (higher than 100 KΩ in the samples). A digital multicenter was used to measure the resistance below 36 $\times 10^6$ Ω. A resistance tester was used to measure the resistance higher than this value. The electrical resistivity was obtained from the second Ohm-law, given in Eq. (1):

$$\rho = \frac{RS}{L} \tag{1}$$

where, ρ is the electrical resistivity, L is the internal electrode distance in meter, R is the measured resistance in omega and S is the electrode conductive area in square meter.

Table 1. Composition of PA 0/16 mixture based on the Dutch standard (RAW 2005).

Sieve size (mm)	Density (g/cm3)	RAW Spec. % retained	Cumm. % ret.	% ret. by weight
22.4–16.0	2.778	0-7	4	4
16.0–11.2	2.774	15-30	25	21
11.2–8.0	2.762	50-65	57	32
8.0–5.6	2.765	70-85	80	23
5.6–2.0	2.781	85	85	5
2.0–0.063	2.688	95.5	95.5	10.5
<0.063	2.638	100	100	4.5
Bitumen 70/100	1.032	4.5% by Wt. of aggregates		

Table 2. Properties of three types of steel fiber.

Type	Steel fiber 1	Steel wool 00	Steel wool 000
Diameters (μm)	29.6 - 191.1	8.89 - 12.7	6.38 - 8.89
Length (mm)	< 1	3.2/6.4/9.5	6
Resistivity (Ω cm)	$7 \cdot 10^{-7}$	$7 \cdot 10^{-7}$	$7 \cdot 10^{-7}$
Density (g/cm3)	7.8	7.8	7.8



Figure 1. Induction heating coil.

Induction Heating Test. Heating tests were done to compare the differences of the heating speeds of porous asphalt concrete containing different types of steel fiber.

As shown in Figure 1, the induction heating experiment was performed by using an induction heating system with a capacity of 50 kW and at a frequency of 70 kHz. Each sample was heated for 3 minutes and its temperature change was measured with a 640x480 pixel full color infrared camera. The distance between the coil of the induction machine and the top surface of the heated sample was about 32 mm. The principles of induction heating are electromagnetic induction and Joule heating (Liu 2009). An electrical force (current) will be induced in a conductive sample when the magnetic flux through it changes. The electrical current will generate heat when it flows through the conductive sample.

Cantabro Test. One of the disadvantages of porous asphalt is the loss of particles. Cantabro test was used to evaluate the particle loss resistance of porous asphalt concrete specimens containing conductive fibers. The test was done at 21.5 ºC in a Los Angeles abrasion machine without steel balls according to the European Norm EN 12697- 17. Each specimen was initially weighed (W_1) and placed separately into a Los Angeles drum. Thereafter, each specimen was weighed again after 300 revolutions of the drum (W_2) in order to determine the weight loss during testing. This weight loss is an indication of the cohesive properties of the mix. Lower weight loss means better cohesion and better particle loss resistance. The test results were expressed as a percentage of weight loss in relation to the initial weight as given in Eq. (2):

$$PL = \frac{W_1 - W_2}{W_1} \times 100 \tag{2}$$

where, PL is the particle loss in percent; W_1 is the initial specimen mass in gram and W_2 is the final Specimen mass in gram.

Water Sensitivity Test. Moisture can result in loss of adhesion between the asphalt binder and the aggregate surface, and can accelerate the development of other distresses such as pothole, cracking and ravelling. So, water sensitivity is an indication of the adhesion between bitumen film and aggregate. The test was done to study the effect of steel wool on the adhesion property of porous asphalt specimens.

The specimens were divided into a dry and a wet group. The dry group was directly used to test the indirect tensile strength at 15 ºC according to the European Norm EN 12697- 12. The wet group was soaked in a 40 ºC water bath for 68 hours and then used to test the indirect tensile strength. The indirect tensile strength ratio was used to denote the water sensitivity of specimens. ITSR was calculated with Eq. (3):

$$ITSR = \frac{ITS_w}{ITS_d} \times 100 \tag{3}$$

where $ITSR$ is the indirect tensile strength ratio in percent, ITS_w is the average indirect tensile strength of three wet samples and ITS_d is the average indirect tensile strength of three dry samples.

Results and Discussions

Effect of the Volume Content of Three Types of Steel Fiber on the Electrical Resistivity of Porous Asphalt Concrete. Fiber size determines the number of fibers per unit of batched weight and the number of fibers per cubic meter of asphalt concrete. Since the total weight determines the cost of the fibers, the question arises whether a large number of short fibers are more effective than the same weight of long fibers. To answer this, three different types of steel fiber are used to check which one is better to improve the electrical conductivity of porous asphalt concrete.

The comparisons among the influence of the three types of steel fiber volume content (fiber-bitumen volume ratio) on the electrical resistivity of porous asphalt concrete can be seen in Figure 2. All three resistivity curves show three stages: a high resistivity stage, exhibiting insulating behavior with resistivity higher than 10^9 Ω m; a transit stage, with resistivity suffering a sharp decrease from 10^9 Ω m to 10^4 Ω m; and a low resistivity stage, exhibiting conductive behavior with resistivity of 10^4 Ω m. However, there are some differences among the three resistivity curves. The diameter of steel fiber type 1 is biggest and the diameter of steel wool type 000 is smallest. The transit stage of the resistivity curve comes earlier and is sharper for the type of fiber with smaller diameter. The optimal content of fiber to make porous asphalt concrete conductive is 20%, 12% and 10% for steel fiber type 1, steel wool type 00 and steel wool type 000 respectively, which means that the smaller the diameter, the less fiber is needed to make porous asphalt concrete conductive. Among the three different types of steel fiber, steel wool type 000 is the best one to make porous asphalt concrete conductive.

Effect of Three Types of Steel Fiber Volume Content on The Induction Heating Speed of Porous Asphalt Concrete. Induction heating speed (increase of temperature in a certain time) of porous asphalt concrete containing different types of steel fiber can serve as another factor to select steel fiber used for induction heating. During the heating, the temperature variations on the top sides of the samples were recorded and used to denote the induction heating speed.

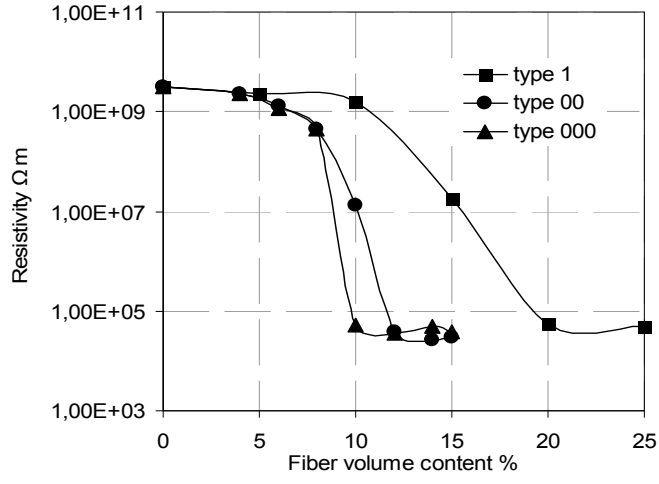


Figure 2. Effect of three types steel fiber volume content on the resistivity of porous asphalt concrete.

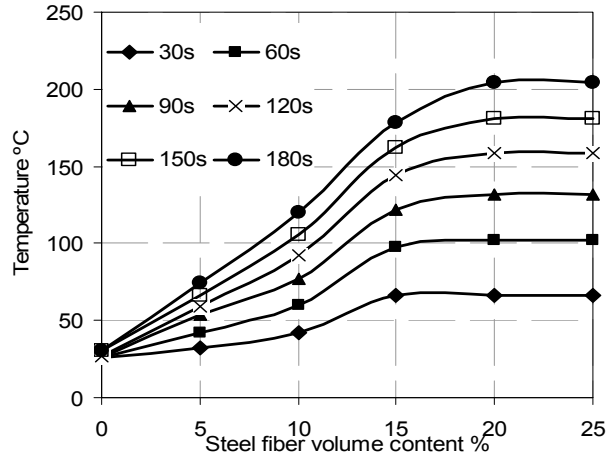


Figure 3. Effect of steel fiber type 1 volume content on the induction heating speed of porous asphalt concrete.

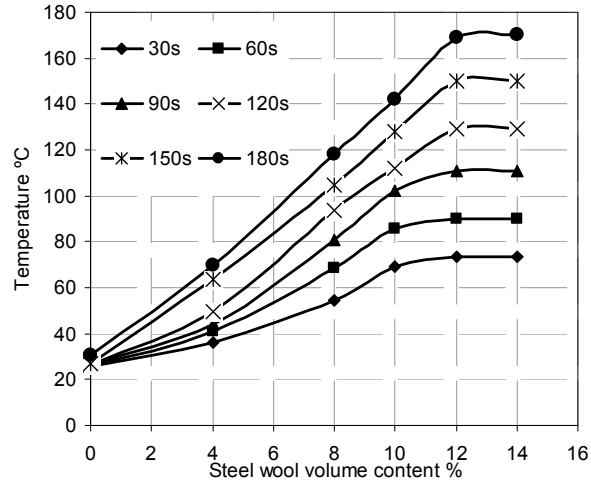


Figure 4. Effect of steel wool type 00 volume content on the induction heating speed of porous asphalt concrete.

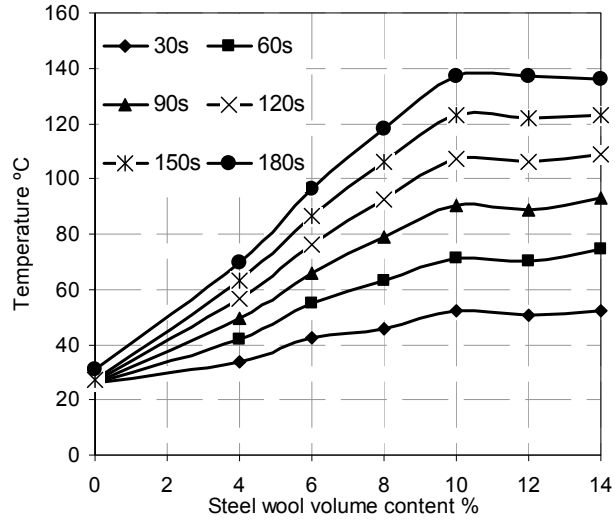


Figure 5. Effect of steel wool type 000 volume content on the induction heating speed of porous asphalt concrete.

As is shown in Figures 3 to 5, the plain samples can't be heated with induction heating. The samples containing steel fiber can be heated with induction heating and the temperature reached within 3 minutes heating increases with the increase of the volume content of steel fiber. However, the temperature of samples can not be increased any faster when the steel fiber volume content reaches a certain value, which can be seen as the optimal content for the highest induction heating speed. The optimal contents are 20%, 12% and 10% for steel fiber type 1, steel wool type 00 and steel wool type 000 respectively. These values coincide with the optimal contents of steel fiber for the minimums of resistivity in Figure 2, which means that minimum resistivity corresponds to maximum induction heating speed. The maximum temperatures reached of samples containing the optimal contents of steel fiber are 207 ºC, 169 ºC and 137 ºC respectively for three different types of fiber. After 3 minutes heating, the temperatures of samples containing 10% of different types of steel fiber reached 120 ºC, 142 ºC and 137 ºC for steel fiber type 1, steel wool type 00 and steel wool type 000 respectively. So, steel wool type 00 is the best one to make porous asphalt concrete induction heatable.

Effect of the Length of Steel Wool on the Electrical Conductivity and Induction Heating Speed of Porous Asphalt Concrete. The differences in Figures 2 to 5 are mainly attributed to the differences in diameters of steel fiber. Figures 6 and 7 show the effect of the length of steel fiber (steel wool type 00) on the electrical resistivity and induction heating speed of porous asphalt concrete.

No apparent differences exist among the three resistivity curves when small amount of steel fiber is added to the porous asphalt concrete. The differences between the three resistivity curves show up when the steel fiber volume content reaches 8% (threshold point), above which the resistivity curves of porous asphalt concrete

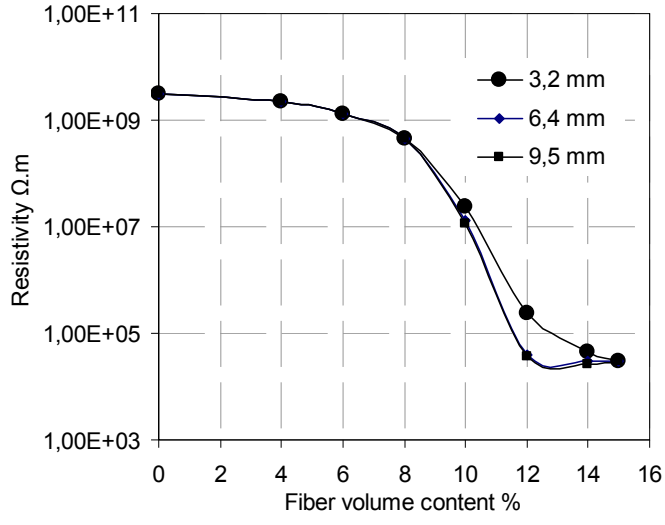


Figure 6. Effect of the length of steel wool type 00 on the electrical resistivity of porous asphalt concrete.

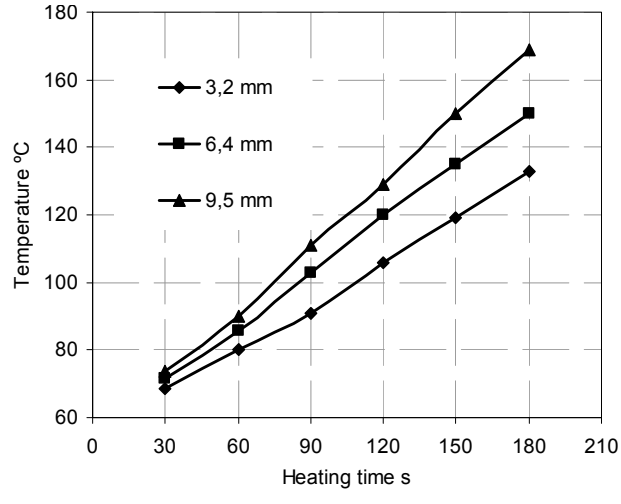


Figure 7. Effect of the length of steel wool on the induction heating speed of porous asphalt concrete containing 12% of steel wool type 00.

decrease sharply. The differences become bigger with the increase of steel wool volume content. The two resistivity curves for steel wool with length of 6.4 mm and 9.5 mm decrease more sharply than that for steel wool with length of 3.2 mm. For steel wool with a length of 6.4 mm and 9.5 mm, porous asphalt concrete becomes electrically conductive when the steel wool volume content reaches 12%. For steel wool with a length of 3.2 mm, 15% of steel wool is needed to make porous asphalt concrete electrically conductive. It seems that longer steel wool is more effective to make porous asphalt concrete electrically conductive and 9.5 mm is long enough for electrically conductive purpose. However, more differences among the induction heating speeds of porous asphalt concrete containing 12% steel wool type 00 with different lengths are found in Figure 7. It is shown that the induction heating speed of porous asphalt concrete with longer steel wool is higher than that of porous asphalt

concrete with shorter steel wool. Steel wool type 00 with the length of 9.5 mm is the best to make porous asphalt concrete electrically conductive and induction heatable.

Effect of Steel Fiber Volume Content on The Particle Loss Resistance of Porous Asphalt Concrete. A particle loss test was done to check if the addition of steel fibers for induction heating purpose worsens the particle loss resistance of porous asphalt concrete or not. The effect of the steel fiber volume content on the particle loss resistance of porous asphalt concrete is shown in Figure 8.

The particle loss of plain porous asphalt concrete samples is 14.84%. It can be seen that the particle loss of porous asphalt concrete decreases with the increase of the fiber volume content and reaches a minimum, after which adding more steel fibers will result an increase of particle loss. Moderate fibers reinforce the mastic, but too many fibers would reduce the bitumen film thickness, causing a bad cohesion between the mastic components. The minimums of particle loss are 8.77%, 8.27% and 8.01% for three different types of fibers and about 15%, 8%, 8% of steel fiber is needed to make the porous asphalt concrete have the optimal particle loss resistance for steel fiber type 1, steel wool type 00 with length of 9.5 mm and steel wool type 000 respectively. At the contents for optimal conductivity (20% for steel fiber type 1, 12% for steel wool type 00 and 10% steel wool type 000), the particle losses are 10.31%, 15.83% and 13.66% respectively. Adding 12% of steel wool type 00 to porous asphalt concrete for optimal conductivity purpose will worsen its particle loss resistance and adding 10% of steel wool type 000 or 20% steel fiber type 1 to porous asphalt concrete for optimal conductivity purpose will increase its particle loss resistance. Among the three different steel fibers, steel wool type 00 with a length of 9.5 mm is best to increase the particle loss resistance of porous asphalt concrete. It is

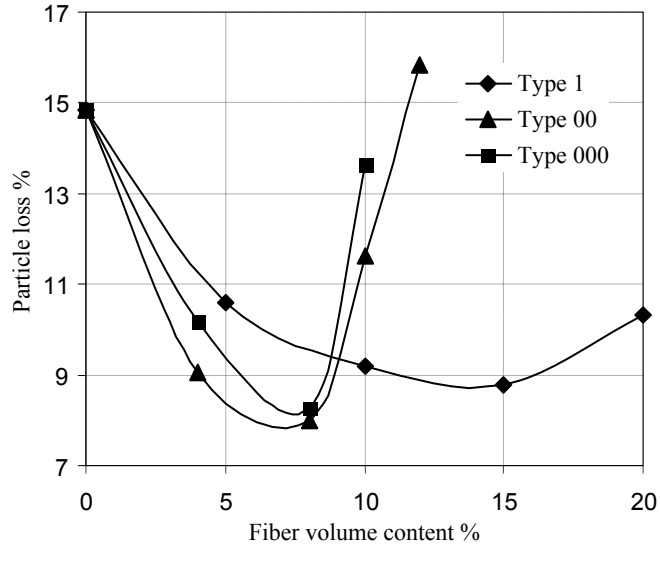


Figure 8. Effect of the steel fibre volume content on the particle loss resistance of porous asphalt concrete.

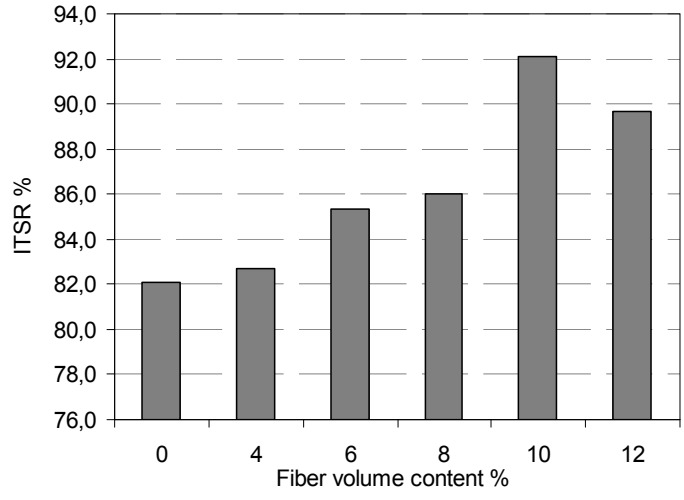


Figure 9. Effect of steel wool type 00 volume content on the water sensitivity of porous asphalt concrete.

considered that ravelling is mostly attributed to the cohesion failure in the mortar (Hagos, 2007) and it is expected that porous asphalt concrete containing steel fiber used for induction heating has good particle loss resistance (higher cohesion strength) intrinsically, which can delay ravelling. So, 8% (by volume of bitumen) of steel wool type 00 is an optimal option to obtain best particle loss resistance in porous asphalt concrete.

Water Sensitivity of Porous Asphalt Concrete Containing Steel Wool Type 00. It has been proved that steel wool type 00 (with a length of 9.5 mm) is the best among the three different steel fibers to make porous asphalt concrete show best particle loss resistance and excellent induction heating. Water sensitivity of porous asphalt concrete containing this fiber is investigated in the section.

It is found in Figure 9 that addition of steel wool to porous asphalt concrete can increase its ITSR. The ITSR increases with the increase of the steel wool volume content and reaches a maximum of 92% with 10% of steel wool type 00, after which adding more steel wool to porous asphalt concrete will decrease its ITSR. The ITSR of plain samples is 82%. For the samples containing 8% steel wool type 00, the ITSR is 86%, slightly higher than that of plain samples, which means addition of 8% steel wool type 00 to porous asphalt concrete for induction heating purpose can also increase its water damage resistance(adhesion between aggregate and mortar). It is also found that in heating test the samples containing 8% steel wool type 00 are more heatable than those containing the same volume other two kinds of steel fiber.

Conclusions

Porous asphalt concrete containing steel fiber with smaller diameter is more electrically conductive than that containing the same content of steel fiber with bigger diameter. Porous asphalt concrete containing longer steel wool is more

electrically conductive and induction heatable than that containing the same content of shorter steel wool. Steel wool type 00 with length of 9.5 mm is more effective to improve the particle loss resistance of porous asphalt concrete than the other two types of steel fiber. Adding 8% of steel wool type 00 (by volume of bitumen) to porous asphalt concrete will result in best particle loss resistance and better water damage resistance. Meanwhile, 8% of this fiber is more effective to make porous asphalt concrete induction heatable than the same weight other two types of steel fiber. In conclusion, 8% of steel wool type 00 with length of 9.5 mm is considered as the best option used for induction heating in porous asphalt concrete. However, more research is needed to find out the possibility for large scale application in the filed.

References

Álvaro García, Erik Schlangen, Martin van de Ven and Quantao Liu (2009). "Electrical conductivity of asphalt mortar containing conductive fibers and fillers." *Construction and Building Materials*,10(23),3175-3181.

Little, D. N., Bhasin, A. (2007). "Exploring mechanisms of healing in asphalt mixtures and quantifying its impact." *Self Healing Materials an Alternative Approach to 20 Centuries of Materials Science, Springer Series in Materials Science*, 100, 205-218.

EN 12697-12 (2004). Bituminous mixtures - Test methods for hot mix asphalt – Part 12: Determination of the water sensitivity of bituminous specimens.

EN 12697-17 (2004). Bituminous mixtures- Test methods for hot mix asphalt- Part 17: Particle Loss of porous asphalt specimen.

Hagos, E.T. (2007). "The effect of aging on Binder Properties of Porous Asphalt Concrete" *Ph.D thesis*, *Delft University of Technology,* The Netherlands.

Bonnaure, F., Huibers, A. H., and Boonders, A. (1982). "A laboratory investigation of the influence of rest periods on the fatigue characteristics of bituminous mixes." *Journal of the Association of Asphalt Paving Technologists*, 51, 104-128.

Padmos C. (2002). "Over ten years experience with porous road surfaces." *ISAP 9th International conference on Asphalt Pavement*s, [4-3-4].

Liu, Q. T., Schlangen, E., García, Á., and van de Ven M. (2009). "Induction heating of electrically conductive porous asphalt concrete." *Construction and Building Materials*, doi:10.1016/j.conbuildmat.2009.12.019.

Liu, Q. T., Schlangen, E., van de Ven M., and García, Á. (2010). "Healing of conductive porous asphalt concrete via induction heating." *EATA08 special issue of the international journal Road Materials and Pavement Design*, Accepted.

Jo, S., and Richard, K. (2001). "Laboratory evaluation of fatigue damage and healing of asphalt mixtures." *Journal of Materials in Civil Engineering*, 13(6), 434-440.

Effect of Alternative Materials on Engineering Properties of Alluvial Soil for Construction of Road Subgrade

Tapash Kumar Roy[1], B.C. Chattopadhyay [2] and S.K. Roy [3]

[1] Sr. Lecturer, [2] Prof., [3] Prof., Department of Civil Engineering, Bengal Engineering & Science University, Shibpur. Howrah-711103, West Bengal. India;
Email: tapash@civil.becs.ac.in, ccbikash@ yahoo.co.in, sudip@ civil.becs.ac.in

Abstract

With the increasing constructional activities in road sector, situation has arisen where conventional good quality soils for subgrade are in short supply in many locations due to variety of reasons. Particularly in densely populated areas this problem is very accute. On the other hand, due to rapid industrialization throughout the world, the production of huge quantity of waste materials creates not only the environmental problem but also the depositional hazards. Safe disposal of the same is very vital issue and such situation can be addressed by the bulk utilization of these materials mainly in the field of civil engineering applications. A systematic detailed investigation was undertaken by the authors for using rice husk ash (RHA) and pond ash with soils to enhance the quality of the soil as subgrade material. CBR value in soaked and unsoaked condition increases in a small amount for addition of RHA, but with addition of lime in RHA mixed soils, better increment has been observed in the CBR values both in the soaked and unsoaked conditions. Further mixing pond ash with the soil showed increment in CBR both in unsoaked and soaked condition and mixing of pond ash upto 150% has shown that the gain in CBR value is more than double for soaked and unsoaked conditions.

Introduction

Due to rapid development in the industrial sectors, huge quantities of waste materials are being produced as by-product and such wastes are to be located over dumping zones near the plants. These wastes not only need depositional land, but also create accumulation hazard and pollute environment around the plants. Demand of electricity is increasing gradually everywhere and for fulfilling this demand, a large number of coal based thermal plants are being set up. However these coal-based

thermal power plants all over the world face serious problems of handling and disposal of ash. It is estimated that thermal power plants in India produce about 45 million tons of ash a year. With decreasing coal quality and increasing power generation, coal ash production is projected to reach 90-100 million tons a year by the turn of the century. Only 2% of the ash produced is used for some gainful purposes; the rest is to be disposed as waste material (*www.worldbank.org*). So as the huge quantities of fly ash are being produced in the country resulting necessity of utilizing fly ash gainfully (Rao 1999), different areas of application of fly ash in Civil Engineering Construction are being made with different degrees of success. Similarly rice is one of the most cultivated and consumed cereal in the world. The husk surrounding the kernel of rice is known as rice-husk accounts for approximately 20% by weight of the harvested grain (paddy) (Mahin, 1990). Annual production of rice husk in India is about 22.31 million tons and study indicates that one ton of paddy produces about 250 kg of rice husk (Chandra *et. al.,* 2005). However a majority of mills use the rice husk as fuel to generate steam to parboil rice with very low combustion efficiency; and as a by-product rice husk ash is produced. Rice husk ash is being generated in yearly frequency in large quantity, about 70 million tones of RHA is produced annually worldwide (*www.ricehuskash.com*) and is commonly left as waste materials. So safe disposal of these types of waste materials adversely affecting the environment and provision for required large storage area are major concerns now-a-days. On the other hand, conventional materials usually used for construction of subgrade of pavement are in short supply in many locations due to various reasons like non-availability of adequate lands to supply fill materials, uneconomical lead, land acquisition problems etc. In view of the shortage of commonly used cohesive soils as fill materials for road construction, need has arisen to use suitable alternative materials for same purpose.

Literature Review

Over the last decades, many works are reported indicating the potential of using different waste materials in different forms in many civil engineering applications. Some works are also reported in here by showing their possible use in different ways. Senol *et al.* (2002) evaluated the effects of fly ash used for stabilizing the soft subgrade and experimental results observed that engineering properties, such as unconfined compressive strength and resilient modulus increase substantially after addition of fly ash. Utilization of fly ash has been long recognized in the construction of embankment as fill materials in reclamation works or in road construction for subbase course (Reddy *et al.* 1999) or base courses (Tike *et al.* 1999). An experimental study has been made by Roy and Chattopadhyay (2005) to use pond ash by modifying their behavior particularly compaction characteristics and CBR in combination with locally available cohesive soil. Results of tests for California Bearing Ratio of pond ash mixed with ordinary silty clay, a locally available soil for various proportion of mixing. Arora et al (2005) carried out a series of tests on locally available sandy soil and flyash mixtures prepared with cement and lime activators. Roy and Chattopadhyay (2008) conducted an experimental program for exploring the possibility of utilization of the alternative materials like rice husk ash

and pond ash by mixing with soil for the construction of road subgrade as cost effective mix. An experimental investigation was carried out by Heeralal et al. (2006) to study the effect of lime & fly ash on poorly graded sand. An experimental work done by Garchar and Trivedi (2005) to study the stabilizing effects of fly ash on an expansive soil and result indicated that addition of fly ash reduced the swelling characteristics and improve the CBR value of expansive soil. Bhasin et al. (1985) reported the improvement in the soaked CBR of soil on addition of rice husk ash (RHA) & lime sludge and further extended their investigation on stabilization of black cotton soil by using RHA with or without lime. Ravi Shankar et al. (2008) made a study to modify the characteristics of lateritic soil with addition of pond ash and cement. A small quantity of cement when added with pond ash and rice husk mixed alluvial soil showed a significant change on compaction characteristics and CBR values and the improvement occurred in the soaked CBR value is nearly 200% for mixed material of rice husk ash, pond ash, soil and cement in the proportion of 20%, 60% &20% respectively with the addition of cement as 1% than that of the soaked CBR value of mixed soil (RHA: pond ash: soil as 20:60:20) without cement (Roy and Chattopadyyay, 2008). Roy and Chattopadyyay (2008) also performed the experimental study on the sandy soil mixing pond ash and rice husk ash with addition of cement and better performance has been observed on the CBR value for the mixed soil comprises of Rice husk ash, pond ash and Silver sand of ratio 20:60:20 with addition of cement 3%. Hence utilization of waste materials e.g. pond ash, rice husk ash etc. in bulk quantity in the construction of road project will reduce the environmental pollution as well as hazard of accumulation. So keeping all the above in view, a detailed experimental investigation was undertaken by the authors for using different waste materials like rice husk ash, pond ash etc. with soils to enhance the quality of the soils as subgrade material. Different results of experimental study on behaviors of different waste materials mixed with soils along with some chemical as reported in literature, was also used to seek confirmation of results obtained by the authors.

Materials Used

For knowing the efficacy of using pond ash and rice husk ash as admixture in soils while using the mixture as subgrade materials for roads, some of the experimental results of different properties published in different literature have been furnished below. Jha and Gill (2006); Roy & Chattopadhyay (2005); Roy et.al.(2008); Reddy & Singh (2008) have performed some investigation to study the effect of fly ash, pond ash, rich husk ash (RHA) on the subgrade soil with or without lime or cement. The physical properties of the materials used by the said investigators are given in the table 1.

For knowing the effect of mixing RHA with soil, Jha and Gill (2006) mixed RHA with soil in the proportion of 6%, 12% and 18% and further for achieving the better improvement, different increasing proportions of lime added with the RHA mixed soil in the ratio of 3%, 5% and 7%.

Table 1. The physical properties of the Soil, Pond ash and Rice husk ash used after Jha and Gill (2006); Roy & Chattopadhyay (2005); Roy et.al. (2008); Reddy & Singh (2008).

Sl. No	Properties	After Jha & Gill (2006)		After Roy & Chattopadhyay (2005); Roy et.al. (2008)			After Reddy & Singh (2008)	
		Soil	RHA	Soil	RHA	Pond ash	Soil	Fly ash
1	Atterberg's limit-							
	Liquid limit (%)	61.2	Non-	47.0	Non-Plastic	Non-Plastic	47.0	
	Plastic limit (%)	34.5	Plasti	32.0			32.0	
	Plasticity Index (%)	26.7	c	15.0			15.0	
2	Grain size Distribution							
	Sand size particles (%)	12.0	-	Nil	72.0	31.25	9.80	
		58.0		68.0	28.0	63.50	76.0	F type
	Silt size particles (%)	30.0		32.0	Nil	5.25	0	fly
	Clay size particles (%)						14.2	ash
							0	
3	Light compaction-	16.2		15.5			17.3	
	Proctor's Maximum Dry Density (kN/m3)	0	8.50	0	11.30	11.4	0	
	Optimum Moisture	24.4	31.80	23.0	28.00	32.5	18.8	
	Content (%)	0		0			0	
4	Specific gravity	2.69	1.95	2.71 3	2.06	2.193	2.54	2.08
5.	CBR-							
	Unsoaked (%)	3.65	8.75	12.9 1	9.90	30.00	-	-
	Soaked (%)	2.40	8.15	4.90	5.39	24.00	3.21	-

Preparation of Mixed Samples

However Roy et.al. (2009) mixed RHA with soil in the ratio of 5%, 10%, 15% and 20%. Reddy & Singh (2008) studied the effectiveness of soil blended with fly ashes and cement in various proportions. Roy & Chattopadhyay (2005) performed an experimental program to map the modification in compaction and strength properties of the typical South Bengal fill materials of locally available yellowish clayey silt on mixing into different proportions of pond ash collected from Kolaghat Thermal Power Plant of West Bengal in India and a design mix proportion was determined by using pond ash in increasing proportion from 0% to 400% and decreasing proportion of soil from 100% to 0%. All the proportions have been showed in the Table 2.

Discussions on Experimental Results

Improvement in the characteristics of the soils due to mixing of reported different alternative materials particularly for compaction characteristics and CBR values at OMC are discussed below.

Table 2. Details of Mixing Proportion of Fly ash, Pond ash and Rice husk ash with soil.

Mix. No	As per Jha & Gill (2006)		Roy et.al. (2009)	Reddy & Singh (2008)		Roy & Chattopadhyay. (2005)
	RHA (%)	Lime (%)	RHA (%)	Fly ash (%)	Cement (%)	Pond ash (%)
1	0	0	0	0	0	0
2	6	0	5	20	0	25
3	12	0	10	35	0	67
4	18	0	15	50	0	150
5	6	3	20	35	3	400
6	12	3		35	5	
7	18	3		35	7	
8	6	5				
9	12	5				
10	18	5				
11	6	7				
12	12	7				
13	18	7				

(Ater Jha and Gill (2006); Roy & Chattopadhyay (2005); Roy et.al.(2008); Reddy & Singh (2008).

Compaction Characteristics

The effect on compaction characteristics properties of soils for mixing RHA, fly ash, pond ash, and lime in varying percentage are shown in the following Table 3 and 4. From the experimental results of Jha & Gill (2006), it is observed that MDD and OMC of the soil decreases gradually with the increasing amount of RHA. However with addition of RHA in increasing proportion as 6%, 12% and 18%, MDD decreases and OMC increases gradually for added percentage lime in increasing amount. For addition of lime 3% with the RHA mixed soil, the MDD and OMC become 15.84 kN/m^3, 15.25 kN/m^3, 14.41 kN/m^3 and 18.75%, 20.83%, 23.67% respectively for increasing amount of RHA as 6%, 12% and 18%. Similar observations are made for mixing of cement 5% and 7% with RHA mixed soils. But Roy et al. (2009) showed that addition of RHA with soil in increasing proportion MDD decreases and OMC increases gradually. They also added that OMC increases only 1.5% more than the corresponding value of soil for achieving MDD which slightly lesser than the said value of soil.

Investigational results of Reddy & Singh (2008) indicated that mixing of pond ash in increasing ratios with the soil, MDD decreases and OMC increases gradually and for fly ash added with the soil upto 50%, MDD and OMC become 15.64 kN/m^3 and 19.63% respectively but the corresponding value of soil 17.30 kN/m^3 and 18.80%. Similar trends have been showed for addition of cement as 3%, 5% and 7% with the

fly ash mixed soil. In the Table 4, observation has been made for mixing of fly ash as 35% with the soil and MDD and OMC represent as 15.89 kN/m^3, 15.79 kN/m^3, 15.68kN/m^3 and 19.90%, 20.42%, 21.00% respectively for added amount of cement 3%, 5% and 7%. Similar observations have been made by Roy & Chattopadhyay (2005) for the results of MDD and OMC for mixing pond ash with the soil.

Table 3. Compaction characteristics of RHA mixed soil.

After Jha & Gill (2006)					After Roy et al. (2009)			
Mix. No.	RHA (%)	Lime (%)	MDD (kN/m^3)	OMC (%)	Mix. No.	RHA (%)	MDD (kN/m^3)	OMC (%)
1	0	0	16.20	24.40	1	0	15.50	23.00
2	6	0	16.00	23.50	2	5	15.30	23.65
3	12	0	15.50	20.50	3	10	15.00	23.90
4	18	0	14.75	18.00	4	15	14.90	24.14
5	6	3	15.84	18.75	5	20	14.70	24.50
6	12	3	15.25	20.83				
7	18	3	14.41	23.67				
8	6	5	15.78	18.76				
9	12	5	14.81	20.84				
10	18	5	14.90	23.42				
11	6	7	15.59	19.92				
12	12	7	14.56	21.00				
13	18	7	14.00	24.58				

(After Jha & Gill (2006); Roy et.al. (2009).

Table 4. Compaction characteristics of fly/pond ash mixed soil.

After Reddy & Singh (2008)					After Roy & Chattopadhyay. (2005)			
Mix. No.	Fly ash (%)	Cement (%)	MDD (kN/m^3)	OMC (%)	Mix. No.	Pond ash (%)	MDD (kN/m^3)	OMC (%)
1	0	0	17.30	18.80	1	0	15.50	23.0
2	20	0	16.57	19.29	2	25	15.04	23.5
3	35	0	16.14	19.50	3	67	14.60	23.5
4	50	0	15.64	19.63	4	150	13.40	24.0
5	35	3	15.89	19.90	5	400	12.50	28.0
6	35	5	15.79	20.42				
7	35	7	15.68	21.00				

After Reddy & Singh (2008); Roy & Chattopadhyay. (2005).

Table 5. CBR of RHA mixed soil.

After Jha & Gill (2006)					After Roy et.al. (2009)			
Mix.	**RHA**	**Lime**	**Unsoake**	**Soaked**	**Mix**	**RHA**	**Unsoake**	**Soaked**
1	0	0	3.65	2.40	1	0	12.91	4.90
2	0	3	13.00	16.75	2	5	13.10	5.13
3	6	3	18.75	18.00	3	10	13.45	5.54
4	12	3	20.50	19.50	4	15	14.30	5.97
5	18	3	20.25	18.75	5	20	14.92	6.24
6	0	5	26.68	33.42				
7	6	5	31.32	33.84				
8	12	5	36.37	34.68				
9	18	5	35.11	33.42				
10	0	7	19.90	24.40				
11	6	7	24.80	27.60				
12	12	7	29.20	28.00				
13	18	7	27.80	25.80				

After Jha & Gill (2006); Roy et.al. (2009).

Table 6. CBR of fly ash and pond ash mixed soil.

After Reddy & Singh (2008)				After Roy & Chattopadhyay (2005)			
Mix.	Fly ash	Cement	Soaked	Mix.	Pond	Unsoaked	Soaked
1	0	0	3.21	1	0	13.0	5.0
2	20	0	2.00	2	25	17.0	5.5
3	35	0	1.72	3	67	19.0	9.5
4	50	0	1.50	4	150	29.0	12.0
5	35	3	15.08	5	400	29.5	15.0
6	35	5	22.87				
7	35	7	41.36				

After Reddy & Singh (2008); Roy & Chattopadhyay (2005).

California Bearing Ratio (CBR) at OMC

The effect on CBR value of the mixed soil shown in the Table 2 in soaked and unsoaked condition for addition of chemicals in varying percentage have been shown in the following Table 5 and 6.

From the reported values of CBR of Jha & Gill (2006), it is observed that unsoaked and soaked CBR of only soil are 3.65% and 2.40% respectively. However addition of lime of with the soil, CBR increased significantly in soaked condition compared to those of unsoaked condition and this value becomes 13%, 26.68% and 19.90% in unsoaked condition and 16.75%, 33.42% and 24.40% in soaked condition for mixing of lime 3%, 5% and 7% respectively. But addition of RHA of increasing ratio with the lime mixed soil, the CBR increased remarkably in both unsoaked and soaked condition and this value showed the higher range in unsoaked condition. compared to that in soaked condition except 6% RHA mixed soil added with 3% and

5% lime. Roy et.al. (2009) showed that for chosen soil without any admixture, the CBR is 12.91% and 4.9% in unsoaked and soaked condition respectively.

But addition of RHA with the soil in greater proportion, a monotonic increase in the CBR value in both soaked and unsoaked condition has been observed. In this case CBR in unsoaked condition have shown the higher amount compared to that in the soaked condition.

Reddy & Singh (2008) have shown that the CBR for mixing fly ash with the soil decreased gradually in soaked condition, but addition of cement with the fly ash mixed soil indicated the significant increment in the soaked CBR in all the cases due to pozzalanic action among the soil, fly ash and cement.

Roy & Chattopadhyay (2005) have shown the CBR for mixing pond ash with the soil in both unsoaked and soaked condition in the table 6. The said value increased gradually for mixing pond ash with the soil in increasing amount. But remarkable increment has been observed for addition of 150% pond ash with soil particularly in unsoaked condition and this value becomes 29.0%. Further addition of pond ash, CBR value does not increase significantly particularly in unsoaked condition. So it ideal to mix pond ash with locally available soil by 150% (by weight) at which it is not needed any increase in water demand for compaction at OMC in the field compared to that required for virgin soil, but gain in CBR value is more than double for soaked and unsoaked conditions.

Conclusions

From the results of investigation reported above, the following conclusions can be made:
(i) Consumption of waste materials like fly ash, pond ash or RHA in bulk quantity in the construction of road project can be made with reducing the accumulation hazard and environmental pollution from these wastes.

(ii) Addition of reported waste materials in increasing proportion with the conventional soil decreases the maximum dry density of the mixed soil with or without mixing of lime or cement. However the optimum moisture content of the mixed soil increases gradually with the increased percentage of waste materials as well as chemicals such as lime or cement.

(iii) In strength characteristics, CBR value in soaked and unsoaked condition increases in a small amount for addition of RHA, but with addition of lime in RHA mixed soils, better increment has been observed in the CBR values both in the soaked and unsoaked conditions. However addition of fly ash with the soil in increasing amount have not shown any improvement in the result of CBR but mixing of cement with the fly ash mixed soil indicated the significant increment in the soaked CBR in all the cases due to pozzalanic action among the soil, fly ash and cement. On the other hand, mixing pond ash with the soil showed increment in CBR

both in unsoaked and soaked condition and mixing of pond ash upto 150% has shown that the gain in CBR value is more than double for soaked and unsoaked conditions.

References

Arora, S., and Aydilek, A.H., (2005). "Class F Fly-Ash-Amended Soils as Highway Base Materials." *Journal of Materials in Civil Engineering,* 17,(6), ASCE, 640-649.

Bhasin, N.K, Goswami, N.K., Oli, P., Krishan, N., Lal, N.B. (1985). "Utilization of Waste Material in Construction of Roads in Rural Areas." *Seminar on Road and Road Transport in Rural area,CRRI, New Delhi*. 260-267.

Chandra, S., Kumar, S. and Anand, K. R., (2005). "Soil Stabilization with Rice HuskAsh and Lime Sludge." *Journal of Indian Highway, Indian Road Congress, (May)* 87-98.

Garchar, V. and Trivedi, S.S. (2005) "Effect of Fly Ash on Engineering Properties of Expansive Soils." Proceedings of Indian Geotechnical Conference, Ahmedabad, 321-323.

Heeralal, M., Sadaiah, E., and Rao, Uma M., (2006). "Stabilization of Poorly Graded Sand with Lime-Fly ash." *Proceedings of the Indian Geotechnical Conference, Vol.I,* 755-758.

http://www.worldbank.org/html/dec/Publications/Abstracts97/03esd/esd22.html
Jha, J.N., and Gill, K.S., (2006). "Effect of Rice Husk Ash on Lime Stabilization." *Journal of the Institution of Engineers (India). November, Vol.87,* 33-39.

Mahin, D. B., (1990). "Energy from Rice Residues." *Biomass Energy and Technology Report. Winrock International Institute for Agricultural Development. Arlington, USA.*

Rao, A.V.N. (1999). "Characteristics of Fly ash its applications – A brief review." Proceedings`of the national seminar on Fly ash characterization and its Geotechnical applications, Fly ash Mission, TIFAC, Bangalore. 1-8.

Ravi Shankar, A.U., Suresha, S.N. and Kashinath, B. (2008). "Characterisation of Lateritic Soil Modified With Pond Ash and Cement." *Journal of Indian Highways, June,* 36(6), 21-27

Reddy, R.M., Kumar, N.D., Prasad, K.N. and Rao, K.M. (1999). "Beneficial use of fly ash in sub base courses." *Proceedings of the national seminar Fly ash characterization and its Geotechnical applications, Fly ash Mission, TIFAC, Bangalore.* 203-208.

Reddy, M.J. and Singh, B. (2008). "Experimental Investigation of Influence of Fly Ash and Cement on a Subgrade Soil." *Proceedings of the Indian Geotechnical Conference at Bangalore.* 186-189.

Roy, T. K., and Chattopadhyay, B. C. (2008). "Utilization of Rice Husk Ash and Pond Ash for Improving Subgrade in Road Construction." *Journal of the Institution of Engineers (India).* 89, 42-46.

Roy, T.K., and Chattopadhyay, B.C. (2008). "Behaviour of Sandy Soils Strengthened with Pond and Rice Husk Ash In Presence of Cement." *Proceedings of the Indian Geotechnical Conference at Bangalore.* 194-197.

Roy, T.K., Chattopadhyay, B.C. (2005). "Utilization of Alternative Materials for Civil Engineering Uses in Road." *Proceedings of the Nnational Conference on Geotechnique in Environmental Protection (GEN-05), IGS.*

Roy, T.K., Chattopadhyay, B.C. (2008). "A Study on the Effect of Cement on Alluvial Soil Strengthened With Pond And Rice Husk Ash For Construction of Road Subgrade." *Proceedings of the 12th International Conference of International Association for Computer Methods and Advances in Geomechanics at GOA.*

Roy, T.K., Chattopadhyay, B.C. and Roy, S.K. (2009). "Improvement of Subgrade of Roads with Addition of Rice Husk Ash." *Proceedings of the National Seminar on Emerging Trends in Ground Improvement, IGS, Kolkata*. 61-63.

Senol, A., Bin-Shafique, Md.S., Edil, T.B., and Benson, C.H., (2002) "Use of Class C Fly Ash for Stabilization of Soft Subgrade." *Fifth International Congress on Advances in Civil Engineering*, Istanbul Technical University, Istanbul, Turkey.

Tike, G.K., Kumar, S., Seehra, S.S., and Sikdar, P.K. (1999) "Utilization of Fly Ash in Construction of Base Courses." *Proceedings of the National Seminar on Fly ash Characterization and its Geotechnical Applications, Fly ash Mission, TIFAC, Bangalore.* 222-227.

Effects of Contaminants on Skid Resistance of Asphalt Pavements

Ping Cao[1], Xinping Yan[1,2,] Xiuqin Bai[1] and Chengqing Yuan[1]

[1] Reliability Engineering Institute, School of Energy and Power Engineering, Wuhan University of Technology, Wuhan 430063, China
[2] Prof., Engineering Research Center of Transport Safety, Ministry of Education, Wuhan 430063, China; PH: +86-27-86554969; FAX: +86-27-86549879; Email: xpyan@whut.edu.cn

Abstract

Skid resistance is one of the most important properties of asphalt pavements. Contaminants in asphalt pavements affect skid resistance at a certain extent. The detailed effects of contaminants on skid resistance of asphalt pavements were investigated. The core samples of AC-13, SMA-13, SHRP -12.5 and OGFC-13 compressed by the Superpave Gyratory Compactor (SGC) in the laboratory were used as pavement samples, which were usually applied in China. The lube oil, sand, earth, grit and water were chosen as contaminants. By testing skid numbers by British Pendulum Tester of the clean samples and the samples with contaminants in three contamination situations (little-covered, half-covered and all-covered), the effects of contaminants on skid resistance of dry and wet asphalt pavements were analyzed. It was proved that contaminants affected skid resistance of asphalt pavements distinctly. The knowledge gained in this study would provide guidance for the maintenance of asphalt pavements.
Key Words: asphalt pavements; contaminants; skid resistance; effects

Introduction

Tire-pavement skid resistance has long been recognized as the primary factor of controlling vehicle direction and speed, as well as ensuring short braking distances.

Lack of tire-road friction leads to skidding, which is one of the most important factors of potential traffic accidents (Fulop, 2000, Alexandros, 1998). Around the world, more than 1 million persons were killed yearly due to traffic accidents. Although high percentage of these accidents were due to drivers' errors, but highways had a significant effect on the high percentage of traffic accidents. The most important factor in the highways affecting traffic accident rates is the skid resistance. Accident rates increased in the rainy season especially after the initial rain showers. One of the main reasons for this increase was attributed to the low skid resistance of the highway surfaces (Ibrahim, 2007). Tire-pavement skid resistance was decided by tire, road surface and medium between tire and road surface. As medium, contamination (snow, oil, water, etc.) was one of the factors affecting road skidding resistance properties (Haas, 1994). Lindenmann (2006) and Kuttesch (2004) investigated the effect of water on skid resistance of asphalt pavement and discussed the relationship between skid resistance and wet traffic accidents. Norheim (2001) studied the effects of the structure and properties of ice and snow on the friction of aircraft tires on movement area surfaces. Wambold (1998) researched tire friction on contaminated surfaces and analyzed the effect of snow on skid resistance of asphalt pavements. Yaron (2005) used water blasting equipment to remove the contaminants on the road surfaces in the study on alleviating the skid resistance problem. Skid resistance of the road increased after the contaminants were removed.

Contaminants on road surfaces mainly came from roadside, passing vehicles and rain, etc. The lube oil, sand, earth, grit and water were chosen as contaminants in this study. The core samples of AC-13, SMA-13, SHRP -12.5 and OGFC-13 made by the Superpave Gyratory Compactor (SGC) in the laboratory were used as pavement samples, which were usually applied in China. By testing the British pendulum numbers of the clean samples and the samples with contaminants, the effects of contaminants to skid resistance of dry and wet asphalt pavements would be analyzed.

Experiments

Preparation of Road Samples. Gradations AC13 (Asphalt Concrete), SHRP12.5 (Superpave in Strategic Highway Research Program), OGFC13(Open Graded Asphalt Friction Course), SMA13 (Stone Matrix Asphalt) were usually used as pavements in asphalt road construction in China. Therefore, four categories of asphalt pavements (AC13, SHRP12.5, OGFC13, and SMA13) were chosen as road samples in the following experiments. Basalt aggregate (from Jinshan County in Hubei Province) was used in four asphalt concrete mixtures. The used gradations for these mixtures were the median recommended aggregate gradations by China Ministry of Communications for design of highway asphalt pavement (JTJ 014-97, 1997). The used gradations for four mixtures were shown in Table 1.

Table 1. Used gradation for four categories of asphalt pavements (%).

Category	Sieve opening size /mm									
	0.075	0.15	0.3	0.6	1.18	2.36	4.75	9.5	13.2	16
AC13	5.8	7.2	8.6	13.6	20.5	32.2	54.4	79.0	95	100
SMA13	10.6	11.7	12.5	13.4	14.7	16.9	29.3	65.2	91.6	100
OGFC13	4.5	5.2	5.8	7.4	9.5	13.1	25.7	61.4	90.7	100
SHRP12.5	5	7.3	7.8	13	20.4	32.8	52.7	76.1	94.2	100

Core samples with the diameter of 150 mm and height of 50 mm were compacted using Superpave Gyratory compactor. The gyratory compactor was used because it simulated field compaction, which can compact samples of 150 mm diameter, and the sample height can be controlled(Ibrahim, 2007). The samples with the diameter of 150 mm were used since the recommended contact path for the British Pendulum shoe on the asphalt concrete surface is between 124 mm and 127 mm (ASTM E303-93, 2000).

Skid Resistance Tests. For properties of cheapness and easily-operating, British Pendulum Tester was selected in the experiment since it was widely used for testing road surface skid resistance in traffic engineering in China. Modified British Pendulum Tester in the testing method of Ibrahim(2007) was adopted in this paper. The method and reason of Ibrahim(2007) to modify British Pendulum Tester were introduced as the followings. Because the British Pendulum was just designed to test field asphalt concrete surfaces and not designed to test asphalt concrete laboratory samples, its base was adjusted to accept 150 mm samples by constructing a special fixture to hold the samples firmly in place. In addition, the height adjusting screws of the pendulum were replaced by longer ones to accept the 50 mm height samples.
ASTM E303-93(2000) "Standard Test Method for Measuring Surface Frictional Properties using the British Pendulum Tester" test procedure was followed in testing skid resistance of road samples.

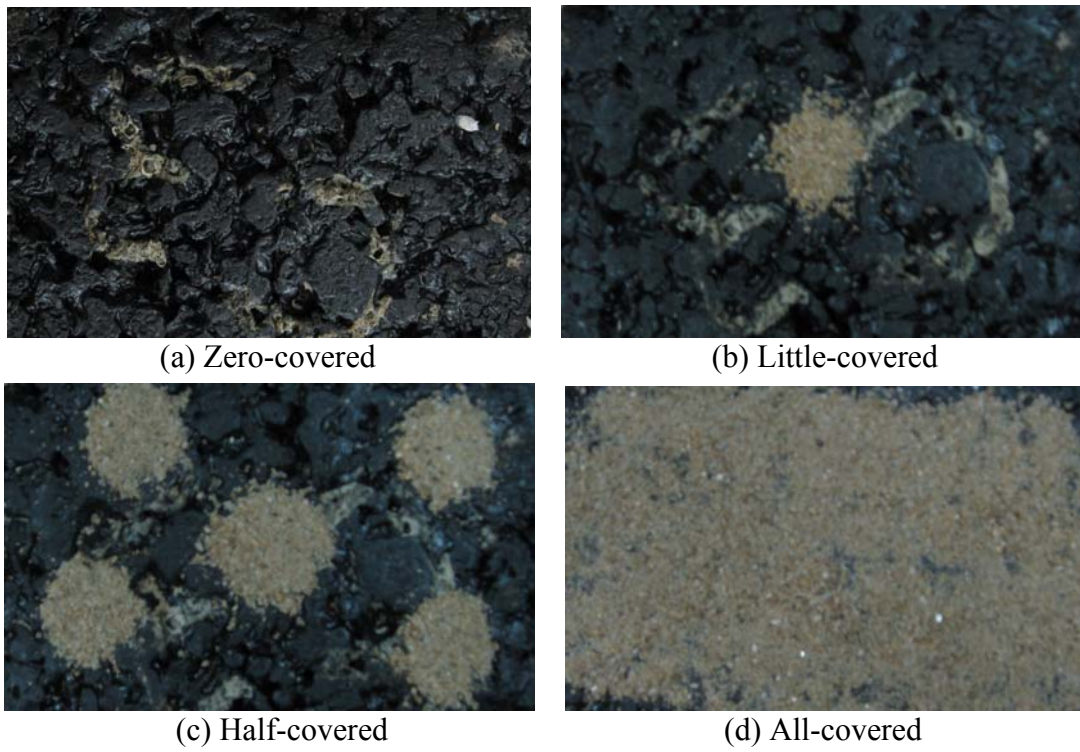
Contaminants Selection. Materials affecting skid resistance of asphalt pavement were defined as contaminants in this study. Materials such as lube oil leaked from vehicles, sand and grit brought from roadside by wind or flowing water, earth from tires of passing vehicles coming from earth road, water from rain would influence skid resistance of asphalt pavement. So lube oil, sand, grit, earth and water were chosen as contaminants in the experiments. Water was clean itself, but it affected skid resistance. So it was considered as a contaminant here. Shell SAE 90 automotive gear oil was used for simulating lube oil leaking from vehicles. Sand used in the experiments was composed of particles with diameter ranged from 0.01 mm to 0.2 mm. Therefore, grit had particles with diameter ranged from 0.5 mm to 2 mm

different from sand. Since earth could exist in all kinds of shapes, earth powder was used as a representative of earth. When water was added into earth powder, earth powder became mud in fact.

Results and discussion

The rubber block of British Pendulum shoe passed a quadrate area (126 mm×75 mm) when testing Skid Numbers of asphalt pavement laboratory samples. The quadrate area was called as contact area. Four expressions zero-covered, little-covered, half-covered and all-covered were used, to investigate effects of the contaminated degree of samples on skid resistance. When contact area on road sample surfaces was clean, it was described using zero-covered. When a small part of contact area was covered by contaminant, little-covered was used. Half-covered was used to describe the situation that a half part of contact area was covered by contaminant. When contact area was covered by contaminant completely, All-covered was used. These four situations were shown in Figure 1.

To study the effect of grit on skid resistance of AC13 asphalt pavement, Skid Numbers of AC13 road sample were tested by British Pendulum Tester when the sample was in four contamination situations. In each contamination situation, Skid Numbers of both dry sample and wet sample were tested. Experiments to study the



(a) Zero-covered (b) Little-covered

(c) Half-covered (d) All-covered

Figure 1. Four contamination situations of road samples.

effect of sand and earth on AC13 sample were done similar to grit. Experiments to investigate effects of sand and earth on skid resistance of SHRP12.5, OGFC13 and SMA13 were done similar to the front.

Water had good liquidity, so it was difficult to control the area covered with water on road samples. Skid numbers were only tested when all of contact area was covered by water on road samples.

When lube oil polluted road samples, it was hard to be removed. Because road samples were needed in other experiments, only AC13 asphalt pavement sample was used here. The effect of lube oil on skid resistance of AC13 asphalt pavement could provide references for investigating effects of lube oil on skid resistance of the other asphalt pavements.

Effects of grit on skid resistance of road samples. Effects of grit on skid resistance of road samples were shown in Figure 2. BPN was the abbreviation of British Pendulum Number which was the scale reading of British Pendulum Tester.

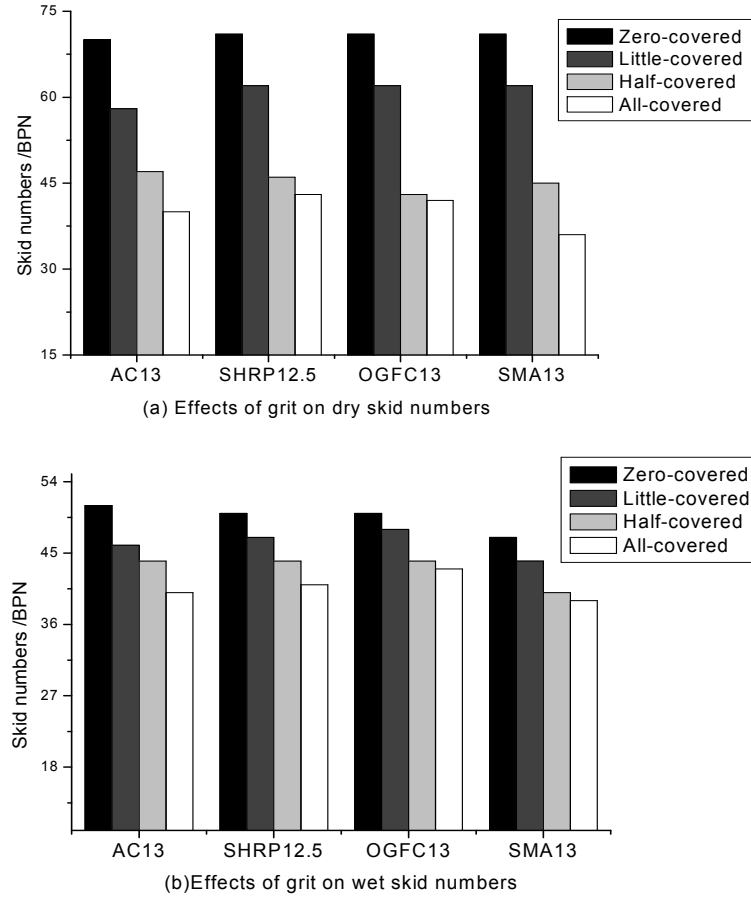


Figure 2. Effects of grit on skid resistance of four categories of road samples.

The effects of grit on skid resistance of dry road samples were shown in Figure 2(a). Dry skid numbers of four categories of road samples had a 13%-17% decrease in little-covered situation. The larger decrease of 33%-39% of skid numbers was found in half-covered situation. Skid numbers had the largest reduction of 39%-49% in all-covered situation. The effects of grit on skid resistance of wet road samples were shown in Figure 2(b). The 4%-10% reduction of wet skid numbers was found in little-covered situation. The larger decrease of 12%-15% was found in half-covered situation. In all-covered situation, skid numbers had the largest reduction of 14%-20%. From the front analyses, it was found that both dry skid numbers and wet skid numbers were affected obviously by girt and reduced with the increase of the area covered by grit in contact area. Dry skid numbers had much larger reduction than wet skid numbers.

The reasons for the effects of grit on skid resistance of asphalt pavements could be that grit changed the friction conditions between road surfaces and the rubber block of British Pendulum Tester. When road surfaces of asphalt pavements were clean, sliding friction occurred between the rubber block and road surfaces. But three-body friction occurred among the rubber block, grit and road surfaces, when road surfaces were covered by grit. The particles of grit rolled between the rubber block and road surface. Sliding friction force was obviously larger than rolling friction force, so skid numbers reduced when road surfaces were contaminated by grit. More area of contact area was covered by grit, more grit particles rolled between the rubber block and road surface. So skid numbers reduced with the increase of area contaminated by grit. When grit was added into wet road surfaces, the rolling of grit particles were restricted in a certain degree by water. Simultaneously grit particles could bread up water film between rubber block and road surface and it was useful for restricting the reduction of wet skid numbers. Therefore, reduction of wet skid numbers was smaller than that of dry skid numbers.

Effects of Sand on Skid Resistance of Road Samples. Effects of sand on skid resistance of road samples were shown in Figure 3. It was found that sand had similar effects on skid numbers of asphalt pavements. The effects of sand on dry skid resistance of road samples were shown in Figure 3(a). Dry skid numbers of four categories of road samples reduced 9%-16% in little-covered situation, 29%-32% in half covered situation and 30%-33% in all-covered situation. Figure 3(b) showed the effects of sand on wet skid resistance of asphalt pavements. Reduction of wet skid numbers about 4%-10% in little-covered situation, 8%-12% in half covered situation and 10%-14% in all-covered situation was discovered. From zero-covered situation to all-covered situation, the reduction degree of dry skid numbers and wet skid numbers increased. Sand had much larger effects on dry skid numbers than that on wet skid numbers.

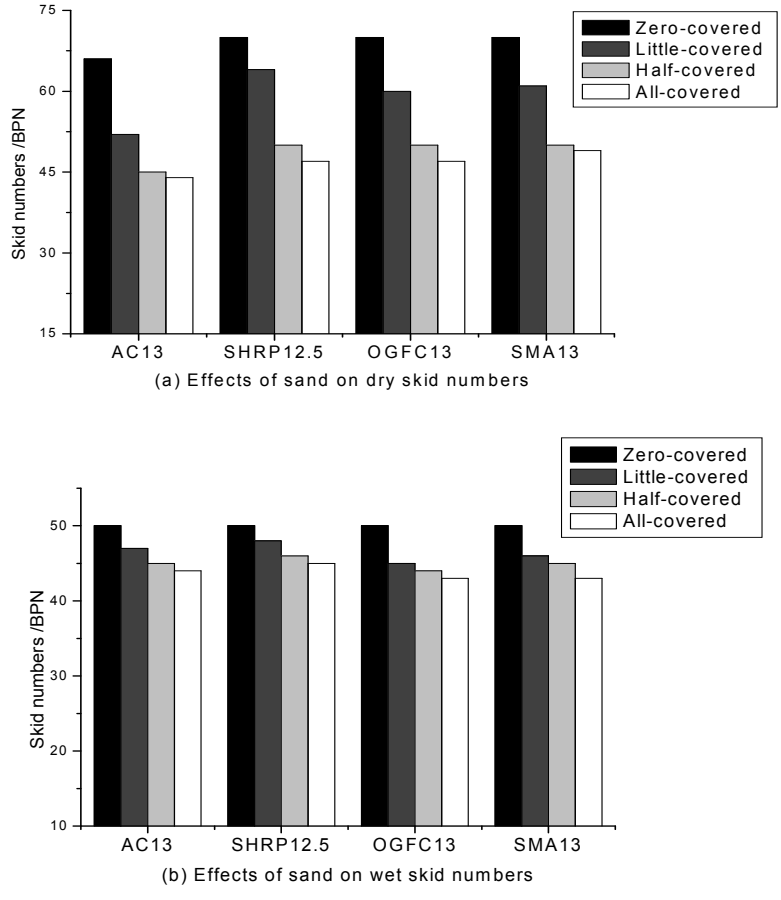


Figure 3. Effects of sand on skid resistance of four categories of road samples.

The reasons for effects of sand on skid resistance of asphalt pavements were different from that of grit. When sand was added into road surfaces, sand filled the gap between asperities in road surfaces for smaller particle diameter than grit. This affected the contact between rubber block and road surfaces. For this reason, dry skid numbers and wet skid numbers decreased when road surfaces were covered with sand.

Definitely, effects on skid numbers for sand particles filling the space between asperities were smaller than that for girt particles rolling between rubber block and road surface. So effects of grit on skid resistance of asphalt pavements were worse than that of sand. Sand particles with smaller diameter could break up water better than grit particles, so sand affected wet skid numbers less than grit did.

Effects of earth on skid resistance of road samples. Effects of earth on skid resistance of road samples were shown in Figure 4. The effects of earth on dry skid resistance of road samples were shown in Figure 4(a). Dry skid numbers of four

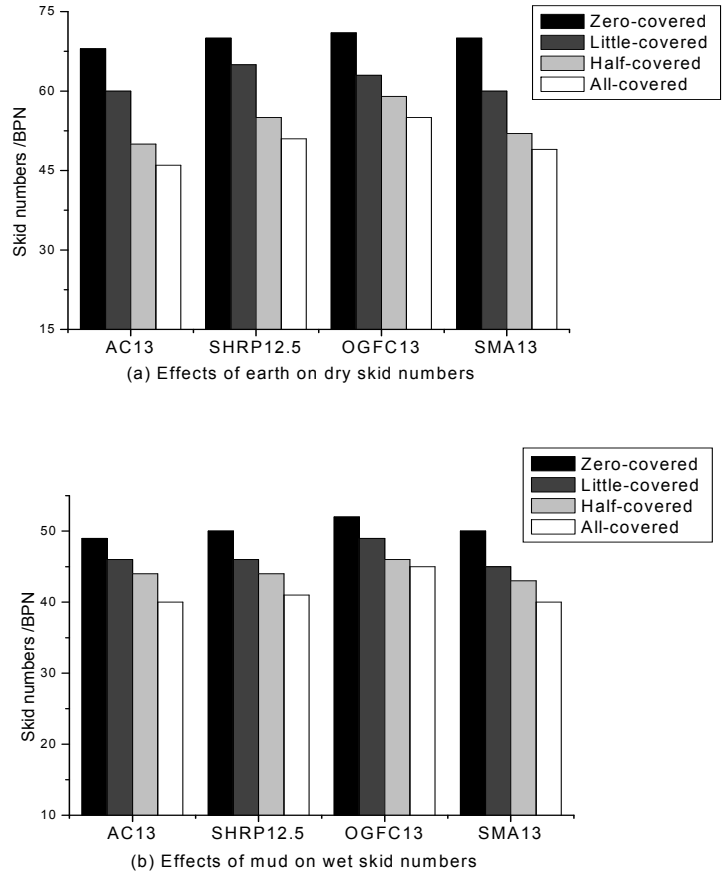


Figure 4. Effects of earth on skid resistance of four categories of road samples.

categories of road samples reduced 7%-14% in little-covered situation, 14%-26% in half covered situation and 27%-32% in all-covered situation. Figure 4(b) showed the effects of mud on wet skid resistance of asphalt pavements.

Wet skid numbers reduced 6%-10% in little-covered situation, 10%-14% in half covered situation and 18%-20% in all-covered situation. The effects of earth on skid resistance were similar to that of sand.

The reasons for effects of earth on skid resistance were also similar to that of sand. Difference only existed in the reason for effects of mud (wet earth) on wet skid numbers. When mud was added on road surfaces, mud mixed with water formed muddy water film. This kind of muddy water film had better lubrication role. So mud had worse effects on wet skid resistance than sand.

Effects of water on skid resistance of road samples. The effects of water on skid resistance of road samples were shown in Figures 2 to 4. Skid numbers in

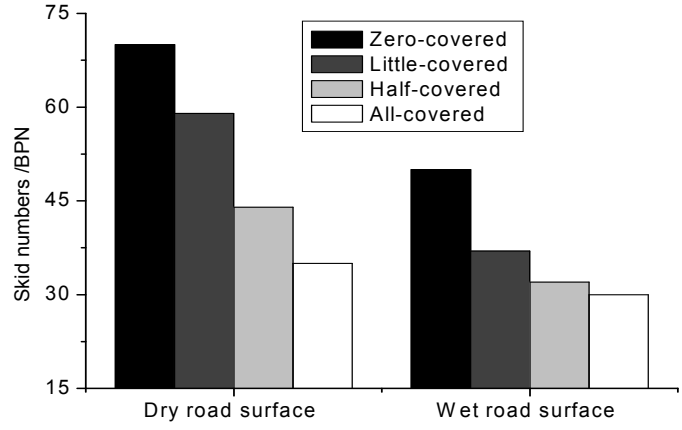


Figure 5. Effect of lube oil on skid resistance of AC13 sample.

zero-covered situation in Figure 2(a), Figure 3(a) and Figure 4(a) were dry skid numbers of four clean road samples. Wet skid numbers of road samples were skid numbers in zero-covered situation in Figure 2(b), Figure 3(b) and Figure 4(b). It could be found that skid numbers of road samples reduced 26%-29% when all contact area was covered by water. Skid resistance of asphalt pavements was affected hardly by water.The reason for effects of water on skid resistance of asphalt pavements might be lubrication role produced by water film.

Effect of lube oil on skid resistance of AC13 samples. The effect of lube oil on skid resistance of AC13 samples was shown in Figure 5.

Dry skid numbers of AC13 asphalt pavement sample reduced 16% in little-covered situation, 37% in half-covered situation and 50% in all-covered situation. Wet skid numbers reduced 26% in little-covered situation, 32% in half-covered situation and 40% in all-covered situation. It was found that lube oil had the worst effects on both dry skid numbers and wet skid numbers of AC13 sample among grit, sand, earth, water and lube oil. Wet skid numbers had about 30% reduction when a small part of contact area was covered by lube oil specially.

The bad effects of lube oil on both dry skid resistance and wet skid resistance of AC13 asphalt pavement attributed to good lubrication role of lube oil and the mixture of water and lube oil.

Conclusions

The effects of contaminants on skid resistance of asphalt pavements were analyzed. The reasons for the effects of contaminants on skid resistance were discussed. The following conclusions can be obtained.

(1) Skid resistance of asphalt pavements was hardly affected when road surfaces were covered by contaminants. Lube oil had the worst effect on skid resistance of asphalt pavements.

(2) Effects of contaminants on skid resistance of asphalt pavements increased with the increase of area covered by contaminants in contact area. The largest effect of contaminants on skid resistance occurred in the situation that all contact area was covered by contaminants.

(3) Contaminants had a potential threat to traffic safety. The maintenance department of asphalt pavements should remove contaminants in time by reasonable methods to ensure traffic safety when road surfaces were found to be contaminated.

Acknowledgments

The authors would like to express their sincere gratitude to the National Natural Science Foundation of China for funding this study (No. 50678140).

References

Alexandros, G. K. and Olympia, K. P.(1998). “Fractal evaluation of pavement skid resistance variations i: wetting surface.” Chaos, *Solitons & Fractals*, 1875-1890.

ASTM E303-93(2000). *Standard Test Method for Measuring Surface Frictional Properties Using the British Pendulum Tester*, West Conshohocken, ASTM.

Fulop, I. A., I. Bogardic, Gulyas, A. and Tarpay, M. C. (2000). “Use of Friction and Texture in Pavement Performance Modeling.” *Journal of Transportation Engineering*, 243-248.

Haas, R., Hudson, R. and Zaniewski, J. (1994). *Modern Pavement Managemen*t, Krieger Publishing Company Press, Malabar, 23-27.

Ibrahim, M. A. (2007). “Evaluating skid resistance of different asphalt concrete mixes.” *Building and Environment*, 325–329.

JTJ 014-97(1997). *Specifications for Design of High Way Asphalt Pavement*, Ministry of Communications of China, Beijing.

Kuttesch, J. S. (2004). *Quantifying The Relationship between Skid Resistance and Wet Weather Accidents for Virginia Data*, Virginia polytechnic institute and state university Press, Virginia, 1-3.

Lindenmann, H. P. (2006). "New findings regarding the significance of pavement skid resistance for road safety on Swiss freeways." *Journal of Safety Research,* 395–400.

Norheim, A., Sinha, N. K. and Yager, T. J. (2001). "Effects of the structure and properties of ice and snow on the friction of aircraft tires on movement area surfaces." *Tribology International*, 617–623.

Wambold, J. C. and Andresen, A. (1998). "The measurement and theory of tire friction on contaminated surfaces." *Transportation Conference Proceedings*, 94-98.

Yaron, R. and Nesichi, S. (2005). "Alleviating the skid resistance problem- The Israeli experience." *The First International Conference on Surface Friction for Roads and Runways*, Christchurch Zealand, 135-145.

Using Mineral Fibers to Improve Asphalt and Asphalt Mixture Behavior

Lihua Zhao[1,2], Jingyun Chen[2] and Shengwu Wang[1]

[1] School of Civil and Security Engineering, Dalian Jiaotong University, Dalian 116028, China
[2] Dalian University of Technology, School of Civil Engineering Water Resources, Dalian, 116024, China

Abstract

To explore a new type of basalt mineral fibers on the impact of asphalt mixture, fiber improving asphalt behavior was researched first. Through the test of asphalt mixture of three indicators with fiber content 0%, 0.2%, 0.3%, 0.4%, 0.5%, and 0.6%, characteristics of composite materials (fiber asphalt)were analyzed based on Einsteins viscosity equation. The effect of different fiber content on the Marshall stability of asphalt mixture, the residual splitting strength and water stability of Marshalls was also researched. The research indicated that basalt fiber had a significant toughening effect on asphalt, and penetration and the softening point to evaluate fiber asphalt mixture's viscosity properties of composite materials was appropriate. MS and MS0 results showed that basalt mineral fibers in the AC-16C were not more than 0.4% volume. With the increase of fiber content, the splitting strength of the asphalt mixture increased. When content > 0.30%, increase was not obvious but modulus and stiffness continued to increase to resist deformation capacity. Composite micromechanics theory explained the basic principles of basalt mineral fibers to improve the behavior of asphalt mixture strength.
Key Words: road engineering; mineral Fiber； asphalt viscosity； fiber content；

Introduction

For the past few years, fiber has been used widely in the research of asphalt mixtures. Many scholars(Reed B F.1996; Airey G D.2004; Dipl-Ing.2003; Cooley

L.A.2002)have done a lot of work and achieved plenty of research results on strengthening the function of an asphalt mixture with lignin fiber, synthetic fiber, and asbestos fiber, which were widely used on plus paving top facing, mending crevices, etc. Short-cut basalt mineral fiber is a new-style fiber, which can remedy the defect of low strength and elasticity coefficient of organic textile fiber, bad thermostable performance, environment pollution of mineral fiber like asbestos etc., and it can be recycled also. In the 1990s, the first road of concrete pavement with mineral fiber (basalt textile fiber) was paved in George Asia of the USA and made a good road performance. But just like one coin has two sides, it shows a discrepancy in different conditions. So, further research is needed. Currently, it is a very popular research topic when it comes to improving asphalt concrete with mineral fiber (Peng 2009).
The research at present is starting macroscopically (Fu et al. 2007; Xiang and Liu 2005; Chen et al. 2002) , using experiment as the main means to research the behavior of asphalt mixture with fiber. This test is based on experiments, analyzed the theory of using basalt mineral fibers to improve asphalt viscosity with generalized comositerite theory in rheology, and analyzed the basic function of asphalt mixture through the Marshalls experiment, the strength of the splitting experiment and water stability experiment, and also explained the rationale of basalt mineral fibers improving the strength performance of an asphalt mixture with compound material mesomechanics theory.

Viscosity Function of Basalt Fiber Improving Asphalt

Generalized composite rate theory. The mathematical function of composite material generalized composite rate is as follows (Ji SC.A, 2004):

$$M^j = \sum_{i=1}^{N} (V_i M_i^j) \tag{1}$$

where "M" is a mechanical property, like Young modulus (E), shear modulus (G), volume modulus (K), or ductility flow intensity(σ_f)； "j" is microstructure modulus with regard to two-phase forms made up with strong phases and weak phases. When a strong phase supports a structure, $j > 0$; when a weak phase supports structure, $j < 0$; "V" is volume fraction.

With regard to solid and liquid suspension system supported with weak phase, Eq. 1 can be written:

$$M = \left[(1 - V_S) M_w^j + V_S M_s^j\right]^{1/j} = M_w \left[1 + V_S \left(\left(\frac{M_S}{M_w}\right)^j - 1\right)\right]^{1/j} \tag{2}$$

where "*s*" stands for solid granule, and subscript "*w*" stands for weak solid granule. Einstein (1906) described liquid viscosity increase because of stream surface curved and then energy dissipated when solid granule existed in his classical thesis "Discussing about determine of numerators' sized"(Einstein, 1906). Batchelor and

Green (1972) described energy dissipate because of mutual reciprocity of suspension granule (just like collision, reciprocity, attraction, or exclusion) leads to the increase of liquid viscosity.(Batchelor and Jones et al. (1991)，Patankar and Hu (2001) described suspension granules displacement in shear surface during the course of shear flowing, which leads to liquid viscosity varying. The theory above can be shown using Eq. 2. Einstein made some supposition and then deduced the famous viscosity formula of mixture of solid and liquid as shown in Eq.3:

$$M = M_w\left(1 + K_E V_S\right) \tag{3}$$

M stands for the viscosity of mixture; *Mw* stands for the viscosity of original liquid, K_E is Einstein coefficient, which is relative to the ratio of length to diameter (L/D); and *Vs* is volume fraction of solid granule. Eq.3 was inferred by Einstein based on some hypothesis. Though real suspending liquid is hard to satisfied ideal condition, plenty of dilute suspending liquid proved this function's value.

Experimental result of fiber improving bituminous. The experiment chose 90# road asphalt of the Liao river oil field and short-cut basalt mineral fiber exclusively used for road asphalt mixture, which was made by the Zhejiang GBF company, and the technical index was shown in Table 1. Mineral fiber was added into asphalt with the percentages of 0%, 0.2%, 0.3%, 0.4%, 0.5%, and 0.6%. After mixing and stirring with a small sized agitator, the mixture was mixed and stirred again by manpower to make sure that mineral fiber dispersed in asphalt evenly. Then the experiment of needle penetration, softening point, and ductility was conducted according to experimental standards (JTJ052-2000). The results were shown in Table 2.

Table 1. Property of basalt fiber.

Technique Data	Technical parameter	Chemical constituent	Percentage composition (%)
Fiber length (mm)	6	SiO_2	57.2
Fiber diameter (μm)	15	Al_2O_3	18.2
Rate of absorbent oil (%)	70	CaO	5.5
Maximum moisture content (%)	0.5	MgO	5.0
Relative density (g/cm^3)	2.463	Na_2O+kO	4.6
Tensile strength (MPa)	3000-4840	TiO	1
Modulus (GPa)	93.1-110	Fe_2O_3+FeO	8.1
Moisture absorption (%)	0.5	Others	0.4

Table 2. Property of asphalt binder.

Test index Unit	Penetration (25 °C， 100g， 5s)0.1mm				softening point (ball and ring method) °C				ductility (5cm/min,25 °C)cm	
	1dkp		2dkp		1dkp		2dkp		A	B
Fiber content (%)	0	89.4	90.2	88.4	89.8	51.9	51.5	52.0	52.3	117.6	131.5
	0.2	86.4	85.8	85.9	85.7	54.4	54.8	56.3	55.9	35.1	23.8
	0.3	82.0	80.2	81.3	81.5	56.7	56.3	57.7	58.2	17.2	19.5
	0.4	74.0	73.6	71.3	72.4	59.2	58.9	60.5	60.3	18.5	20.0
	0.5	62.7	63.0	63.8	62.9	60.9	61.2	62.1	61.8	20.7	15.2
	0.6	56.6	51.7	54.3	53.4	63.1	63.5	64.3	63.9	14.3	28.9

From the experimental data above, we can see that needle penetration, ductility, and softening point of asphalt mixture changed when fiber was added. Therefore, needle penetration and softening point show obvious linear law, as shown in Figures 1 and 2. Needle penetration and softening point are indirect expressions of asphalt viscosity and widely used because of its simple experiment. Asphalt mixture is a two-phase composite material that is made up with asphalt and fiber. Its viscosity function can be analyzed using composite material's generalized rule of mixture. With regard to asphalt mixture, the scale of weak phase asphalt is tremendous, it is in domination position. So j<0 in formula (2),and $(M_S/M_w) \to \infty$， so $(M_S/M_w)^j \to 0$ Eq. 2 simplified to:

$$M = M_w(1 - V_S)^{1/j} \tag{4}$$

where, M means needle penetration (or softening point) of asphalt mixture, Mw means needle penetration(or softening point) of asphalt, Vs means mass fraction of fiber. We can easily get the general function (5)from figure ure . More significant conclusions is that it is suited to use needle penetration and softening point to appraise the viscosity of asphalt mixture with basalt fiber.

$$M = M_a(1 + m_S + C) \tag{5}$$

In Eq. 5, M means fiber asphalt mixture's needle penetration (P) and softening point ($T_{R\&B}$), Ma means asphalt's P or $T_{R\&B}$, m_s means fiber's mass fraction , C means constant.

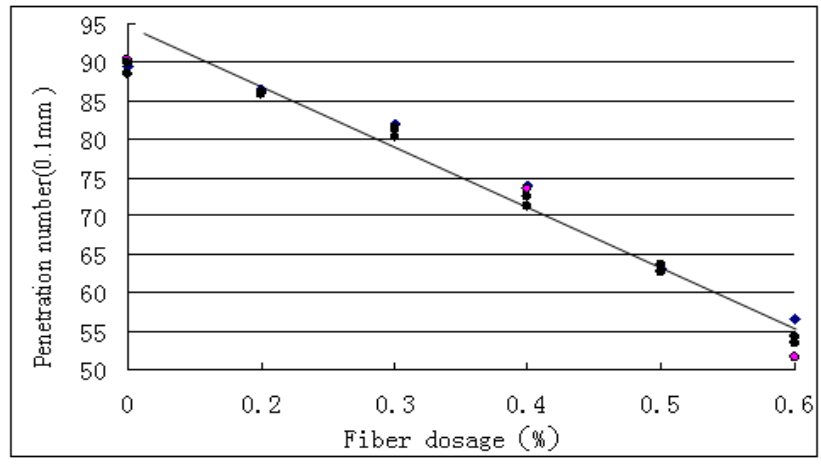


Figure 1. Relation curves between fiber content and needle penetration.

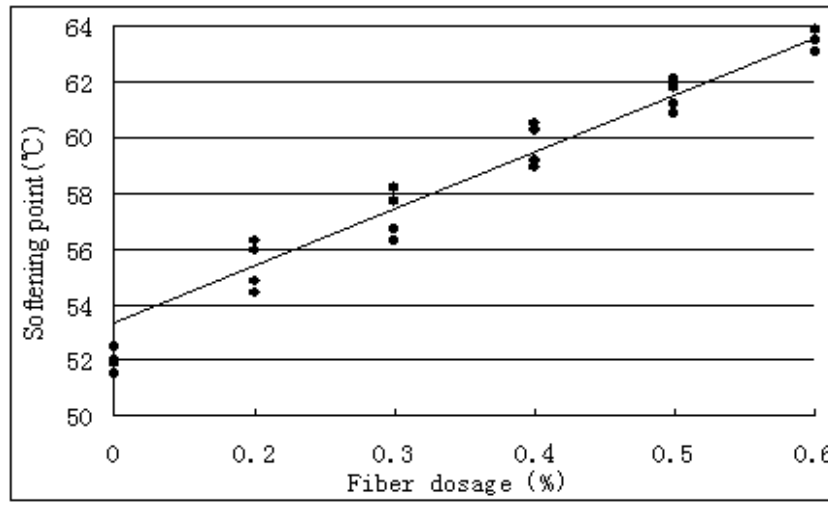


Figure 2. Relation curves between fiber content and softening point

Asphalt mixture ductility rapidly reduces when adding fiber, because fiber has no stickiness itself, and asphalt and fiber's adhesion is lower than that of asphalt's. Therefore, slipping happened in the interface between asphalt and fiber in ductility experiments, and using ductility index to appraise the viscosity function of asphalt mixture with fiber is not effective.

Basalt fiber to improve performance of asphalt mixture

Aggregate and gGradation. The coarse aggregate (19-2.36 mm) used in the experiment is limestone in Dalian, china, and fine aggregate (0.075-2.36mm) is clean river sand, and the filler is the limestone powder. Aggregate technical index is in line with experimental protocols require (JTG E42-2005). Aggregate gradation is the median recommended AC-16C, as shown in Table 3.

Table 3. Gradation of AC-16C dense-graded asphalt mixture.

Mesh size (mm)	16	13.2	9.5	4.75	2.36	1.18	0.6	0.3	0.15	0.075
Passing rate (%)	98	83	68	53	41	30	22	16	11	6

Table 4. Marshall stability of AC-16C dense-graded asphalt mixture.

Fiber content (%)	0	0.2	0.3	0.4	0.5
Optimum proportion of asphalt (%)	5.1	5.5	5.6	5.4	5.3
Marshall stability (KN)	17.2	17.7	15.3	14.1	8.5
Residual stability (%)	76.4	77.6	86.4	91.82	87.3

Marshall test analysis. Based on the optimal oil-stone ratio from Marshall asphalt mixture testing, six standard samples were molded for the Marshall test and divided into two groups, one group was soaked in water with a constant temperature 60℃ for 30-40 minutes, another was soaked in water with a constant temperature 60℃ for 48 hours. Marshall stability tests were conducted under the different curing conditions to calculate the Marshall residual stability MS_0. Experimental results are presented in Table 4.

Based on the test, the analysis of experimental results is below:

With the increase of fiber content, the optimal oil-stone ratio increased first the decreased, which indicate that when fiber content≥0.4%, the existing mixing process has difficulty dispersing the fiber uniformly, resulting into the following groups of fiber specific surface area not increasing and effective fiber fraction not enhancing. The maximal Marshall stability is 17.7kN when the fiber content is 0.20%. From the results of Marshall stability, there is a reasonable content of basalt mineral fibers in the asphalt concrete.

Admixture fibers can improve the water stability of asphalt mixture, but when fiber content>0.40% residual stability declined slightly. It indicates that there is also an optimal content of basalt mineral fibers for water stability of asphalt concrete.

Splitting test analysis. Splitting test was used to determine the mechanical properties of asphalt mixture at a specified temperature and loading rate. The experiment uses the standard Marshall test sample with temperature 15 ^{o}C, the asphalt concrete splitting test results with different fiber content were shown in Table 5.

Table 5. Splitting strength of AC-16C dense-graded asphalt mixture.

Fiber content (%)	0	0.2	0.3	0.4	0.5
Breaking load (KN)	7.2	8.5	9.7	9.9	10.1
Splitting strength (mpa)	0.740	0.839	0.958	0.979	0.997
Stiffness modulus (mpa)	149.314	204.145	245.559	285.323	313.602

From the test results of Table5, we can see that fibers can increase the asphalt mixture splitting strength and stiffness modulus. It can prove that basalt mineral fibers can enhance the anti-cracking performance of an asphalt mixture, and fiber asphalt mixture increased the capacity of resisting deformation.

Increasing of splitting strength of fiber asphalt mixture reflects the fiber tensile strength of asphalt mixture of "reinforced" effect. According to the theory of composite micromechanics strength (Du 1998), splitting tensile strength of fiber asphalt mixture can be functioned:

$$f_s = f_{sfp}\rho_{fp}(1 - K_a\rho_a) + f_{ab}(S_a/V_a) \tag{6}$$

where, f_{sfp} is the tensile strength of fiber asphalt; ρ_{fp} is the percent by volume of fiber asphalt; K_a is the biggest stress concentration factor of aggregate embedding and crowding, and for AC mixture, $K_a = 0$; ρ_a is aggregate's percent by volume; f_{ab} is aggregates-asphalt interface bonding strength; S_a/V_a is unit volume within the aggregate ratio of surface area and volume.

Mixing fiber improved the value of f_{sfp}, because the skeleton structure network of random distributed fibers, on the one hand, increased the viscosity of asphalt; on other hand, this space net structure can be across of the inner pore and crack of asphalt mixture and form a "bridge fibers", making crack propagation of the energy release rate decrease, which plays the role of delaying or stopping the asphalt mixture crack. Mixing fiber increased the value of S_a/V_a, because the fiber diameter is only 15μm, which will greatly increase the surface area of aggregates. From the analysis above, the strength of composite micro-mechanics theory proved the basic principles that adding fibers can improve splitting strength of asphalt mixture.

Conclusions

Through the analysis of test results of the admixture of basalt mineral fiber asphalt and asphalt mixture, some valuable conclusions were obtained:

(1) Based on generalized composite rate analysis theory for fiber-modified asphalt viscosity, it is applicable that the index of needle penetration and softening point

evaluate asphalt mixture fibers viscosity.

(2) The changes of optimal asphalt content reflect the dispersion of fiber of asphalt mixture, and adding an appropriate amount of fiber causes the increase of fiber optimal asphalt content of asphalt mixture.

(3) Marshall stability and residual stability results indicate that the basalt mineral fibers in the asphalt mixture should not be greater than 0.4%. The issue that fibers in the asphalt mixture improve the low-temperature behavior, high temperature behavior, and fatigue performance remains to be studied further.

(4) Asphalt concrete splitting strength increased with the admixture of fiber, but the dosage was higher than 0.30%, the increase was not obvious, and the ability to resist deformation increases with the fiber increase.

(5) According to composite materials micromechanics strength theory analysis, fiber-increased splitting strength of asphalt mixture is the result that fiber-increased asphalt cement strength and interface strength between aggregate and asphalt.

References

Airey G D,Rahimzadeh B, and Collop A C.(2004) “Linear rheological behavior of bituminous paving materials.” *Journal of Materials in Civil Engineering*, (4), 212-220.

Batchelor GK,Green JT.(1972). “The determination of the bulk stress in a suspension of spherical particles to the order”, *Fluid Mech*.56, 401-427.

Chen Hua-xin,Zhang Zheng-qi, and Hu Chang-shun.(2002) “Mechanism research of the fiber reinforced asphalt concrete.” *Journal of Chang’An University(Natural Science Edition*, 22(6), 5~7.

Cooley L.A.Brown E.R, and Watson D.E.(2002) “Evaluation of open-graded friction course mixtures containing cellulose fibers.” *Transportation Research Record,* (1723), 19-25.

Dipl.-Ing,Ludwig, and Steve B.B.(2003). “The influence to the mechanical properties of bituminous mastic and,thus to open grain asphalt using synthetic fibers.” *AAPT 12th International Flexible Pavements Conferenc,* 1-9.

Du Shan-yi.(1998). *Micromechanics of composites*, Science Press.

Einstein A.(1906). Eine neue Bestimmung der Molekuldimensionen.Ann.Phys.(Leipzig), 19, 289-306.

FuLi-qiang, Wang Zi-ling, and Zhang Rui.(2007). "Modification Effect Analysis of SBS and Fiber." *Jounal of Highway and Transportation Research and Developmen*, (5), 26-29.

Ji SC.A generalized mixture rule for estimating the viscosity of solid-liquid suspensions and mechanical properties of polyphase rocks and composite materials, Geophys.Res., 109:B10207,doi:10,1029/2004JB003124.

Jones DAR,Leary B,Boger DV.(1991). "The rheology of a concentrated colloidal suspension of hard spheres." *Colloid Interface Sci*,147, 479-495.

JTJ 052-2000,Highway engineering asphalt and bituminous mixture experiment regulations.

Patankar N,Hu HH.(2001). "Rheology of a suspension of particles in viscoleastic fluids." *Non-Newtonian Fluid Mech*.96, 427-443.

Peng Guang-yin,Qian Zhen-dong Fu Dong-liang.(2009) "Research on Pavement Performance of the short-cut basalt fiber-reinforced asphalt mixture." Petroleum Asphalt, (1), 8-11.

Reed B F,James L,Burati J R.(1996). "Polyester fibers in asphalt paving mixtures." *Journal of the Association of Asphalt Paving Technologists*.65(l), 65-66.

Shu Xiang, Liu Li-xin.(2005). "Difference and Comparison of Bitumen Modification and Mineral Fibers." *Highway*, (9), 171-175.

Dynamic Analysis of Damage Behavior of Vibration Compaction on Concrete Bridges

Xin Yu[1] and Yuhua Fu[2]

[1] Hohai University, School of Civil Engineering and Transportation, Xikang Raod #1, Nanjing, P.R.China, 210098, Email: hhu_yuxin@163.com
[2] Suzhouraocheng Freeway Co.,Ltd. , Suzhou, P.R.China, 215134

Abstract

To study whether vibration compaction on concrete bridge decks can cause damage to a bridge structure, based on the theory of vibration modal, the paper analyzed dynamic vibration modal and damping of common types of bridges. The analysis using dynamic vibration theory calculated the top four orders vibration modal and corresponding frequency of a common bridge and compared them with the frequency of a vibratory roller. The results showed that the fundamental frequency of the common bridge types first-order vibration modal was between 5–7 Hz (Zhao et al. 2008), while the frequency of a vibratory roller is 30 Hz. Usually, the top 30 orders of frequency of the common bridge can't reach the frequency of the vibratory roller. As a result, the probability of roller vibration damaging the overall bridge is small. Vibrating rollers can be used on the deck pavement compaction.
Key Words: vibration compaction; concrete bridge deck; vibration modal analysis; vibration modal theory; damage behavior

Introduction

Bridge deck is a typical vibration system with small damping and large amplitude. The bridge will vibrate under the vehicle load. If the load frequency is the same as the frequency of the bridge deck, the bridge deck will be damaged because of bridge resonance (Xie. 2007). Worried about the bridge deck damage, the vibration of the vibratory roller is often prohibited in China when compacting on the PCC bridge deck pavement which may result in insufficient compaction and large porosity of bridge deck pavement (Specification. 2004). Inadequate compaction is an important reason for the early damage of bridge deck pavement leading to early damages of bridge deck pavement such as water damage and so on. Studying the impact factors of bridge deck pavement compaction has practical significance.

Based on kinetic analysis, using vibration modal analysis theory, this paper calculated the first few orders vibration modal and the corresponding frequency of the typical bridge structure, then compared them with the typical frequency of a vibratory roller and analyzed the possibility of vibration compaction on bridge deck pavement resulting in bridge damage because of the resonance. The last bridge deck pavement compaction process was recommended.

Numerical Simulation of Typical Bridge

Plate girder bridge, T girder bridge and box girder bridge are typical bridges used on freeway in china. The body of the Plate girder bridge is often precast hollow slab, generally with the span under 25 meters; T girder bridge with longitudinal stiffening rib, generally the span is between 25-35 m; box-girder bridge with cabinet structure for large-span bridge structure, generally span between 40-80 m with the largest span up to 100 m.

This paper assumed that the cement concrete bridge materials modulus was 30000 Mpa with Poisson's ratio of 0.2 and simplified the constraint as fixed at both ends. The load of the vibratory roller was placed at the center of the bridge, and roller load used the common engineering parameters.

Load field setting of bridge deck pavement. The vehicle load is very complicated, generally simplified as a triangle or semi-sine wave load during analysis (Toshiyuki et al. 2009). In this paper, the load variation followed the simplified form given by Huang Yang-Xian in the Kenlayer procedure, which simplified the vehicle dynamic load acting on the deck pavement as equivalent round load, and the load distributed as uniformly distributed half-sine wave as shown in Eq. 1:

$$q(t) = q_{\max} \sin^2(\pi t / T) \tag{1}$$

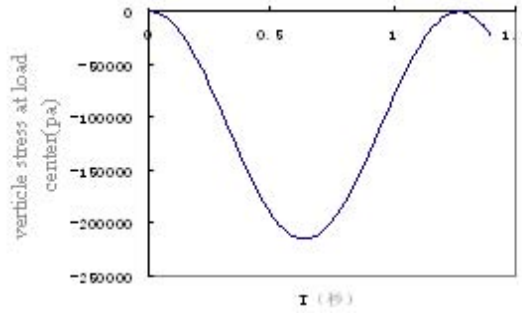
Single-cycle load function can be expressed through Eq. 2:

$$q(t) = q_{\max} \sin^2(\omega t) = q_{\max} \sin^2(\pi t / T) \qquad 0 \le t \le T \tag{2}$$

$$q(t) = 0, \qquad t > T$$

$$T = 12a / v \qquad \omega = \pi / T$$

In Eq. 2: a is tire contact area radius, and v is vehicle running speed.



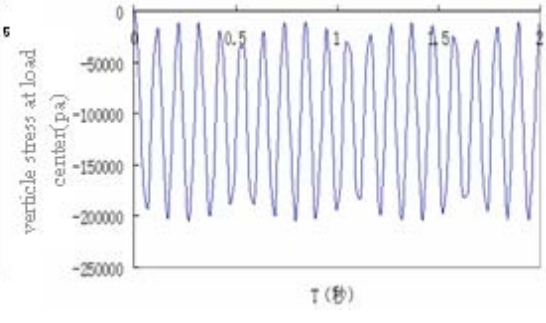


Figure 1. Surface stress of stationary roller. **Figure 2.** Surface stress of vibratory roller.

The paper assumed the roller load range is 2 m in length with a width of 0.2 m during the bridge deck pavement construction. The stationary roller load is 10 tons, and load set degree is 0.25 Mpa. With the roller speed 3–5 km/h for example, through the equation mentioned above, the roller dynamic load frequency is about 2 Hz. Assumed the frequency of vertical vibratory roller is 30 Hz with a pressure force of 100 KN, the surface stress distribution of roller with vibration and no vibration are

shown below, respectively.

Establishment of typical bridge model **Plate girder bridge.** The deck of a plate girder bridge is hollow precast plate, simplified as rectangular entity for facilitating the calculation. The paper assumed the bridge deck was 15 m in width with the span of 20 m and 30 m, as shown in Figures 3 and 4.

T girder Bridge. We assumed the width of a T girder bridge deck was 15 m with the web spacing as 2 m, and the web height was 1.7 m. The thickness of webs and the deck were the same as 0.3 m. The sectional form was shown in Figure 5. The model of a 20 m span was shown in Figure 6, and the model of a 40 m span was shown in Figure 7.

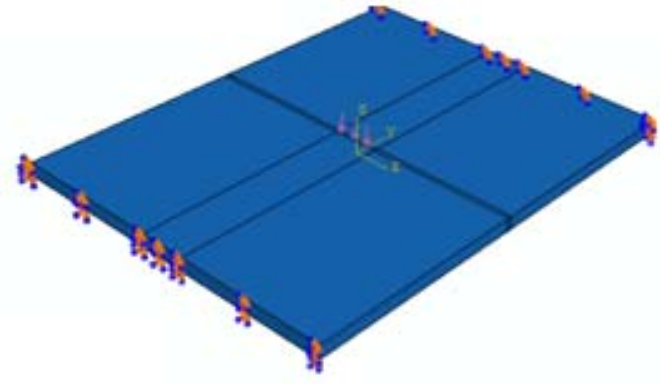


Figure 3. 20 m plate girder bridge deck model**.**

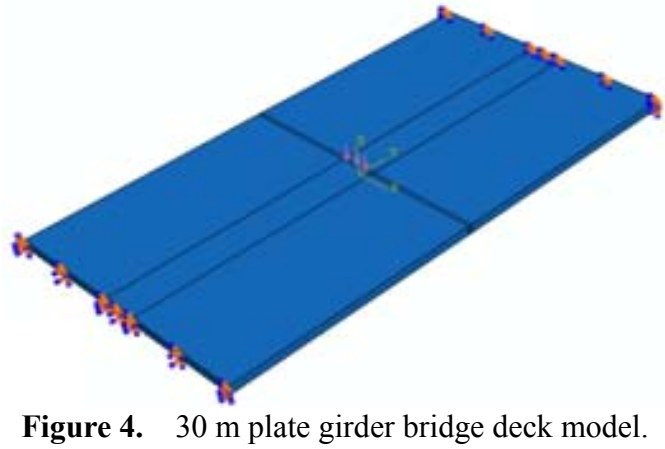


Figure 4. 30 m plate girder bridge deck model.

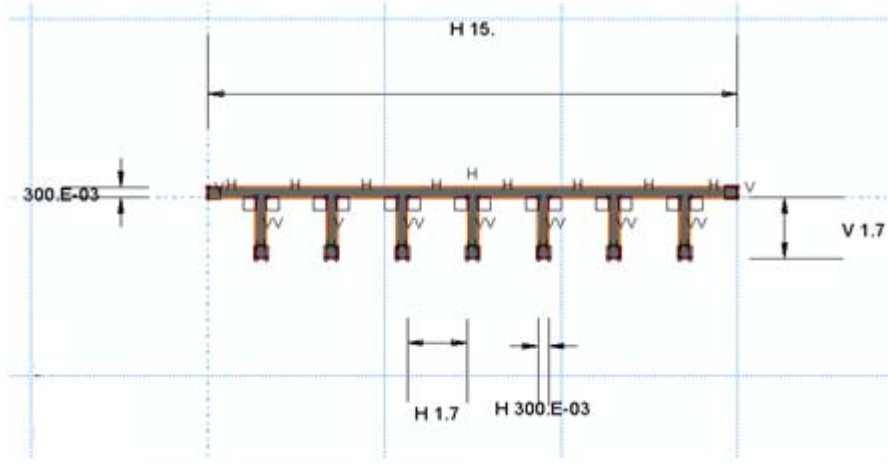


Figure 5. The section size of T grider.

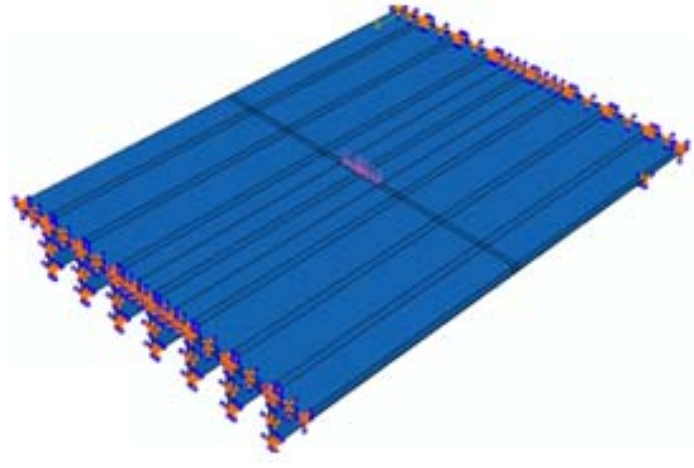


Figure 6. 20 m T girder bridge mode.

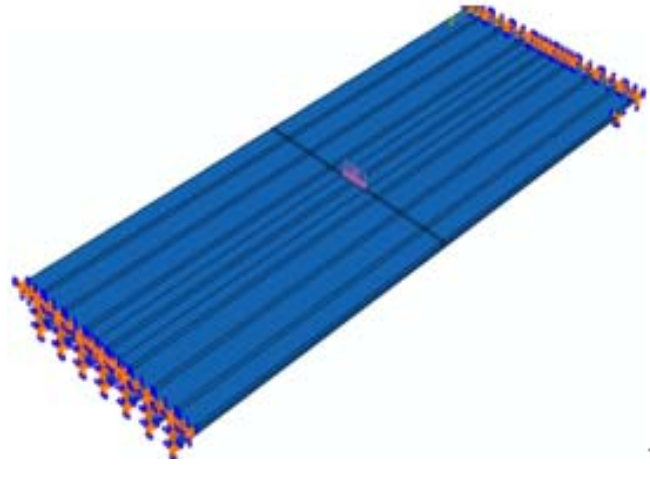


Figure 7. 40 m T girder bridge model.

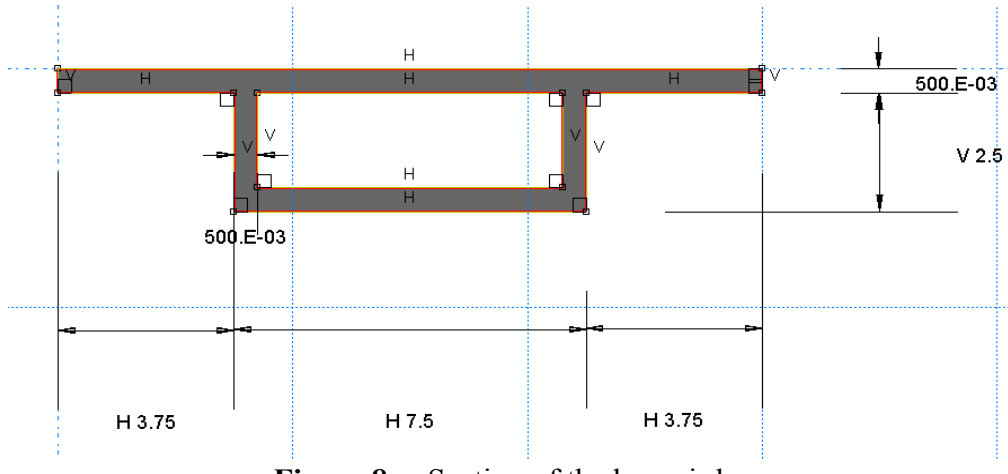


Figure 8. Section of the box-girder.

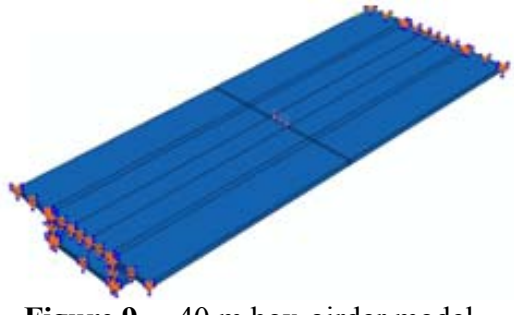


Figure 9. 40 m box-girder model.



Figure 10. 80 m box-girder model.

Box-girder Bridge. Assumed the box-girder had a width of 15 m, a height of 2.5 m, and a width of 7.5 m. The thickness of both the bridge deck and the box was 0.5 m. The sectional form of the box-girder was shown in Figure 8. Assumed the span was 40 m and 80 m, the model of 40 m span was shown in Figure 9, and the model of 80 m span was shown in Figure 10.

Damping Analysis of Typical Bridge

Damping is a basic parameter of structure dynamic analysis, which reflects the energy dissipation characteristic of structure in the structural system vibration process, and damping has a great impact on the accuracy of dynamic analysis results (Gao et al. 2009). For the vibration analysis of bridge structures, the main consideration is the upper damping of the structure. Damping analysis methods usually have two categories, including viscous dampers and hysteretic damping. This paper analyzed viscous damper. α and β are Rayleigh damping constants, which are in proportion with the quality and the stiffness, respectively. α damping relates to quality, which is inversely proportional to the frequency and mainly affects the low-order vibration mode. While β damping relates to stiffness, has a linear relation with frequency, and mainly affects higher-order vibration mode. It has almost no impact on low-frequency. Damping

Table 1. Different span damping of plate girder, T girder, box girder.

Damping coefficient	plate girder 20 m	plate girder 30 m	T girder 20 m	T girder 40 m	box girder 40 m	box girder 80 m
α	0.300257	0.146293	0.736213	0.21949	0.386827	0.143262
β	0.008143	0.015924	0.003396	0.01139	0.006242	0.014184

matrix [C] was calculated with α multiplying the mass matrix [M] and β multiplying the stiffness matrix [K]. Damping matrix [C] as well as α and β were solved by the following equations (Fu et al. 2000):

$$[C] = \alpha [M] + \beta [K] \tag{3}$$

$$\alpha = \frac{2\omega_i \omega_j \left(\omega_j \xi_i - \omega_i \xi_j\right)}{\omega_j^2 - \omega_i^2} \tag{4}$$

$$\beta = \frac{2\left(\omega_j \xi_j - \omega_i \xi_i\right)}{\omega_j^2 - \omega_i^2} \tag{5}$$

where ω_i, ω_j were two-stage natural frequency randomly. ξ_i, ξ_j were damping ratio corresponding to the given vibration mode. i, j were the order corresponding to the vibration mode.

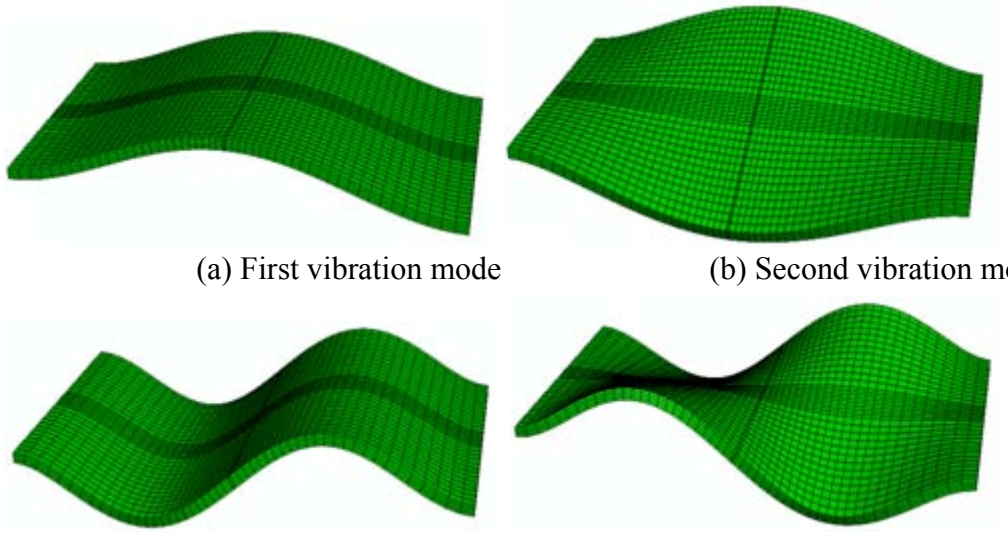
The first two order natural frequency were achieved through modal analysis, and then were taken into the two formula above to calculate α damping and β damping corresponding to bridge of different span and type separately. The results were shown in Table 1.

From Table 1 we can see that the structures are small damping structures because the results of the damping coefficient are less than 1. The effect of the damping to the accuracy of dynamic analysis results can be neglected.

Modal Analysis of Typical Bridge

Modal analysis is an important branch of vibration engineering, which is based on vibration theory, mainly solves the model natural frequency of the structure, and is the basis of transient analysis and structural damping analysis.

According to the vibration principle, the vibration mode of objects becomes complex from lower frequency to high frequency. The modes corresponding to low-frequency are shown by the vibration with overall and larger amplitude. When frequency increases, the vibration mode becomes complex, but mainly for the local and decentralized vibration forms, and the overall rate declines (Binti et al. 2005). The external vibration affects the low-frequency vibration mode of objects mainly (Wang et al. 2004). The fourth-order vibration mode of three kinds of typical bridges is analyzed as follows. The generic vibration mode of plate girder bridge is shown in Figure 11.

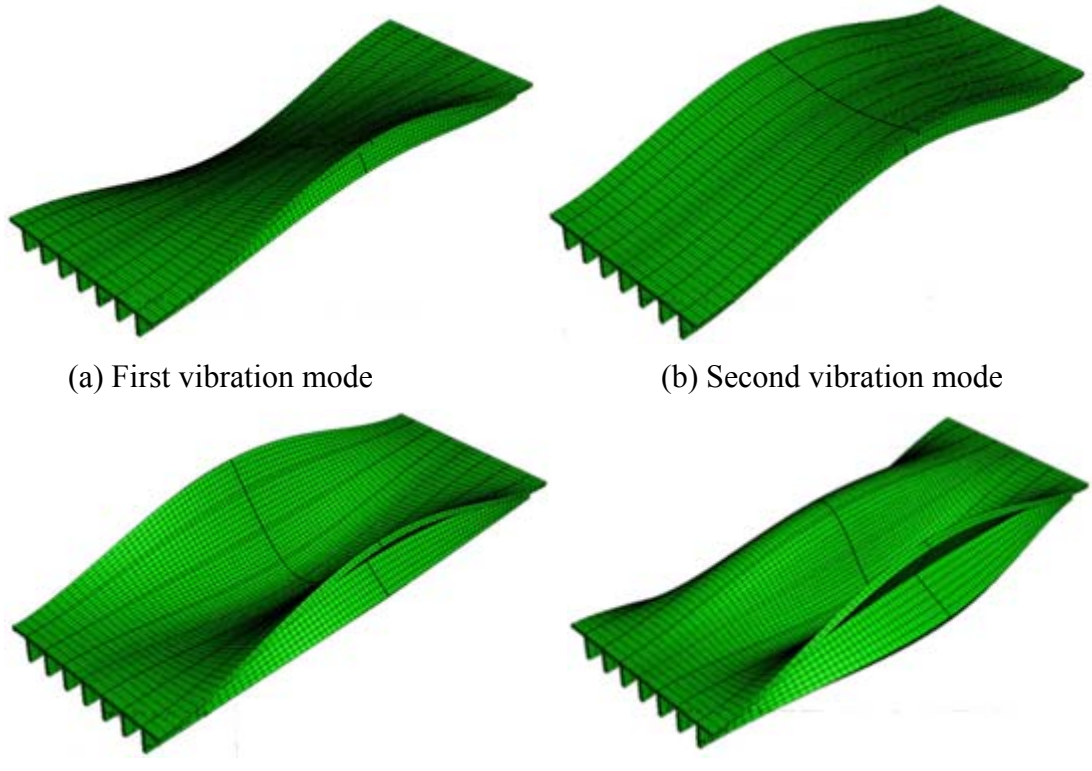


(a) First vibration mode (b) Second vibration mode

(c)Third vibration mode (d) Forth vibration mode

Figure 11. The vibration mode of Plate girder bridge.

The generic vibration mode of T girder bridge is shown in Figure 12.

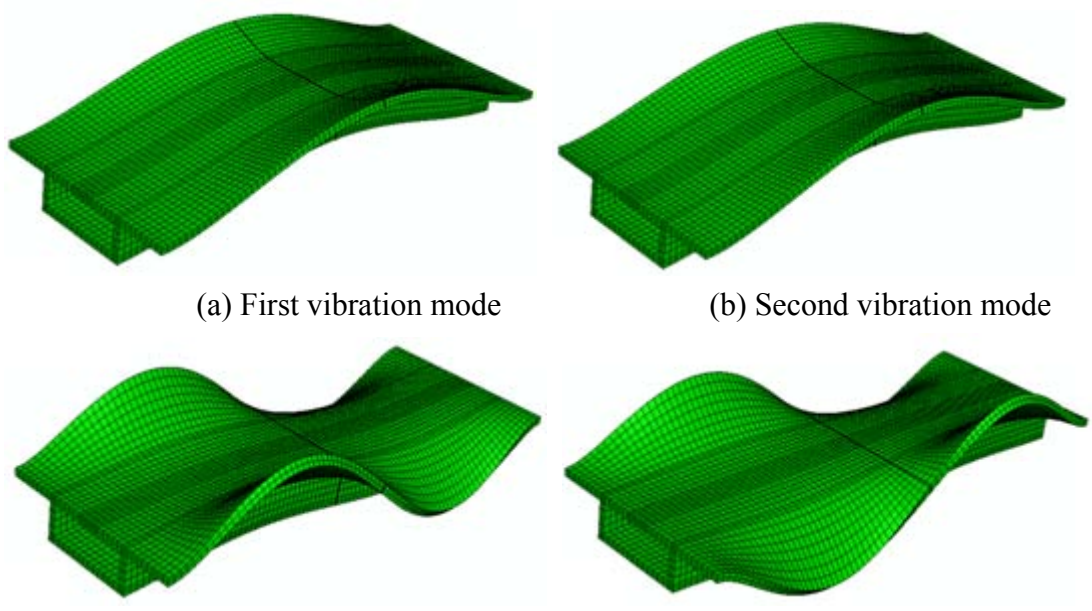


(a) First vibration mode (b) Second vibration mode

(c)Third vibration mode (d) Forth vibration mode

Figure 12.. The vibration mode of T girder bridge.

The generic vibration mode of box girder bridge is shown in Figure 13.



(a) First vibration mode (b) Second vibration mode

(c)Third vibration mode (d)Forth vibration mode

Figure 13. The vibration mode of box girder bridge.

With the assumption parameters mentioned above, the paper calculated the frequency corresponding to the pre-fourth-order vibration modal of different spans and different bridge types. The results were shown in Table 2. From the results, we can see the fundamental frequency of the first-order vibration mode of the bridge deck is less than 7 Hz, while a T girder bridge produces a frequency more than 15 Hz because of shell buckling. The frequency of vertical vibratory roller is 30 Hz, which is within the No. 30 order frequency of the common bridge.

Table 2. Vibration mode frequency of plate girder, T girder, box girder with different span (Hz).

Frequency	plate girder 20 m	plate girder 30 m	T girder 20 m	T girder 40 m	box girder 40 m	box girder 80 m
First vibration mode	5.23	2.32	14.625	4.36	6.53	2
Second vibration mode	7.05	3.96	14.83	4.42	9.49	5.05
Third vibration mode	8.12	6.41	16.67	5.86	12.533	5.46
Forth vibration mode	11.27	11.59	17.72	9.67	13.867	6.18

According to vibration theory, as the bridge frequency of low-order vibration mode has great difference with the universal frequency of vibratory roller, the probability of bridge resonance caused by the roller vibration is very small.

The low-order vibration mode frequency of a T girder bridge is higher compared to the other two bridge types, so the probability of local resonance occurring is larger than plate girder and box girder bridge. As to span, the shorter the bridge span is, the larger the probability of resonance occurring under vibration compaction.

At present, box girder bridges are usually used for large-span bridges, and T girder bridges are seldom selected. So, for large span box girder bridge deck pavement, vibratory compaction can be used. However, if the bridge structure is T girder with a small span, try not to use the vibratory compaction on bridge deck pavement.

Conclusion

(1) By viscous damping results analysis, α damping and β damping for different span and different types of bridges are less than 1. The effect of the damping to the accuracy of dynamic analysis results can be neglected.

(2) The fundamental frequency of the first-order vibration mode of the bridge deck was less than 7 Hz. The frequency of vertical vibratory roller is 30 Hz, which is within the No. 30 order frequency of the common bridge.

The low-order vibration mode frequency of T girder bridge is higher compared with the other two bridge types, so the probability of local resonance occurring is larger than the plate girder and box girder bridge. As to span, the shorter the bridge span is, the larger the probability of resonance occurring under vibration compaction.

(3) For large span box girder and plate girder bridge deck pavement, vibratory compaction can be used. However, if the bridge structure is a T girder bridge with a small span, try not to use the vibratory compaction on bridge deck pavement.

References

Zhao, M., Yu, X., and Li, Y.-T (2008). *Cement Concrete Bridge Deck Asphalt Pavement Defects Prevention and Treatment Technology Research*, Jiangsu Provincial Communication's Department Project.

Xie, G.-M. (2007). *Mechanics of Engineering Vibration*, National Defense Industry Press.

People's Republic of China industry standards. (2004). "Asphalt pavement construction technical specifications." (JTG F40-2004). *The People Traffic Press*.

Toshiyuki, O. Sumio, G.-N. and Shuichi, M. (1985). "Analysis of coupling vibration of a bridge and moving vehicle system by integral equation." *Memoirs of the Kitami Institute of Technology,* 17.

Gao, F., and Jin, J. (2009). "Research on structural damage based on modal analysis." *Journal of Southwest Jiaotong University*.

Fu, Z.-F., and Hua, H.-X. (2000). *Modal Analysis Theory and Application*, Shanghai Jiao Tong University Press. 2000.

Binti, N. and Bhkari, M. (2005). *Modal Analysis of Concrete Bridge Decks Subjected to Free Vibration*, Universiti Teknologi Malaysia.

Wang, H., Qiao, J. (2004). "Finite element analysis and experimental study on dynamic characteristics of bridge structures." *Journal of Highway and Transportation Research and Development*.

Creep Tests and Viscoelastic Analysis of Rubber Asphalt Concrete

Yuliang Chen[1]

[1] Lecturer, School of Engineering & Architecture, East China Jiaotong University, Nanchang, Jiangxi, China, 330013. Phone: (86731) 230-9023; Email: chenyuliangjx@yahoo.com.cn

Abstract

The paper uses rubber asphalt concretes trabecular, carries on the bending creep experiment on the MTS810 material testing machine, and the viscoelastic analysis. The bending creep experiment and the viscoelastic analysis indicated that the rubber asphalt concretes' creep compliance is lower, namely under the unit load condition the rubber asphalt concretes' distortion is small. Along with stress level's increase, the creep strain speed growth is slower, and the rubber asphalt concrete has a good resistance to permanent distortion ability.
Key Words: road engineering; rubber asphalt concrete; creep; viscoelastic; stress level; creep compliance

Introduction

With many years' successful application both in China and abroad, rubber asphalt concretes, as a type of pavement material used in road engineering is obvious in environmental protection and public welfare function. Itnot only can solve the social problem caused by waste tires, it can also reduce the pavement thickness, prolong service life of the pavement, delay the reflection crack, reduce the driving noise, and be fine winter flexible. It is one of the effective ways to solve heavy load transportation and early damage problems. As far as highway construction is concerned, using rubber asphalt concrete can save the construction investment. In China, we construct the high quality bituminous pavement with limited financial and physical resources (Peng and Lu 2001). Therefore, rubber asphalt concrete is the best prospect for highway application.

Currently, researchers mainly concentrate on two aspects: 1) the rubber and the asphalt concrete action mechanism, and 2) the rubber asphalt concrete pavement

performance (Zhou et al. 2007). The research mostly emphasizes the macroscopic angle and uses tests as the main reason to research rubber asphalt concretes' pavement performance. Correlation theory analyses are few, and research is needed about rubber asphalt concretes with high temperature creep and the viscoelastic aspect. Therefore, the paper uses rubber asphalt concretes trabecular, carries on the bending creep experiment on the MTS810 material testing machine, and the viscoelastic analysis.

Test Materials and Method

Test Materials. The aggregate in test are basalt of some material field in Henan province, and the technical indexes are shown in Table 1. The mineral powder in the test is the mineral powder of limestone in Liuyang city of the Hunan province, and the technical indexes are shown in Table 2. The binder is a rubber-modified asphalt provided by afactory in the Henan province, and the technical indexes are shown in Table 3.

Test Method. The commonly used creep test method has the single axle compressive creep, the splitting creep, the bending creep, and so on. The single axle compression and the splitting creep test method influencing factors are many, finally the changeability is big. The bending creep experiment loads are small, the experiment

Table 1. Technical indexes of aggregate.

Test items		Test results	Secification requirement
1	Crushing value(%)	13.3	≤26
2	Abrasion value(%)	19.1	≤28
3	Apparent relative density(g/cm^3)	≥2.744	≥2.60
4	Water absorption(%)	≤1.31	≤2.0
5	Adhesion with asphalt(grade)	5	≥4
6	Sturdiness(%)	≤3.3	≤12
7	Content of fine, long, and flat(%)	≤5.32	≤12
8	<0.075 mm grain content(%)	0.6	≤1
9	Burnish value（PSV）	47.0	≥42
10	Impact value(%)	15.6	≤28
11	Content of soft stone(%)	≤2.88	≤3

Table 2. Technical indexes of mineral powder.

Test items			Test results	Secification requirement	Remarks
1	Apparent density(g/cm^3)		2.732	≥2.50	T0352
2	Water content(%)		0.232	≤1	T0103
3	Size range(%)	<0.6 mm	100	100	T0351
		<0.15 mm	97.2	90-100	
		<0.075 mm	77.8	75-100	
4	Hydrophilic coefficient		0.600	<1	T0353

Table 3. Technical indexes of modified asphalt.

Test items			Test results	Secification requirement
1	Penetration(100 g, 25 °C, 5s) (0.1mm)		76.3	60-80
2	Penetration index	(%)	0.17	≥-0.6
3	Ductility (5 °C, 5cm/min)	(cm)	37	≥40
4	Softening point	(°C)	72	≥50
5	Kinematic viscosity135 °C	(Pa.s)	1.206	≤3.0
6	Flash point(118)(COC)	(°C)	314	≥230
7	Solubility(trichloroethylene)	(%)	99.7	≥99
8	Elastic recovery(25 °C)	(%)	94	≥75
9	Rotation film heating test 163℃	Mass loss (%)	0.0330	±1
		Penetration ratio (%)	75	≥65
		Ductility 5 °C (cm)	23	≥15

slightly simple and so on, recommends for the present experimental regulations and American SHRP. Therefore, this paper uses the bending creep testing method.

The test sample forming method uses the wheel roller method, then cuts 300×30×35 mm trabecular. The test sample's effective span is 200 mm, and the test carries on the bending creep experiment on the MTS810 multi-purpose material testing system with a temperature of 20℃. Because the bending creep test result has been a big influence by the stress level, stress level choice too large or too small causes the distortion development quickly or excessively slow, the test sample either breaks very quickly or develops slowly and is unable to spread out, making data analysis difficult. Therefore, before the experiment, the researcher must carry on the disposable load destructive test to the different asphalt mixture. Using 0.1, 0.2, 0.3, and 0.44 stress ratios of the bending failure load carries on the comparative trial, observes the hour distortion rate, when the stabilization period (total creep curve for linear stage) surpasses 1h, use the strength level to carry on the bending creep experiment. This test result indicated when the stress ratio chooses 0.2, the test sample distortion is slow. When the stabilization period surpasses 1h, the test sample does not break, therefore, the creep test uses 0.2 stress ratio to loading.

Test Result Analysis

When the stress ratio is 0.2, the experimental time and mid-span tensile strain curve is shown in Figure 1.

As shown by Figure 1, 20℃, under 0. 2 stress ratios the rubber asphalt concretes' bending creep curve, is the same with the plain asphalt concretes bending creep, its creep curve may also divide into the migration time, the stabilization period and the breakdown time. The migration time's creep growth, the stabilization period strain speed maintains stable, belongs to the linear deformation stage, and breakdown time

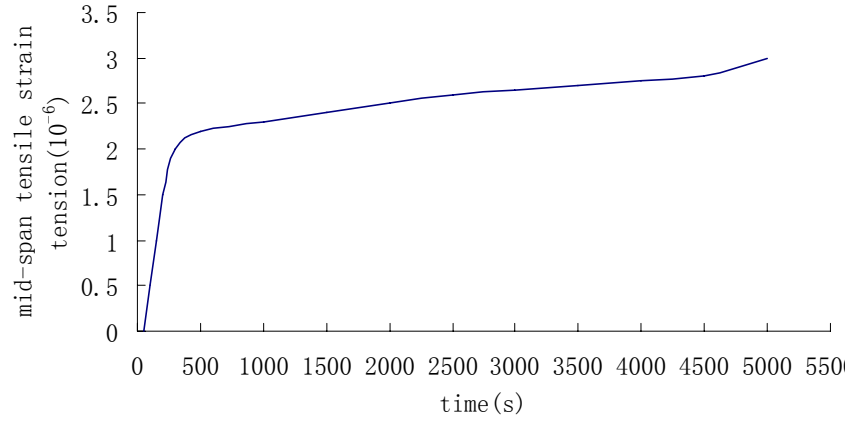


Figure 1. 20 °C, under 0.2 stress ratio, time and mid-span tensile strain tension strain curve.

total creep sharp growth. Through the creep test result analysis, may obtain the material creep rate, establishes the rheological model, obtains the rheological parameter, evaulates high temperature rheological performance of material.

Creep Compliance. The creep compliance expresses distortion under the unit stress. The creep compliance is bigger, and the distortion under the unit load condition is bigger. Min et al. (2004) proposed may use the creep compliance of 1000s (J1000) as design parameter of pavement. According to the standard [Standard test methods of bitumen and bituminous mixtures for highway engineering (JTJ 052-2000), China], the curving creep compliance equation is as shown in Eq. 1., and can obtain the creep compliance when asphalt concrete 1000s, as shown Table 4. From Table 4 we can see the rubber asphalt concretes' creep compliance under 0.2 stress ratio loading, and the 1000s are 13.3% the plain asphalt concretes' creep compliance. Therefore, the rubber asphalt concretes creep compliance compared to plain asphalt concretes is lower, which indicated that under the unit loading, the rubber asphalt concretes distortion is smaller than the plain asphalt concretes by far.

$$J(t) = \frac{4bh^3}{L^3 F_0} d(t) \times 10^6 \tag{1}$$

Creep Rate. The migration time, the stabilization period creep strain $\varepsilon(t)$, is mainly decided by the creep strain rate ε_{speed} and load time t (as shown in Eq. 2). We can use the strain rate of the creep curve stabilization period to charaterize material creep characteristic, because the creep strain rate is related to distortion of the stabilization period stage, and directly reflects the ability of the material resistance distortion under repeat loading.

Table 4. Creep compliance of asphalt mixture.

Type	Creep compliance(1/MPa)
Plain asphalt concrete	0.060
Rubber asphalt concrete	0.009

Table 5. Low temperature bending creep rate of asphalt concrete.

Type	Average creep rate (1/s)
Plain asphalt concretes	5.27×10^{-7}
Rubber asphalt concretes	3.15×10^{-7}

$$\varepsilon(t) = \varepsilon_e + \frac{1}{\beta}\varepsilon_{speed}t + \alpha(1 - e^{-\varepsilon_{speed}t}) \quad (2)$$

where: $\varepsilon(t)$ is creep strain; ε_e is immediate acknowledgment elastic deformation; t is time; β，α is the experiment constant.

Low temperature bending creep test results of the plain asphalt concretes and the rubber asphalt concretes is shown in Table 5.

From Table 5 we can see, under the same stress level, the rubber asphalt concretes' creep strain rate is the plain asphalt concretes 59.7%, which indicates the rubber asphalt concretes' resistance distortion ability is stronger than plain asphalt concretes' resistance distortion ability under repeat loading.

Viscoelastic Analysis. The common viscoelasticity model includes: Maxwall model; Kelven model; Ven der poel model; Burgers model, and so on. In these four models (Lougheed and Papaglannakis 1996; Nolank 1995), the Burgers model can well reflect the distortion characteristic of asphalt concrete, manifesting especially in the load process. But the Burgers model also has a flaw: The asphalt concrete permanent deformation is attributed for time linear function, but in fact the viscosity increase is along with the load time s extension, thus permanent deformation s increase follows along with the decrease in time. Therefore this model cannot well reflect distortion characteristic of the asphalt concrete in the entire load, the unloading process.

This paper uses Tongji University to propose "the four unit five parameters" model (Xu 1992), the four unit five parameters" model considered the asphalt concrete has "the consolidation effect," namely its viscosity increase is along with the load time s extension "The four unit five parameters" model has carried on the following revision to the Burgers model: Exterior dashpot of the model is expanded to one generalized dashpot, which is paralleled by many different viscosity dashpots, and along with the load time s extension, the numbers of parallel dashpots increase according to a certain rule. Therefore, this dashpot s viscosity increases along with the load time, then makse up the Burgers model flaw effectively.

According to strain value obtained from the experiment, we can back calculate viscoelasticity parameters E_0, E_1, A, B, and η_1 of asphalt concrete through the iterative method (as shown in Table 6).

From Table 6 we can see, the rubber asphalt concretes relaxation time will be

Table 6. Asphalt concrete viscoelastic parameters(20℃).

Type	E_0 (MPa)	E_1 (MPa)	A (10^5MPa)	B (10^{-4}MPa)	η_1 (10^{-4}MPa.s)
Plain asphalt concretes	601	194	2.05	1.7	3.8
Rubber asphalt concretes	1200	776	8.9	4.7	7.08

obviously longer than the plain asphalt concretes relaxation time, which indicated that the rubber asphalt concretes have the good resistance to permanent deformation ability.

Conclusions

The paper uses rubber asphalt concretes trabecular and carries on the bending creep experiment and the viscoelastic analysis. The conclusions are given as follows:
(1) Compared with plain asphalt concretes, the rubber asphalt concretes distortion is small under the unit load condition.
(2) The resistance to deformation ability of rubber asphalt concretes is stronger than that of plain asphalt concretes under repeated loading.
(3) Compared to the plain asphalt concretes, the rubber asphalt concretes have the better resistance to permanent deformation ability.

References

Bahia, H. U., Anderson, D. A., and Christensen, D. W. (1992). "The bending beamrheometer; A simple device for measuring low-temperature rheology ofasphalt binders." *AAPT*, 61,117-153.

Chehab, G. (2002). "Characterization of asphalt concrete in tension using a viscoelastoplastic model." *Ph.D. dissertation*, North Carolina State University, Raleigh, NC.

Chehab, G. R., Kim, Y. R., Schapery, R. A., Witczak, M. W., and Bonaquist, R. (2003), "Characterization of asphalt concrete in uniaxial tension using a viscoelastoplastic model." *Paper presented at the Association of Asphalt Paving Technologists Annual Meeting.*

Gibson, N. H., Schwartz, C. W., Schapery, R. A., and Witczak, M. W. (2003), "Viscoelastic, viscoplastic, and damage modeling of asphalt concrete in unconfined compression." *Paper presented at the Transportation Research Board Annual Meeting.*

JTJ 052-2000 (2000). "Standard test methods of bitumen and bituminous mixtures for highway engineering." *Beijing: People's Republic of China Ministry of transportation.*

Kaloush, K. (2001). "Simple performance test for permanent deformation of asphalt mixtures." *Ph.D. dissertation*, Arizona State University, Tempe, AZ.

Lin, X. F., and Wu, Q. I. (1992). "Rubber modified road asphalt and microscopic structure." *Journal of Synthetic Rubber Industry*, 23(3),166-169.

Lougheed, T. J., and Papaglannakis, A. (1996). "Viscosity characteristics of rubber-modified asphalts." *Journal of Materials in Civil Engineering*, 8,153-156.

Motola, Y. (2003). "A study of the effect of shear stress reversal on the behavior of asphalt concrete." *M.Sc Thesis (in Hebrew)*, Technion, Haifa, Israel.

Mun, S., Chehab, G. R., and Kim, Y. R. (2007). "Determination of time-domain viscoelastic functions using optimized interconversion techniques." *Road Materials and Pavement Design*. Lavoisier, In Print.

Nolank, L. (1995). "Low temperature nature of PE modifier binder and asphalt concrete mix." *AAPT*, 1995, 4(5),64-67.

Peng, H., and Lu, W. M. (2001). "Research on performance and technology of the rubber powder modified asphalt mixture." *China Journal of Highway and Transport*, 14(12),4-7.

Uzan, J. (1996) "Asphalt concrete characterization for pavement performance prediction." *Assoc. Asphalt Pavement Technol*., 65,573-607.

Xu, S.-F.(1992). "A rheological model representing the deformation behavior of asphalt mixtures." *Mechanics and Engineering*, 14(1),37-40.

Zhou, Z. G., Tan, J., and Li, X. L. (2007). "Fiber modification rubber asphalt high temperature stability test research." *Journal of China and Foreign Highway*, 12(06), 171~173.

Shrinkage Performance of Cement-treated Macadam Base Materials

Yan Wang[1], Xiang Ma[2] and Zhilin Sun[3]

[1] Instructor, PhD, Faculty of Built Environment and Engineering, Jinling Institute of Technology, 99 Hongjing Road, Nanjing, Jiangsu province, China, 211169; Email: wangyan@jit.edu.cn
[2] Ph.D., College of Transportation Engineering, South-east University, 2#, Sipailou, Nanjing, Jiangsu province, China, 210018; Email: mx_fst@hotmail.com
[3] Ph.D., School of Traffic and Transportation Engineering, Changsha University of Science & Technology, 960#, 2nd south-wanjiali road, Changsha, Hunan province, China, 410004, Email: sunzhilin1979@163.com

Abstract

Shrinkage performance is always considered as the most important characteristic of cement-treated macadam base. By using different gradation types and different levels of cement stabilized macadam mixture, the paper conducts indoor performance tests to study the mixture's shrinkage performance. Based on the tests, estimation formulas on shrinkage coefficients of the cement stabilized macadam mixtures are then elaborated. Results of the test indicated that mixes possess the lowest dry-shrinkage when 3%~4% cement dosage are used, while temperature-shrinkage increase with cement dosages are added. Otherwise, more close-grained mixes with that have a better dry-shrinkage performance, while the temperature-shrinkage performance of these kinds of materials is worse than the sketch-grained mixes. Further analysis and test results show that shrinkage stress in base rises as cement dosage increased; a smaller cement dosage of base materials, as long as strength meets the requirement to reduce shrinkage crack in base, is feasible. Large numbers of project experience indicate that dry-shrinkage mainly occurs within 20 days when base construction complete. Moreover, compared to the continually cold winter, cracks easily form, when the temperature changes suddenly even in warm weather. Results of an indoor test show the same speciality.

Key Words: highway engineering; cement-treated macadam materials; shrinkage performance; estimate model

Introduction

The total highway mileage in China increased to 63 thousand kilometres by the end of 2008. More than 90% of pavements are semi-rigid asphalt pavements. As a major type of semi-rigid base structure, the cement stabilized macadam base is most often than not deployed. However, the semi-rigid base is prone to crack and dehisce due to temperature gradients and humidity changes, and with the intermittent impacts of vehicle load and changes in temperature and humidity, the primary ruptures in the base tend to extend and expand up to the asphalt surface, resulting in serious reflection cracks. Conversely, the asphalt surface is apt to crack under the combined impacts of temperature changes and traffic load, and the surface cracks will then gradually develop downward to the base structure through longtime or heavy use (Guthrie, 2008). Engineering practice shows that cracks in the structure of cement stabilized macadam mixture as base and asphalt pavement as surface have become a major drawback for the pavement structure (Sheng, 2007).

Furthermore, with rain and water easily flowing into the cracked ruptures, the pavement structure, which is under the continued impact of vehicle loads, will inevitably have erosions, net-shaped cracks, and other forms of damages. Therefore, the shrinkage rupture in the cement stabilized macadam base is the major factor that affects the overall usable performance and service life of pavements (Smith, 2006). One of the effective approaches to reducing the quantity of cracks and even impeding the emergence of cracks is to improve the contractility performance of the material (Li 2006). In view of the lack of systematic achievements on the contractility of the cement stabilized macadam in the existing researches, which has partly resulted from the fact that there are great uncertainties and variabilities when molding and testing cement stabilized macadam mixture (Wang 2008), the paper explores and clarifies laws of changing contractility of the cement stabilized macadam under different conditions of humidity and temperature.

Test Analysis

The Raw Material's Characteristics and Gradation Designing. Indoor tests will select from 2.5%–5.0% cement dosage (marked as "value"), different grade with CTB mix material specified in this regulation facilitates with shrinkage test, including fine gradation, namely the median gradation of the standard (marked as "z"), the rough gradation of the standard gradation range (marked as "c") and the skeleton gradation determined by the compacting test (marked as "g") (JTJ E42-2005). The gradations are shown in Table 1.

Table 1. Composition of different aggregate gradations.

Gradations	Sizes of Sieve Pores (mm)							
	31.5	26.5	19	9.5	4.75	2.36	0.6	0.075
z	100.0	94.7	82.6	54.3	42.0	27.2	13.9	2.5
c	100.0	92.5	75.2	43.1	30.6	14.8	7.5	1.4
g	100.0	-	47.6	36.0	26.4	16.2	8.3	1.5

The cement employed is the 32.5# composite silicate cement, the fineness degree of which is 3.2%, primary solidifying time of which is 5.3 h while the final solidifying time is 7.4 h. The aggregate consists of No. 1, No. 2, No. 3, and No. 4. The crushing value of the aggregate is examined to be 22.8%; the liquid limit and plasticity index should also be examined if the grain diameter is below 0.6 mm. The rest of the capability data of the cement and aggregate all meet the criterion.

Dry-shrinkage Performance Test. The dry-shrinkage experiment is performed at the temperature of 25℃ and with a humidity of 50%. The beam specimen chosen for the experiment measures are about 100 mm*100 mm*400 mm (Si, 2007). Strain gauges are pasted on both sides of the specimen to test the straining change. The strain data are collected by computers while the water content change of the specimen is recorded manually. The test time interval is once every 6 h on the first day, once every 12 h from the second to the fifth day, once every 24 h thereafter until the water content basically stops changing. Based on the change value of dry shrinkage strain and water content, dry-shrinkage coefficient $\overline{\alpha_i}$ was calculated by Eq. 1 as follows:

$$\overline{\alpha_i} = \frac{\varepsilon_i - \varepsilon_0}{\omega_0 - \omega_i} \tag{1}$$

where, ε_i ε_0 is dry-shrinkage strain of initial and ith time, με; ω_i, and ω_0 is water content of initial and *i*th time, %.

Influence on Dry-shrinkage Characteristics Caused by Water Loss Rate and Exposition Time. The water loss rate is obtained by dividing the water loss by the dry weight of the specimen and is used to evaluate the dry-shrinkage characteristic of different mixtures under certain conditions of water loss. Test log reveals that water loss rate of the first three days account for 60% of the total amount. This conclusion coincides with the opinion in the literature. Water loss rate will slow down since that time. Figures 1 and 2 show the rules that the average dry-shrinkage coefficient changes with the water loss rate and the exposition time.

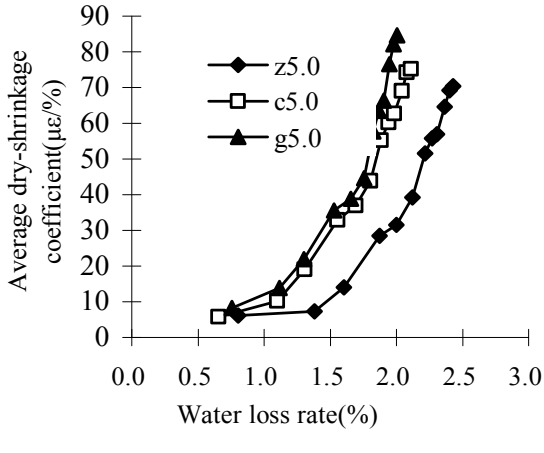


Figure 1. Changes of dry-shrinkage coefficient with water loss rate.

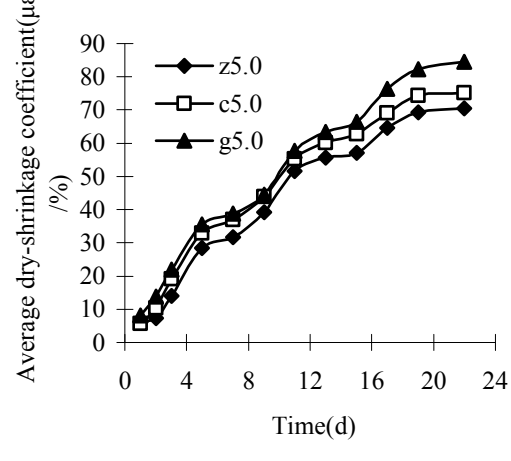


Figure 2. Changes of dry-shrinkage coefficient with exposition time.

We can see from the figure, with the water loss rate and exposition time increased, the "$\overline{\alpha_i}$" for different mixtures is all augment synchronous. However, an increasing slope of $\overline{\alpha_i}$ value becomes steeper beyond 2.0% in Figure 2; while the increased amplitude of average dry-shrinkage coefficients keep in a steady condition with exposition time changed. Combined with the rules expressed in Figures 2 and 3 with the conclusion of the water loss rate law,, it is easy to find out that to do a better job maintaining during first several days when construction completed, so that the surface and inside water in the cement-treated base will be contained as large as possible. It will be an effective way to decrease shrinkage in base, consequently, shrinkage crack in base and reflecting crack in asphalt pavement will be controlled.

Influence on Dry-shrinkage Characteristics Caused by Cement Dosage and Gradation. As we know that both the cement dosage and gradation are the key factors for every capability of cement-treated base materials. Test results of the mixture's dry-shrinkage performance with different gradations and cement dosage is shown in Figure 4.

Viewing from Figure 3, along with an increase in cement dosage, the dry shrinkage coefficient for different mixture has one min value next to 3% cement dosage. On the other hand, influence of gradation is quite different when the cement dosage varies. In other words, higher cement dosages correspond to smaller shrinkage distinguished for materials with different gradation, while less cement adopted in mixture, shrinkage performance will show a large difference. Anyway, the mixtures with fine gradation possess better shrinkage performance relative to other two type materials. Here is one thing worth mentioning that because the strength and some other capacity of mixture will perform better with higher cement usage, the contractors always adopt a high cement dosage of up to 6% in practice. From the test results we come to a conclusion that the fine gradation mixture with 3~4% cement dosage has an optimal dry-shrinkage performance.

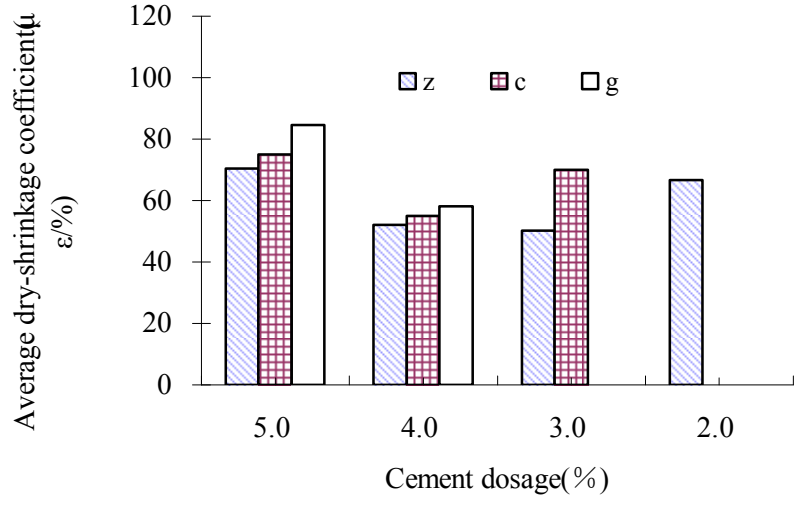


Figure 3. Changes of dry-shrinkage coefficient with cement dosage and gradation.

Table 2. Calculated temperature-shrinkage coefficient.

Types of mixture	Temperature (ºC)							
	50	40	30	20	10	0	-10	-20
z5.0	10.5	11.2	8.7	7.9	7.3	6.5	4.8	4.6
z4.0	9.9	10.9	8.3	8.2	7.0	6.1	5.1	3.7
z3.0	9.1	10.4	8.3	7.7	6.8	5.8	4.2	3.7
z2.0	8.7	10.3	7.7	7.8	6.6	5.3	4.2	3.0

Temperature-shrinkage Performance Test. The temperature-shrinkage experiment is performed in the high and low temperature conditions box with the temperature change range from 55℃ to -25℃. The beam specimen chosen for the experiment measures are the same with the dry-shrinkage test. The selection of temperature alternation is 10℃, and the change rate is 1℃/min. The temperature will be maintained for 2 hours once it changes to the setting. Also, the strain data are collected by computers, and the temperature change parameter and setting is auto-controlled by programming (Guthrie, 2002).

The curing time of the specimen is 28 days; three of same specimens were prepared for the same type of materials. First of all, the initial temperature is set to 55℃. Place the specimen in the conditions box for 3 h to begin the test. As soon as the temperature changes, strain data for every specimen will be tracked synchronously. Temperature-shrinkage strain for every range is considered as the value collected at the time that constant temperature is about to change. The temperature-shrinkage coefficient β can be calculated by Eq. (2) as follows (Miller, 2000):

$$\beta = \frac{\Delta\varepsilon}{\Delta t} + \beta_s \tag{2}$$

where: Δt is the difference of temperature, ℃; $\Delta\varepsilon$ is the difference of strain corresponding to Δt , με; β_s is linear expansion coefficient of self-temperature-compensated strain gauge.

According to the test results, the temperature-shrinkage coefficient for different mixtures and temperature conditions were calculated, as shown in Table 2. It can be observed that temperature coefficient decreases as the test temperature drops, that is to say, negative linear correlation can find out between the shrinkage coefficient and temperature; on the other hand, when the temperatures are the same, the shrinkage coefficient in various mixtures has no significant difference. With the cement usage being higher, the index has a few increases. The large coefficient occurs in high circumstance temperature, and shrinkage capacity of materials in summer is stronger than in winter. So we can come to the conclusion that temperature changes suddenly in summer and is worse than the sustained low temperature status.

Influence on Temperature-shrinkage Characteristics Caused by Cement Dosage and Gradation. For the certain type mixture, the average shrinkage coefficients were calculated under different temperatures to analyze the influence of cement dosage and gradation. Figure 4 lists the results.

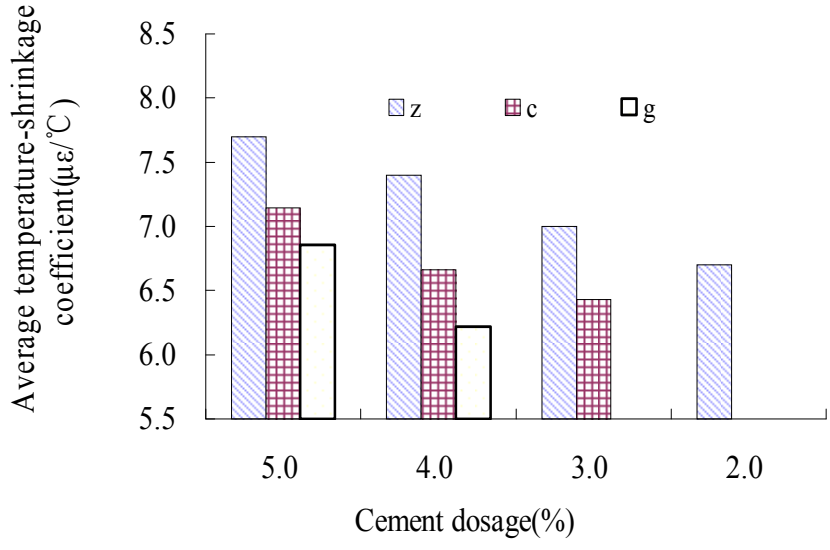


Figure 4. Calculated average temperature-shrinkage coefficient.

According to this image, we can find that the average shrinkage coefficient increased linearly with cement usage added in the mix. Also we know that based on the calculated value, cement dosage ranged from 2.0~5.0%, while the average temperature shrinkage coefficient varies only changes between 6.7~7.7με/℃. That is to say, cement dosage has no significant influence on temperature shrinkage coefficient. However, when the gradation of the mixture is changed, even the cement dosages are the same; its shrinkage characteristic has some difference among the mixture. Specifically speaking, the coefficients decrease successively from the fine, rough gradation to skeleton gradation; the mixture with more coarse aggregate possesses a better shrinkage performance when the temperature changes.

Prediction Model

Influence of factors including water loss rate, exposition time, temperature, cement dosage, and gradation are analyzed based on the indoor test. Several useful conclusions were made based on the results. In order to put the outcome of research into practice or to make it a helpful reference in design and construction, we are attempted to estimate shrinkage performance prediction model of cement-treated base materials based on the test results.

Some correlative documents indicated the curve of continuous gradation can be fitted by an exponential function; this conclusion was adopted to estimate the model. In consideration of the rough and skeleton gradation is two specific types that mix in the lab to analyze the effect of gradation to the shrinkage performance, in other words, these two kinds of mixtures were not widely used in engineering, so only the fine gradation results were used in multiple regressions.

In the research, single factor regression method was used to complete the gradation mode first of all, and then using common analysis software Origin to do the multiple

regressions. Dry-, temperature-shrinkage coefficients were deduced, respectively, as Eqs. (3) and (4).

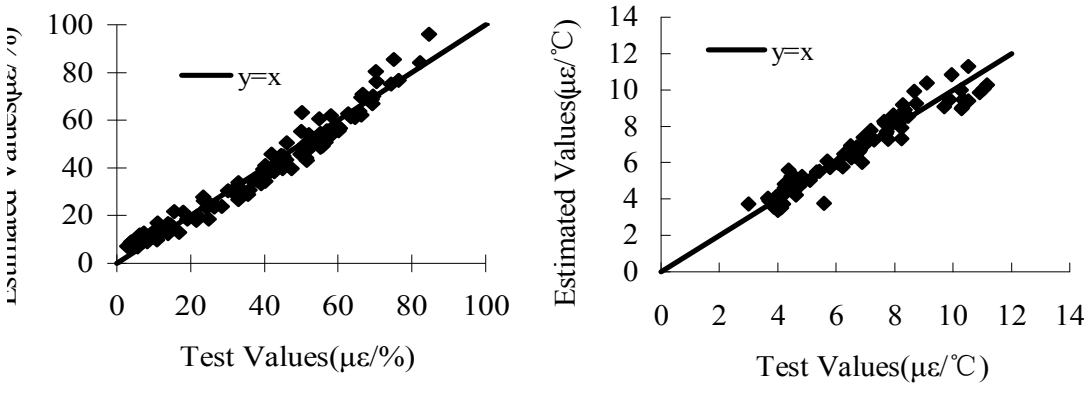
$$y_\alpha = 8.75x_c^2 - 59.97x_c + 151.56 \quad (3)$$

$$y_\beta = 0.33x_c + 6.04 \quad (4)$$

where: y_α, y_β stand for dry- and temperature-shrinkage coefficients, respectively; x_c is cement dosage, %.

To make sure the validity of the model, the estimate value and test results were first drawn in the same chart. Figure 5 shows that estimate values calculated by the equation, and test values are near the line. That is to say, estimate values are close to the test.

Secondly, T-test is adopted to inspect the validity of the model. As t-statistic is employed to test whether a significant difference exits in the hypothesis testing, after the respective bilateral t-tests on the evaluating models paired data on the significance level of 0.05, the result can be seen in Table 3.



(a) Dry-shrinkage (b) Temperature-shrinkage

Figure 5. Comparisons between the estimate and test results.

Table 3. Results of *t*-Test.

	Dry-shrinkage		Temperature-shrinkage	
Mean	37.796	37.659	6.900	6.896
Variance	479.408	471.473	4.844	4.511
Observations	108	108	72	72
Poisson related coefficient	0.981		0.958	
Hypothesized Mean Difference	0		0	
D_f	107		71	
t Stat	0.33		0.060	
P(*T*<=*t*) two-tail	0.74		0.952	
t Critical two-tail	1.982		1.994	

The analyzed results of the T-test show that all the absolute value of t-statistic of the model is lower than the bilateral critical value. It brings forward such a conclusion that the test value and evaluated value don't have any distinct difference; the model is valid and can be adopted to determine the shrinkage performance of different CTB mixtures.

Conclusions

The study has attempted to shed light on shrinkage performance of different cement-treaded base materials. Key conclusions for research are as follows:

- Dry-shrinkage performance of a mix shows best when the cement usage is 3%～4%, that is to say, both higher and lower cement dosage are all disadvantageous for improving the shrinkage performance.
- Rules of temperature-shrinkage performance change with cement dosage is quite differently to dry-shrinkage. Linear increase trend was observed on temperature-shrinkage performance. More cement usage in mix, lager shrinkage deformation logged
- Generally speaking, mixtures with density and coarse gradation have a better shrinkage performance.
- The prediction model established in the research has preferable precision; it can be used for CTB materials design and selection.

References

Guthrie,W.and Reese,G.(2008). "Use of the heavy clegg impact soil tester to assess rutting susceptibility of cement-treated base material under early trafficking*", TRB 2008 Annual Meeting CD-ROM*, No.08-2286.

Guthrie, W. S., Sebesta, and Scullion, T. (2002). "Selecting optimum cement contents for stabilizing aggregate base materials." Report 7-4920-2. Texas Transportation Institute, Texas A&M University System, College Station, TX, February.

Li, Z. (2006). "Research on the performance of cement stabilized aggregate mixture and indexes correlation-ship." *Southeast University*, 34-52.

Miller, G. A., and Zaman, M. (2000). "Field and laboratory evaluation of cement kiln dust as a soil stabilizer." *Journal of the Transportation Research Board, National Research Council*, Washington D.C. , No. 1714, 25-32.

Ministry of Transport-Denmark. (2004). "Mechanistic design of semi-rigid pavements-an incremental approach." *Danish Road Institute*, 5-8.

Sheng, C. (2007). "Performance research and construction control of cement-treated aggregate mixture." *Southeast University*, 31-36.

Si, Z., and Herrera, H. (2007). "Laboratory and field evaluation of base stabilization using cement kiln dust." *Journal of the Transportation Research Board,* National Research Council, Washington D.C.

Smith, B. S. (2006). "Design and construction of pavements in cold regions: state of the practice." *Department of Civil and Environmental Engineering*, Brigham Young University, Provo, UT, December.

Wang, Y. (2008). "Research on the mechanism and design method of cracking in cement-treated base asphalt pavements." *Southeast University*, 3-126.

Vertical Load-carrying Natural Frequency of Railway Continuous Steel Truss Bridges

Jianying Ren[1], Mubiao Su[2] and Qingyuan Zeng[3]

[1] PhD. Graduate student, College of Civil and Architectural Engineering, Central South University of China; Lectuer, Department of Engineering Mechanics, Shijiazhuang Tiedao University of China, Email: aren1977@sina.com
[2] Prof., Structural Health Monitoring and Control Institute, Shijiazhuang Tiedao University of China.
[3] Prof., College of Civil and Architectural Engineering, Central South University of China

Abstract

In this paper, the vertical load-carrying natural frequency of 3×64 m through stud welding railway steel truss bridge is calculated using the vehicle-bridge system model, and the rule with time and some influence factors to the continuous bridge vertical load-carrying natural frequencies are analyzed, when 20 high-speed passenger vehicles with the same parameters are travelling on the bridge. The results show that when the vehicles are distributing on the whole bridge, the bridge vertical load-carrying natural frequency is periodically varied, and the periodic time is $T=L_v/v$. Also, the bridge vertical load-carrying natural frequency is related to the unsprung mass and the stiffness of the unsprung of each wheel-set of the vehicle, the vehicle's length, and the mass of the vehicle. But it is independent of the speed of the train. When the length of the train is shorter than the length of the bridge, the periodic changing stage of the vertical load-carrying natural frequency would never occur, and then the vehicle-bridge system would not incur vertical resonance.
Key Words: continuous steel truss bridge; vehicle-bridge system; vertical load-carrying natural frequency; the vehicle parameters

Introduction

The dynamic response of railway bridges under train loads is one of the fundamental problems to be solved in railway bridge design. In the past two decades, numerous

researchers have studied the train-induced vibration of railway bridges (Wu and Yang 2003; Yang et al. 1997; Yang et al. 2004; Ju and Lin 2003; Xia et al. 2006). The dynamic analysis has been the subject of numerous researches over the years, with an increased interest in the last decade, mainly due to development of new high-speed railway lines and common drive to upgrade the existing infrastructure (Majka and Hartnett 2008). With more and more high-speed railways being built in the world, great efforts have been constantly spent to the subject of dynamic interaction of vehicles and bridges. On the one hand, the train running with high speed induces dynamic impact on the bridges, influencing their working state and service life. On the other hand, the vibration of the bridge in turn affects the running stability and safety of the train, and thus becomes an important factor for evaluating the dynamic parameters in bridge design (Xia and Zhang 2005). The research work on this subject has a long history of more than one hundred years. Especially in the last decades, increasingly sophisticated analytical models have been developed by researchers in China and abroad (Cheli 1989; Mallikarjuna and Savard 1993; Xia 2002).

As a train travels over a bridge at a resonant speed, the response of the bridge tends to increase steadily as there are more loads passing the bridge. This is the so-called "resonance phenomenon" (Yau 2009). After long-term field experiment and the theories research, when vehicles are traveling on the bridge at a high speed, it is known that as vehicle speed increases, the bridge amplitude is larger. When successive vehicles travel on the bridge at velocity of v, the vehicles' (which length is L_v) wheel loads are equal to a dynamical periodical load with fundamental frequency v/L_v act on the beam. When the integer times of the frequency due to weight loads of a series of successive vehicles acting periodically on the bridge agree with the fundamental vertical load-carrying natural frequency, that is $jv/L_v=f_{1y}$ (j=1,2,3…), the sum of free-vibration response components generated by each moving load acting successively on the bridge will result in the resonance (Li and Su 1999). If the vehicle speed is higher, the resonant maximum amplitude is larger, too. Increasing demand on the railway transportation requires some measures to be taken in order to satisfy this, and improvements of the capacity can be achieved either by increasing the speed of the traffic or by increasing the axle loads. However, either of these solutions exerts extra pressure on the existing infrastructure, and the effects of these changes have to be carefully considered.

Therefore, further investigation of the vertical load-carrying frequency is very significant. Up to now, the research of vertical load-carrying frequency on the vehicle-bridge interaction is mainly focused on simply supported bridges, scarcely focused on continuous bridges. Consequently, the main purpose of this paper is to further study the vertical load-carrying frequency of the continuous bridge and the effects of various parameters influencing the vertical load-carrying frequency of the continuous bridge.

Vertical Motion Equation of the Vehicle-Bridge System

The Vehicle-Bridge System Model. Let the first and the last wheel-axle set on the bridge be represented by K and M at time t, and vehicle and bridge interact at the same vertical plane (Figure 1). Each vehicle is considered as a model of a rigid body with four wheel-axle sets, which has two degrees of freedom only in the vertical plane. (Chu et al. 1979). The two degrees of freedom of the vehicle correspond to bounce and pitch motion (Figure 2). The vehicle vertical two series suspension spring is considered as a linear spring, and the stiffness is substituted as an equivalent stiffness (Wang 1983). The mass of the redirector truss is averaged to the rigid body and wheel-axle sets. It is hinged between the vehicles. The vertical displacements of the wheel and the wheel/rail contact point on the rail are always considered uniform**.**

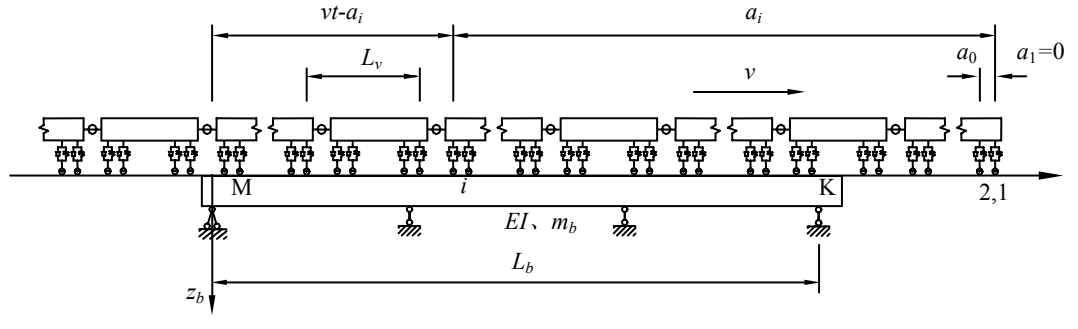


Figure 1. The vehicle-bridge model.

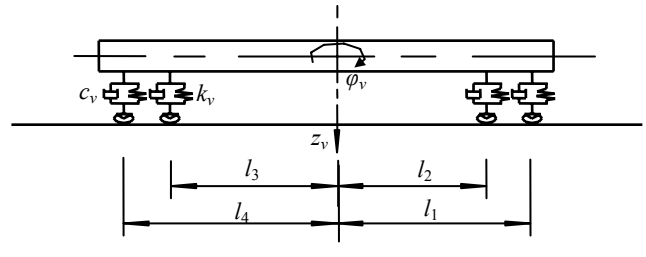


Figure 2 The vehicle model.

Motion Equation. Referring to Figure 2, the vertical motion equations for each vehicle can be expressed as (Su et al. 2001):

$$\begin{cases} m_v \ddot{z}_v(t) + c_v \sum_{j=1}^{4} \dot{u}_j + k_v \sum_{j=1}^{4} u_j = 0 \\ J_v \ddot{\varphi}_v(t) + c_v \sum_{j=1}^{4} \dot{u}_j(\pm l_j) + k_v \sum_{j=1}^{4} u_j(\pm l_j) = 0 \end{cases} \tag{1}$$

where u_j is a deformation of the vehicle jth axle suspension spring, z_v and φ_v are vertical and pitch displacements of the vehicle from the equilibrium position, c_v and k_v represent the spring damping and stiffness of the vehicle's suspension spring for each wheel-axle set, m_v and J_v are, respectively, the mass and pitch moment of inertia

for each vehicle, l_j is the longitudinal distance between the centroid of the vehicle and its *j*th wheel-axle set.

The motion equation for the bridge can be expressed as (Su et al. 2001):

$$\frac{\partial^2}{\partial x^2}(EI\frac{\partial^2 z_b}{\partial x^2})+m_b\frac{\partial^2 z_b}{\partial t^2}+c_b\frac{\partial z_b}{\partial t}= \sum_{i=K}^{M}\{\delta[x-(vt-a_i)](P_i-m_s\frac{\partial^2 z_b}{\partial t^2}+c_v\dot{u}_i+k_v u_i)\} \tag{2}$$

where, EI, m_b, and c_b are the vertical flexural rigidity, unit length mass, and damping coefficient of the bridge, respectively, v is the train running speed, a_i is the longitudinal distance from the *i*th wheel-axle set to the first wheel-axle set of the train (Figure 1), P_i is the static weight of the *i*th wheel-axle, m_s is the unsprung mass of the *i*th wheel-axle set, c_v and k_v are the damping and stiffness of the vehicle's suspension spring for the *i*th wheel-axle set, $\delta(x-\eta)$ is Dirac function, it can be defined as:

$$\delta(x-\eta)=\begin{cases}1, & x=\eta\\ 0, & x\neq\eta\end{cases} \tag{3}$$

u_i is deformation of the vehicle *i*th wheel-axle set suspension spring that is the relative displacement u_i and the relative velocity $\dot{u}_i$. They are expressed as:

$$\begin{cases}u_i=z_v(t)\pm l_i\varphi_v(t)-(z_{bi}-\omega_i)\\ \dot{u}_i=\dot{z}_v(t)\pm l_i\dot{\varphi}_v(t)-(\dot{z}_{bi}-\dot{\omega}_i)\end{cases} \tag{4}$$

where, z_v and $\dot{z}_v$ are respectively the vehicle's vertical displacement and vertical speed corresponding to the *i*th wheel-axle set, l_i is marked in Figure 2, φ_v and $\dot{\varphi}_v$ are respectively the vehicle's pitch displacement and pitch speed corresponding to the *i*th wheel-axle set, z_{bi} and $\dot{z}_{bi}$ are respectively the bridge's vertical displacement and vertical speed, ω_i is vertical track irregularity under the *i*th wheel-axle set.

The vehicle-bridge system motion equation can be obtained by coupling Eq.1 and 2.

Solve Motion Equation. It has been proven that damping scarcely influences the natural frequency and the vibration mode, consequently, the damping can be neglected when the frequency and the mode are calculated (Jiang 1984). For this reason, the damping is neglected in this paper.

Eq. 2 can be solved by using mode superposition method. Where L modes are considered, and the bridge vertical vibration displacement can be expressed as:

$$z_b(x,t) = \sum_{j=1}^{L} q_j(t)\phi_j(x) \tag{5}$$

where, $\phi_j(x)$ is the bridge *j*th mode. Substituting Eq. 5 into Eq. 2 and using orthogonality of the vibration mode, the characteristic equation can be obtained:

$$\left|K - \omega^2 M\right| = 0 \tag{6}$$

To the continuous bridge, five modes are considered.

Example. The situation of the train on the bridge is changed with the time, that is to say, action points of the train's wheel-axle sets acting on the bridge are changed with the time. Consequently, the mass matrix and the stiffness matrix of the vehicle-bridge system are changed with the time. In this paper, the generalized ***Jacobi*** method is used to calculate this system's vertical natural frequency for every time step. And 20 high-speed vehicles are considered in order to make the vehicles distribute on the whole bridge and keep for a moment. The vehicle has the high-speed passenger vehicle parameters, the wheel-axle weight and the wheelbases are marked in Figure 3. This paper calculated the continuous bridge vertical load-carrying natural frequency for the 3×64 m through stud welding steel truss bridge which has two kinds of bridge deck, one is open deck and the other is ballast deck.

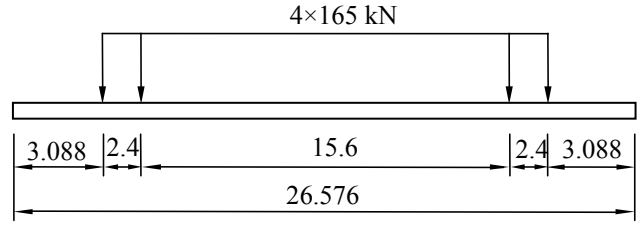


Figure 3. The high-speed passenger vehicle wheel-axle weight and the wheelbase (Unit: m).

***The 3×64 mThrough Stud Welding Steel Truss Bridge*(*Open Deck*).** This bridge is 11 m high, the distance between the maintruss is 5.75 m, and the length of internode is 8 m.

Firstly, the bridge vertical natural frequency is calculated (Table 1). By only studying the bridge's vertical vibration, the bridge model is simplified using a vertical plane model, and the element is a plane truss. One bar in the main truss is one element, the stiffness and the mass are, respectively, the equivalent stiffness and the equivalent mass, and the nodes that connect the elements are hinge joint. There are 93 elements in the bridge model (Figure 4).

There are 20 high-speed passenger vehicles with the same parameters that are traveling on this bridge with the speed of 160 km/h. When the first wheel-axle set is

on the bridge, it is start time. The variation curves with time of five (from 1 to 5) load-carrying natural frequencies are obtained (Figures 5-9).

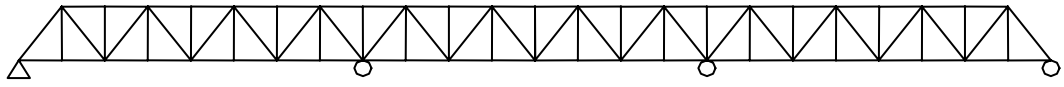


Figure 4. The model of the 3×64 m through stud welding steel truss bridge.

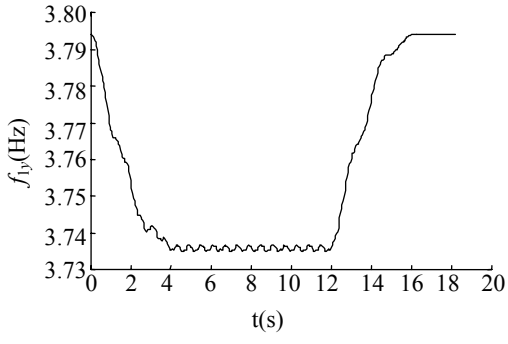


Figure 5. The first vertical load-carrying natural frequency variation curve of the 3×64 m through stud welding steel truss bridge.

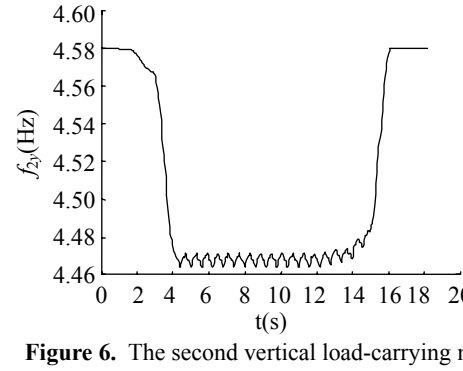


Figure 6. The second vertical load-carrying natural frequency variation curve of the 3×64 m through stud welding steel truss bridge.

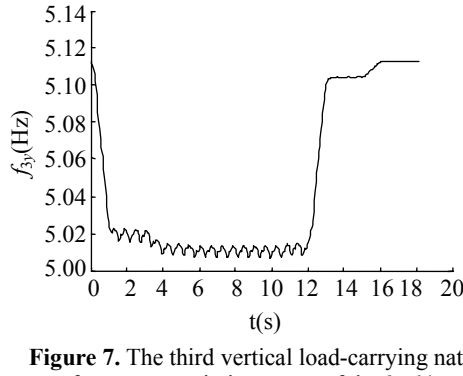


Figure 7. The third vertical load-carrying natural frequency variation curve of the 3×64 m through stud welding steel truss bridge.

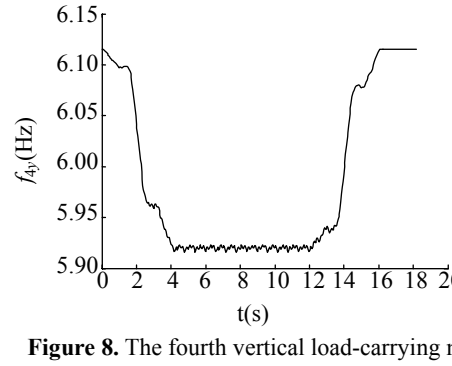


Figure 8. The fourth vertical load-carrying natural frequency variation curve of the 3×64 m through stud welding steel truss bridge.

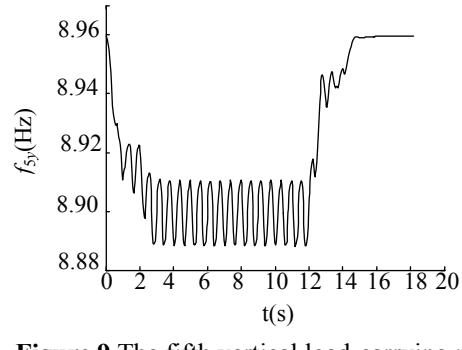


Figure 9. The fifth vertical load-carrying natural frequency variation curve of the 3×64 m through stud welding steel truss bridge.

To the ballast deck continuous through stud welding steel truss bridge, there has been the same regularity, although the numerical value is a little different (Table 1).

Table 1. The vertical load-carrying natural frequency of the railway continuous bridge and the railway simply supported bridge.

Bridge type	Frequency number	Natural frequency/ Hz	Variation range /Hz	Maximum D-value /Hz	Medium value/Hz	δ /%
64 m simply supported steel truss bridge (Su et al. 2001)	The first	3.850	3.764~3.759	0.005	3.762	-2.36
3×64 m through stud welding steel truss bridge (open deck)	The first	3.7942	3.7368~3.7349	0.0019	3.7359	-1.5616
	The second	4.5801	4.4713~4.4639	0.0074	4.4676	-2.5369
	The third	5.1124	5.0140~5.0070	0.0070	5.0105	-2.0605
	The fourth	6.1151	5.9232~5.9155	0.0077	5.9194	-3.2649
	The fifth	8.9596	8.9106~8.8881	0.0225	8.8994	-0.7978
3×64 m through stud welding steel truss bridge (ballast deck)	The first	3.7426	3.6883~3.6865	0.0018	3.6874	-1.5005
	The second	4.5191	4.4153~4.4081	0.0072	4.4117	-2.4558
	The third	5.0431	4.9487~4.9419	0.0067	4.9453	-2.0057
	The fourth	6.0330	5.8490~5.8416	0.0074	5.8453	-3.1726
	The fifth	8.8351	8.7881~8.7663	0.0217	8.7772	-0.7786

where:

$$\delta = \frac{(\text{the minimum of the bridge vertical load-carrying natural frequency-the bridge vertical natural frequency})}{\text{the bridge vertical natural frequency}} \times 100\%$$

Result Analysis. The results about the continuous bridge show that there are three stages in the time-varying curve of the vertical load-carrying natural frequency during the train traveling on the bridge. Firstly, as soon as the train runs on the bridge, the bridge vertical load-carrying natural frequency begins to gradually decline. Secondly, when the vehicles are distributed on the whole bridge, the bridge vertical load-carrying natural frequency is periodically varied. Thirdly, when the vehicles begin to off the bridge, the bridge vertical load-carrying natural frequency begins to gradually increase, until the train completely leaves the bridge, the bridge vertical load-carrying natural frequency gradually returns to the natural frequency.

The reason the bridge vertical load-carrying natural frequency periodically varied is because the position of the wheel-axle set on the bridge is periodically varied when the train travels on the bridge. If the vehicles have the same parameters, for one position (such as midspan), after every time interval $T=v/L_v$, there will be a completely similar wheel-axle set located at this position. $T=v/L_v$ is the changing period of the train act on the bridge, i.e., the changing period of the equivalent mass and the equivalent stiffness of the vehicle-bridge system.

From Table 1 and Figures 5-9, it is determined that the variation range of the vertical load-carrying natural frequency of the 3×64 m through stud welding steel truss bridge is small, the maximum D-value is 0.0225 Hz<0.1 Hz. The D-value between the vertical load-carrying natural frequency and the vertical natural frequency is very small, and the maximum deviation percentage is only 3.2649%. And those values are smaller than those of the 64 m simply supported through stud welding steel truss bridge. From Table 1, for the continuous through stud welding steel truss bridge, the natural frequency and the load-carrying natural frequency of the open deck bridge are all larger than that of the ballast deck bridge.

The variation range of the former four vertical load-carrying natural frequencies of the 3×40 m deck plate steel girder bridge is also small, the maximum D-value is 0.0650 Hz<0.1 Hz, the variation range of the fifth frequency is slightly large, and the D-value is 0.2790 Hz. The D-value between the vertical load-carrying natural frequency and the vertical natural frequency is very small, the maximum deviation percentage is only 4.2153%.

From the abovementioned information, when the train has fewer vehicles, that is to say, when the length of the train is shorter than the length of the bridge, the periodical changing stage of the vertical load-carrying natural frequency would never occur, then the vehicle-bridge system would not incur vertical resonance. And when the girder's mass is greater, the difference between the load-carrying natural frequency and the free frequency is smaller.

Main Influential Factors for the Vertical Load-Carrying Nnatural Frequency

The main influential factors include unsprung mass, suspension spring stiffness of the wheel-axle set, vehicle mass, speed of the train, vehicle length, etc. These factors regularly influence the vertical load-carrying natural frequencies and are in the two continuous bridges. Consequently, only give the influential regularity about the first vertical load-carrying natural frequency of the 3×64 m through stud welding steel truss bridge (open deck).

Unsprung Mass. When the other parameters keep constant and the unsprung mass of the vehicle wheel-axle set is 1800 kg, 2500 kg, and 3200 kg, Figure 10 shows the variation curves with time about the 3×64 m through stud welding steel truss bridge (open deck) first vertical load-carrying natural frequency.

From the curves in Figure 10, it can be obtained when the unsprung mass of each wheel-set of the vehicle is greater, the bridge's vertical load-carrying natural frequency is smaller.

Suspension Spring Stiffness of the Wheel-axle Set. When a train is travelling on the bridge, traditional view considers that the bridge equivalent mass is increased, therefore, the bridge vertical load-carrying natural frequency is certainly less than the natural frequency. In fact, in the vehicle-bridge system, if the suspension spring

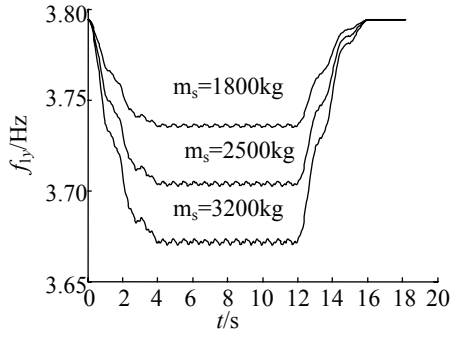


Figure 10. The influence of the unsprung mass.

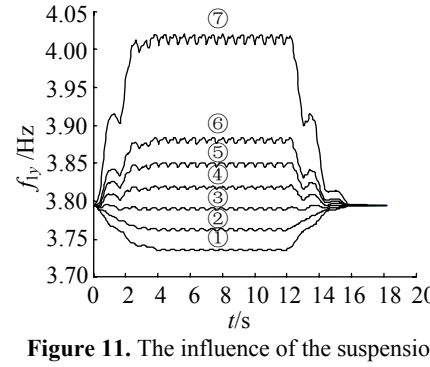


Figure 11. The influence of the suspension spring stiffness of the wheel-axle set.

stiffness is only changed, the bridge vertical load-carrying natural frequency is possibly greater than the natural frequency (Figure 11). Where, curve ①~⑦ are respectively the bridge vertical load-carrying natural frequency variation curve, when the suspension spring stiffness of the calculated vehicle is 1 time, 2 times, 3 times, 4 times, 5 times, and 6 times of the high-speed vehicle.

The curves indicate that when the suspension spring stiffness is greater, the bridge vertical load-carrying natural frequency is greater too. That is to say, when the train on the bridge, the bridge equivalent mass is changed, at the same time, the bridge equivalent stiffness is also changed. Consequently, the bridge vertical load-carrying natural frequency can be greater than the corresponding vertical natural frequency.

The Vehicle Length. If other vehicle parameters aren't changed and when the vehicle length is 26.576 m, 24.576 m, 22.576 m, and 20.576 m, the variation situation of the first vertical load-carrying natural frequency maximum and the minimum is shown in Table 2.

Table 2. The influence of the vehicle length.

Vehicle length L_v/m	The first vertical load-carrying natural frequency/Hz		
	$f_{1\min}$	$f_{1\max}$	D-value
26.576	3.7348	3.7367	0.0019
24.576	3.7311	3.7333	0.0022
22.576	3.7272	3.7287	0.0015
20.576	3.7224	3.7229	0.0005

From Table 2, it can be obtained that when the vehicle is shorter, the maximum $f_{1\max}$ and the minimum $f_{1\min}$ of the bridge vertical load-carrying natural frequency are all smaller.

The Vehicle Mass. If other vehicle parameters aren't changed and when the vehicle mass is 48.4 t, 58.4 t, and 68.4 t, the variation situation of the first vertical load-carrying natural frequency maximum and the minimum is shown in Table 3.

From Table 3, when the vehicle mass increases or decreases 10 t, the variation of the bridge vertical load-carrying natural frequency is merely 0.0002 Hz, and the D-value between the bridge vertical load-carrying natural frequency maximum and the minimum are invariable, and are all 0.0019 Hz. In a word, the vehicle mass hardly influences the value of the bridge vertical load-carrying natural frequency.

The Train Speed. If other vehicle parameters aren't changed and when the train speed is 160 km/h, 140 km/h, and 120 km/h, the variation situation of the first vertical load-carrying natural frequency maximum and the minimum is shown in Table 4.

Table 3. The influence of the vehicle mass.

Vehicle mass/t	The first vertical load-carrying natural frequency/Hz		
	$f_{1\min}$	$f_{1\max}$	D-value
48.4	3.7350	3.7369	0.0019
58.4	3.7348	3.7367	0.0019
68.4	3.7346	3.7365	0.0019

Table 4. The influence of the train speed.

Train speed/ $\mathrm{m \cdot s^{-1}}$	The first vertical load-carrying natural frequency/Hz		
	$f_{1\min}$	$f_{1\max}$	D-value
44.444	3.7348	3.7367	0.0019
38.889	3.7348	3.7367	0.0019
33.333	3.7348	3.7367	0.0019

From Table 4, when the train speed is changed, the bridge vertical load-carrying natural frequency maximum and the minimum are invariable.

Conclusions

Based on the calculation and analyses of the continuous bridge vertical load-carrying natural frequency, some useful results are obtained:

- When the train is travelling on the bridge and the vehicles are distributing on the bridge, the continuous bridge vertical load-carrying natural frequency is periodically varied, and the variation period is $T=v/L_v$.
- The variation regularity of the continuous bridge vertical load-carrying natural frequency is similar with that of the simply supported bridge. As soon as the train runs on the bridge, the bridge vertical load-carrying natural frequency begins to gradually decline. When the vehicles are distributing on the whole bridge, the bridge vertical load-carrying natural frequency is periodically varied. When the vehicles begin to get off the bridge, the bridge vertical load-carrying natural frequency begins to gradually increase, until the train completely leaves the bridge, the bridge vertical load-carrying natural frequency gradually returns to the natural frequency.

- When the unsprung mass of each wheel-set of the vehicle is heavier, the bridge's vertical load-carrying natural frequency is smaller. When the suspension spring stiffness is greater, the bridge vertical load-carrying natural frequency is greater, too, and the bridge vertical load-carrying natural frequency can even be greater than the corresponding vertical natural frequency. When the vehicle is shorter, the maximum $f_{1\max}$ and the minimum $f_{1\min}$ of the bridge vertical load-carrying natural frequency are all smaller. The vehicle mass hardly influences the value of the bridge vertical load-carrying natural frequency. When the train speed is changed, the bridge vertical load-carrying natural frequency maximum and the minimum are invariable.

References

Chu, K., Garg, V. K., and Dhar, C. L. (1979). "Railway-bridge impact: simplified train and bridge model." *Journal of the Structure Division*, 1823-1844.

Diana, G., and Cheli, F. (1989). "Dynamic interaction of railway systems with large bridges. " *Vehicle System Dynamic*, 71-106.

Fafard, M. M., and Savard, M. (1993). "Dynamics of bridge vehicle interaction." *Proc EURODYN'93*, 951-960.

He, X., and Zhang, N. (2005). "Dynamic analysis of railway bridge under high-speed trains." *Computers and Structures*, 1891-1901.

Jiang, C. (1984). "Finite element method basis." *Tsinghua University Press.*

Ju, S. H., and Lin, H. T. (2003). "Resonance characteristics of high-speed trains passing simply supported bridges." *Journal of Sound and Vibration*, 1127-1141.

Li, J., and Su, M. (1999). "The resonant vibration for a simply supported girder bridge under high-speed trains." *Journal of Sound and Vibration*, 897-915.

Majka, M, and Hartnett, M. (2008). "Effects of speed, load and damping on the dynamic response of railway bridges and vehicles." *Computers and Structures*, 556-572.

Su, M., Li, J., and Liang, Z. (2001). "Study on vertical loaded frequencies of railway simply supported beam bridges." *Journal of the China Railway Society*, 76-80.

Wang, F. (1983). "Vehicle dynamics." *China railway publishing house*.

Wu, Y. S., and Yang, Y. B. (2003). "Steady-state response and riding comfort of trains moving over a series of simply supported bridges." *Engineering Structures*, 251-265.

Xia, H. (2002). "Dynamic interaction of vehicles and structures." *Beijing: Science Press*.

Xia, H, Zhang, N, and Guo, W. W. (2006). "Analysis of resonance mechanism and conditions of train-bridge system." *Journal of Sound and Vibration*, 810-822.

Yang, Y. B., Yau, J. D., and Hsu, L. C. (1997). "Vibration of simple beams due to trains moving at high speeds." *Engineering Structures*, 936-994.

Yang, Y. B., Yau, J. D., and Wu, Y. S. (2004). "Vehicle-bridge interaction dynamics." *Singapore: World Scientific Pub Co Inc.*

Yau, J. D. (2009). "Response of a train moving on multi-span railway bridges undergoing ground settlement." *Engineering Structures*, 2115-2122.

Anti-cracking Ability of Asphalt Mixture Added with Polyester Fiber

Xin Yu[1,2] and Lijun Sun [1]

[1] Tongji University, School of Transportation Engineering, Siping Rraod 1239 #, Shanghai, P.R.China, 200092; Email: njyxzq@163.com
[2] Hohai University, School of Civil Engineering and Transportation, Xikang Raod #1, Nanjing, P.R.China, 210098; Email: hhu_yuxin@163.com

Abstract

Cracking is one of the main disease types of asphalt pavement. Currently, adding polyester fibers is one of the measures most commonly used to reduce cracking. Firstly, performance-enhancing mechanism on anti-cracking ability of asphalt mixture adding with polyester fibers was studied. Then, a typical polyester fiber was selected, anti-cracking ability of asphalt binder and asphalt mixture for different fiber lengths and different fiber dosages were tested. BBR test was selected to evaluate anti-cracking ability of asphalt binder while four-point bending beam test was selected to evaluate anti-cracking ability of asphalt mixture. Study showed that anti-cracking ability of asphalt was excellent when the length of polyester fiber is 6 mm, and the optimum dosage is between 0.2 ~ 0.5% of the mixture.
Key Words: Anti-cracking ability; asphalt mixture; polyester fiber

Introduction

Cracking is one of the main disease types of asphalt pavement especially in cold regions, which is mainly due to a combined action of temperature stress and repeated load stress. Cracking occurs when composite stress exceeds the allowable strength of asphalt mixture. To improve the anti-cracking performance of asphalt mixture, we can take measures to improve shrink properties of asphalt mixture or improve tensile strength of asphalt mixture. Studies have shown that mixture added polyester fibers can effectively improve anti-cracking performance of mixture, which can greatly reduce road surface cracking (Peng et al. 2002) (Serfass et al. 1996) (Cleven 2000). In this paper, anti-cracking mechanism of mixture added polyester fibers is first analyzed, then, laboratory tests were taken to study the influence of different dosages and different fiber lengths on the anti-cracking performance. The conclusion can provide guidance in applications of polyester fibers in asphalt.

For asphalt mixture added polyester fiber, we can regard polyester fiber as ellipsoidal inclusion buried in the asphalt concrete. Using Eshelby's research results of equivalent inclusion body, assuming viscous-elastic mode as the Maxwell model described, we can get one-dimensional constitutive equation of asphalt concrete added polyester fiber:

$$\sigma = q/(q-f)[E_m\varepsilon + E_c\int_0^t \varepsilon \exp(-(t-\tau)/\theta)d\tau] \quad (1)$$

$$q = 2(1+v_0)\alpha^2[\ln(2/\alpha) - 0.25\times(5-4v_0)/(1-v_0)] \quad (2)$$

In the equation above,
α --Ratio of diameter and length of fiber; f --Fiber volume ratio; v_0 --Poisson's ratio; E_c --Modulus of asphalt concrete; E_m --Modulus of asphalt concrete added with polyester fiber.

For the mechanism of the anti-cracking performance of asphalt mixture added polyester fibers, several theories are currently widely recognized among which anti-cracking and reinforced theory are representative.

Anti-Cracking Mechanism According to Fracture Mechanics. According to fracture mechanics, defects or cracks exist inside material. Combined action of temperature stress and repeated load stress may produce stress concentration at crack tip, which promotes further expansion of the crack. Fiber has characteristics such as high tensile strength, high modulus, small diameter, and so on. When fiber is uniformly dispersed in the asphalt mixture, a three-dimensional network structure formed which plays a blocking role in crack propagation and constrains further development, thus improves anti-cracking ability of the mixture (Chen et al. 2005).

According to Ramualdi's K Stacking method, stress intensity factor of asphalt mixture crack added fiber can be expressed as:

$$K_{fm} = K_m - K_f \quad (3)$$

where K_{fm} -- stress intensity factor of asphalt mixture added with fiber; K_m -- stress intensity factor of asphalt mixture; K_f -- stress intensity factor of fiber.

From the Eq. 3, we can see the stress intensity factor of asphalt concrete is significantly reduced after adding polyester fiber, thus effectively prevents crack formation and development.

Reinforced Mechanism of Fiber to Mixture. According to composite materials theory, taking into account the coupling influence of fiber length, interfacial adhesion, and other factors, modulus and tensile strength of asphalt concrete can be separately calculated by the following two equations:

$$E_c = E_m V_m + \eta E_f V_f \quad (4)$$

$$\sigma_c = \sigma_m V_m + \eta \sigma_f V_f \quad (5)$$

where, E_m -- modulus of asphalt concrete; E_c -- modulus of asphalt concrete added polyester fiber; E_f -- elasticity modulus of fiber; V_f and V_m -- volume ratio of fiber and asphalt mixture, respectively; E_m--asphalt mixture stiffness modulus. η-- the effective coefficient considering the factors such as fiber length, orientation, distribution, interface bonding, fiber defects, and so on.

As the modulus of fiber is much higher than that of the asphalt mixture, we can see fiber strengthen asphalt mixture like "micro-reinforcement."(Danie et al. 2001). From the analysis above we can also see that the reinforcing effect of fiber is related to fiber properties such as dispersion state, length, interface bonding of fiber and asphalt mixture, and so on.

To analyze the effect of polyester fiber on anti-cracking performance of asphalt, BBR tests were carried out using asphalt binder with different fiber dosage and different fiber length at -12 ℃and-18 ℃, respectively (Hoare et al. 2000). Test asphalt binder chose shell 70# and the polyester fiber used named “Lenovo.”

The two evaluation parameters of BBR testing are creep stiffness modulus (S value) and the change rate of stiffness modulus (m value) when load acting on the asphalt binder. To restrict the pavement crack, SHRP Performance Specifications limited the

S value and m value at 60 s were not greater than 300 MPa and not less than 0.30, respectively(guideline 2001).

Different Fiber Dosage Test Result. The dosages of test fibers were of 1%, 4.2%, and 8.3% (ratio of binder).The length of fiber in the test is 6 mm.

The changing trend of the S value and M value of fiber asphalt binder with different temperature and different fiber dosages is shown in Figure 1.

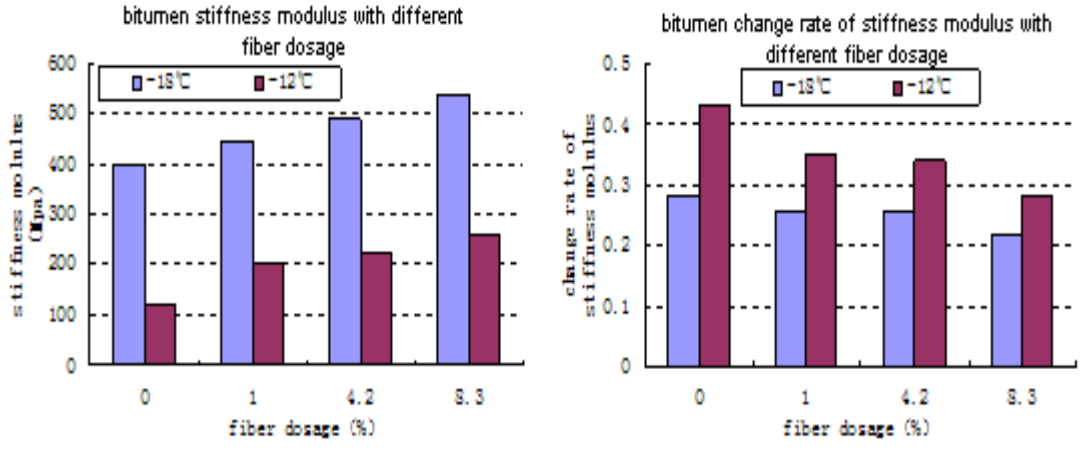


Figure 1. BBR test results of asphalt binder with different fiber dosage.

From the data of Figure 1, the following conclusions were determined:

(1) When temperature rises from -18 °C to -12 °C, the S value of asphalt binder is greatly reduced, and the M value is greatly increased, which shows the temperature sensitivity of the asphalt binder added fiber.

(2) The S value of asphalt binder increased with the increase of fiber dosage, and the M value increased with the decrease of fiber dosage, which indicates that the enhancement role of fiber to asphalt binder was very obvious with the increase of fiber dosage.

(3) According to the SHRP performance specifications, the anti-cracking performance of asphalt binder with no fiber is better than that with fiber, but it is clear that mixtures added fiber will help to increase the anti-cracking performance of the mixture, so the lower limit of asphalt binder stiffness modulus should be proposed to ensure that the asphalt binder had a certain consistency.

Different Fiber Length Test Result. To analyze the affects of polyester fiber length on properties of asphalt binder, the author carried on BBR test at -18°C and -12°C for the fiber asphalt binder and pure asphalt binder. The lengths of the fiber tested were 3 mm, 6 mm, and 10 mm. The dosage of fiber in test is 4.2%.

The changing trend of the S value and M value of fiber asphalt binder with different temperature and different fiber lengths was shown in Figure 2.

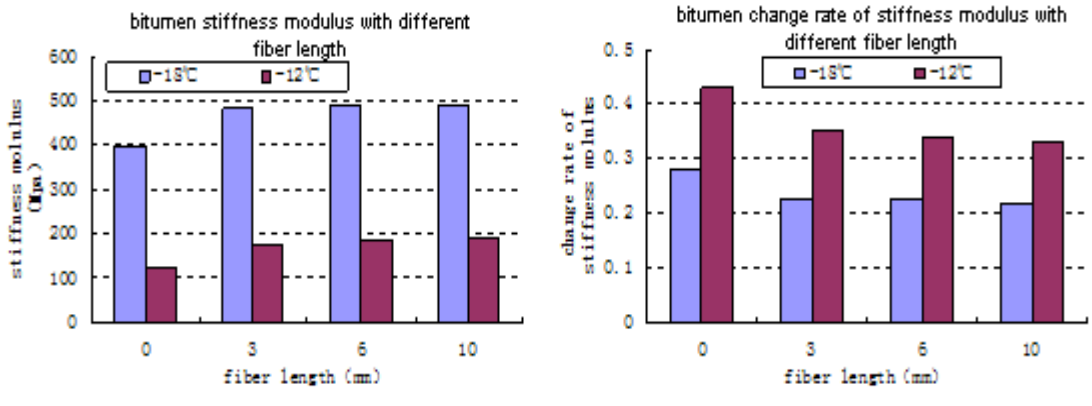


Figure 2. BBR test results of fiber asphalt binder with different fiber length.

From the data of Figure 2, the following conclusions can be determined:
The S value of asphalt binder increased greatly after fiber was added, and the S value grow with the increase of the fiber length, while the M value decreased with the increase of the fiber length.

The S value of asphalt binder increased with the fiber length increased reflected the enhancement of fiber-reinforced role. The main reason of which was that before exceeding the critical length, with the growth of fiber length, the force passed to the

fiber through the interface increased, and the superior mechanical properties of fibers can be more fully exerted. The fiber in the mixture will bound creep behavior of asphalt binder, which decreases the M value, whose magnitude relates to the enhancement effect of fiber. The longer the fiber's length is, the greater the force on the fiber and asphalt binder interface and the larger of the bound force of the fiber to the creep behavior of asphalt binder.

The BBR test results of fiber asphalt binder added polyester fibers with different dosages and different fiber lengths reflected the strength of fiber-reinforced role. As the reinforcing effect increased, the S value of fiber asphalt binder increased while the M value decreased. According to SHRP material performance analysis method, the anti-cracking performance of asphalt binder became worse after adding fibers. It was contrary to the facts because fiber can surely enhance the anti-cracking performance of asphalt binder. It showed that there were some flaws using the SHRP method to evaluate anti-cracking performance of asphalt binder added polyester fibers.

To analyze the effect of polyester fiber on anti-cracking performance of asphalt mixture, the bending beam tests were conducted using different fiber lengths and different fiber dosages.

During testing, the mixture board of size 300 mm × 300 mm × 60 mm was made with a wheel grinding machine and the board density was controlled between 99% to101% of the standard Marshall Compaction density. Then the board was cut to prism trabecular specimens of 50 mm × 50 mm × 250 mm, each kind of mixture prepared four specimens. Trabecular specimens were placed into an incubator 5 hours at -10℃, and then the trabecular specimens were set on MTS testing machine to carry on 4-point bending beam test with a loading rate of 50 mm/min.

The following is the bending beam test result of mixture added polyester fibers:

Effect of Fiber Length to the Anti-Cracking Performance of Mixture. During the test, different fiber lengths of 3 mm, 6 mm and 15 mm were tested, respectively, while the fiber dosage was fixed 2.5‰ of mixture. The 4-point bending beam test results of fiber asphalt mixture with different fiber length at -10 ^{o}C were shown in Table 1.

Table 1. The bending beam test results of mixture with different fiber length.

Fibers length	Flexural failure strength (MPa)	Flexural failure strain (10^{-6})	Bending stiffness modulus (MPa)
3 mm	12.88	3639.10	2849.24
6 mm	13.44	4068.10	3304.40
15 mm	14.74	3869.76	3889.14
No fibers	11.49	3754.43	3369.19

Analyzing the data in Table 1 we can see that fibers can improve the bending tensile strength of asphalt mixture effectively, with the increasing of the fiber length, the bending tensile strength of asphalt mixture increases, which indicates that an appropriate increase of fiber length can increase the force act on the fibers passing through the interface from the asphalt mixture, which helps to exert the characteristics of fiber material thus improve the strength of the asphalt mixture.

From the flexural failure strain results of regarding different lengths, 6 mm fibers were optimal to improve the anti-cracking performance of the mixture.

According to damage mechanics theory, strength and deformation are two important technical parameters of the material. There are some limitations to evaluate the anti-cracking with flexural strain or flexural strength as the control target, a comprehensive parameter need to be found to reflect the strength and deformation.
Based on the material damage criterion, material damage includes three stages: initiation, expansion, and the final termination of cracks. Assuming that the destruction form of material corresponds with the energy change per unit volume, then the material damage can be expressed with a function of strain energy density (Si et al. 2002):

$$\frac{dW}{dV} = \int_0^{\varepsilon_0} \sigma_{ij} d\varepsilon_{ij} \tag{6}$$

where, $\frac{dW}{dV}$-- for the strain energy density function. σ_{ij} ,-- ε_{ij} , represents the stress and the strain components. ε_0 -- the critical strain.

The critical strain energy density of material reflects the energy required for cracking, which can evaluate the anti-cracking performance of the fiber asphalt mixture. The greater the critical strain energy density of the materials, the greater the energy required for destroying the material, and the better the anti-cracking performance of the material is. Critical strain energy calculation results were shown in Table 2.

From the data, the critical strain energy of asphalt mixture added fibers was greater than the asphalt mixture with no fibers, among which the critical strain energy was greatest when fiber length was 6 mm length, so when the dosage is 2.25‰, 6 mm length fiber has the best low temperature performance.

Table 2. The bending beam test results of mixture with different fiber lengths.

Fiber length	No fibers	3 mm	6 mm	15 mm
Critical strain energy density ($kJ.m^{-3}$)	19.73	25.80	29.68	26.94

Table 3. The bending beam test results of asphalt mixture with different dosages.

Fiber dosage	Max load (kN)	Mid-span deflection (mm)	Bending tensile strength (MPa)	Stiffness modulus (MPa)	Failure strain (με)	Standards (με)
No fibers	989.0	0.492	8.07	3126.9	2582.0	>2000
0.2%	1064.0	0.547	8.69	3027.3	2869.1	
0.3%	1159.0	0.622	9.46	2899.2	3263.4	
0.5%	1189.0	0.570	9.71	3244.1	2992.0	

Effect o Fiber Dosage to The Anti-Cracking Performance of Mixture. In order to analyze the effect of polyester fiber dosage on anti-cracking performance of asphalt mixture, different fiber dosages of 0.2%, 0.3%, and 0.5% were carried out, and the fiber length was fixed 6 mm. Different fiber dosages at -10^{o}C were shown in Table 3.

The data in Table 3 shows that the anti-cracking performance of the mixture improved significantly after adding polyester fibers. And with the increase of the fiber dosage, the failure strain of asphalt mixture increases first and then decreases, which indicates that there is a optimum value of fiber dosage between 0.2% to 0.5%, which maximums the asphalt mixture low temperature failure strain.

Conclusion

With regard to the effect of polyester fibers on the anti-cracking performance of asphalt binder and asphalt mixture, through the testing and analysis, the main conclusions obtained are as follows:

(1) There are certain deficiencies with SHRP stiffness evaluation methods to evaluate the BBR test results of fiber asphalt binder, and the evaluation findings are inconsistent with the actual performance.
(2) By analyzing the low-temperature bending beam test data of polyester fiber asphalt mixtures, we can see that polyester fiber can effectively improve the anti-cracking performance of a mixture.
(3) For polyester fibers asphalt mixture, the asphalt mixture has the best anti-cracking performance when fiber length is 6 mm, and the fiber dosage is between 0.2% to 0.5%.

References

Asphalt Institute, Lexington, Kentucky. (2001). *Superpave Mix Design* (SP-2).

Chen, J.-S., and Lin, K.-Y. (2005). “Mechanism and behavior of bitumen strength reinforcement using fibers.” *Journal of Materials Science*.

Cleven, M. A. (2000). *Investigation of the properties of carbon fiber modified asphalt mixtures*, Master dissertation: Michigan Technological University.

Danie, J. S., Kim, Y. R. (2001). "Laboratory evaluation of fatigue damage and healing of asphalt mixtures." *Journal of Materials in Civil Engineering*.

Hoare, T. R., and Hesp, S. (2000). "Low-temperature fracture testing of asphalt binders, regular and modified systems." .*TRB, National Research Council*. Washington, D.C.

Peng, B., Jin. M. (2002). "Studies of fiber-reinforced asphalt mixture performance." *Journal of Chongqing Jiaotong University*.

Serfass, J. P., and Samanos, J. (1996). "F*iber-modified asphalt concrete characteristics, applications and behavior*." AAPT, 65.

Si, Z.,Little, D. N.,Lytton, R. L. (2002). "Characterization of microdamage and healing of asphalt concrete mixtures." *Journal of Materials in Civil Engineering*.

Centrifugal Model Tests on Settlement Controlling of Piled Embankment in High Speed Railway

Changdan Wang[1], Binglong Wang[2], Xu Wang[3], Shunhua Zhou[4]

[1] Ph. D candidate, [2] Prof., [4] Prof., Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, 4800 Caoan Road, Shanghai, People's Republic of China 201804; Email: tjjtwangcd@yeah.net. cn, wangbl8@163.com.cn
[3] Prof., School of Civil Engineering, Lanzhou Jiaotong University, 88 Anning Road(w), Lanzhou Gansu, People's Republic of China 730070

Abstract

The settlement after construction controlling of subgrade is a crucial issue on designing of high speed railway. Based on the experimental study at home and abroad, the centrifuge model test is used to imitate piled embankment in different pile spacing to research and analyze the settlement controlling ability of piled embankment in collapsible loess ground in high speed railway. Then the relationship between pile, lime-soil cushion, soil between piles controlling the settlement are concluded by analyzing the test data. The research results show that untreated ground which can not meet the request on settlement after construction in ballasted track in high speed railway need be reinforced. The total settlement, settlement after construction and settlement ratio of piled embankment get larger with the extending of pile spacing. The settlement of soil is faster and more increasing than settlement of pile, and pile-soil load share ratio get lower and negative skin friction of pile get more increasing.

Key Words: subgrade engineering; collapsible loess; piled embankment; centrifuge model test; settlement

Introduction

The development of high speed railway demands a subgrade that possess such engineering properties as higher strength, greater rigidity, more stable and durable, lower strain and homogenous in longitudinal(Wang 2007). Through the research of condition operation in existed railway shows that : The deteriorating state of existed railway which caused by excessive settlement of subgrade is widespread and

especially in collapsible loess area. Thus the controlling of subgrade settlement becomes core problems and difficulties in railway construction. The investigation of the settlement of subgrade in the high speed railway after construction becomes the main control factor in the high speed railway subgrade engineering. The collapsible loess foundation is an unsaturated and underconsolidated soil with a larger porosity and a low dry density. The most prominent feature of collapsible loess is the additional settlement is rapid and larger by water and self weight of soil or additional stress(Qian and Luo, 1985). Therefore, the research on the settlement controlling is a matter of great significance in theoretical research and engineering application on collapsible loess foundation in high speed railway.

Literature Review

The piled embankment is one of the new techniques of foundation treatment, which is composed of rigid pile, lime-soil cushion and soil between piles. The piled embankment is applied in construction, research and testing more and more in civil engineering (Yan and Zhang, 2001; Gong, 2002; Rao and Zai, 2004). The main differences of piled embankment are as follows in compare with in subgrade engineering and civil engineering. First, the foundation of civil engineering is rigid foundation and the foundation is flexible foundation in subgrade engineering. Second, the static load is in civil engineering and the dynamic load is in low railway subgrad mainly. Third, the target of settlement controlling is total settlement and differential settlement in civil engineering and settlement after construction in subgrade engineering

.

The centrifuge model test is the best physical test method with similarity in the research of geotechnical engineering. And centrifuge model test can imitate the deformation processing similarly with prototype soil in gravity, so the method is used widely in laws of soil deformation and geotechnical research (Huang and Wang, 1998; Lee and Chen, 2002; Zeng, 2003; Wang, 2008). At present, the study on the centrifuge model tests of piled embankment in high speed railway is not seen in the literature yet. The mechanism of interaction between pile and soil between piles is inadequate on composite foundation and need further study.

The centrifuge model test is used to imitate piled embankment in different pile spacing to research and analyze the settlement controlling ability of piled embankment in collapsible loess ground in high speed railway. Then the relationship between pile, lime-soil cushion and soil between piles controlling the settlement are concluded and explore the feasibility of the method by analyzing the test data to provide a new method for theoretical research and engineering practice.

Test Design and Method

Test Equipment. The centrifuge is used in this test is L-30 (20g-ton) in Tongji university. The main technical parameters of the equipment include 1.55 m effective

radius, 200g maximum centrifugal acceleration, 415 mm×370 mm×230 mm size of model box. There are percolation holes in the bottom of model box.

Test Design. Centrifugal acceleration: centrifugal acceleration is 100g in these tests, based on considering the main technical parameters of centrifuge, prototype test conditions, measurement accuracy requirements, model material processing and so on.

Test model: test models are made in terms of 1:100 model rate and use the soil in-site to simulate the actual soil in order to assure the physical and mechanical properties of model materials similar to that of the prototype in these tests(Huang and Wang, 1998; Zeng, 2003). The model materials are IV level self-weight collapsible loess which is got from the III level terrace of the Yellow River in Lanzhou. The main indexes of remold collapsible loess are controlled strictly to simulate the undisturbed collapsible loess. The method of centrifugal model test which uses remold collapsible loess instead of undisturbed collapsible loess to research the self-weight loess collapsibility is practicable on the purpose of settlement study (Wang, 2008). The similar material of stable bearing stratum is sand. Main controlling indexes of testing model is shown in Table 1.

The rigid pile and pile-cap which bear vertical load mainly select similar materials in terms of pile compressive stiffness (Zeng, 2003). Meeting the geometric similarity, aluminium tube are as similar materials that simulate the rigid pile in test. The elastic modulus of aluminium is 7.1×10^4 MPa (The elastic modulus of precast reinforced concrete piles is 3.8×10^4 MPa). The piles are driven and neglected deformation. The similar materials of cushion and geogrid are lime-soil and screen netting respectively. Test-time conversion is $t_m=t_p/N^2$ based on Terzaghi one-dimensional consolidation theory considering the laminar flow, consolidation and other issues (Taylor, 1995; Bao and Rao, 1998). The parameters of subgrade, loads of trains and track structure reference the relevant norms.

Table 1. Main controlling indexes of testing model.

Soil	Density ρ	Moisture content ω (%)	Void ration e	Void ratio n	Saturation S_r (%)	Specific density d_s
Undisturbed collapsible loess	1.53	7.7	0.91	0.48	23.0	2.71
Remolded collapsible loess	1.60	19.0	1.02	0.50	50.7	2.71
Subgrade filling	2.08	19.0	0.55	0.35	93.5	2.71

Test Program. The tests are divided into two groups which are the settlement on non-reinforced remold collapsible loess and settlement on piled embankment under the high speed railway. In these tests the height of subgrade is 6.0 m (Prototype value, similarly hereinafter). The thickness of remold collapsible loess and stable bearing stratum are 15 m and 2 m respectively. The trains and track structure loads of the high speed railway are simulated as soil column. The length of pile is 16m, the diameter of pile is 40 cm, the thickness of pile lining is 10cm. The pile spacing (s) are set as groups(s=2D (the diameter of pile), 4D, 6D). The thickness of lime-soil cushion is 1 m. The time of post-construction settlement is from 200th day to 1970th day in consideration of construction period, technical conditions and surcharge preloading, etc. The test model is shown as Figures 1 and 2.

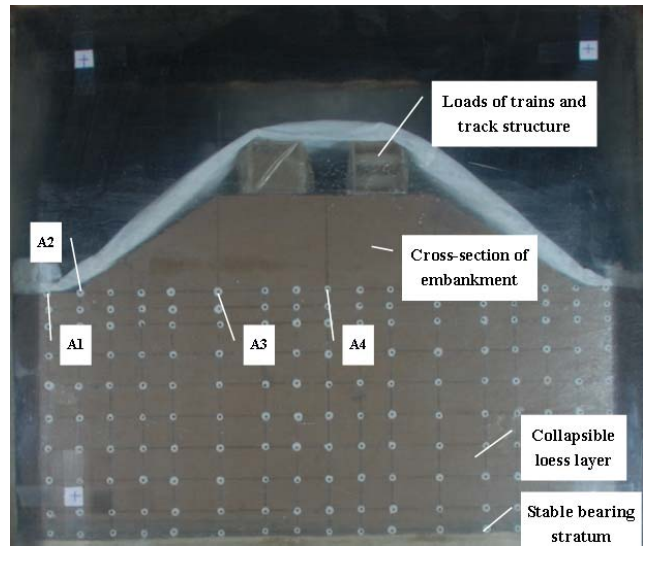


Figure 1. The model of test on non-reinforced remold collapsible loess.

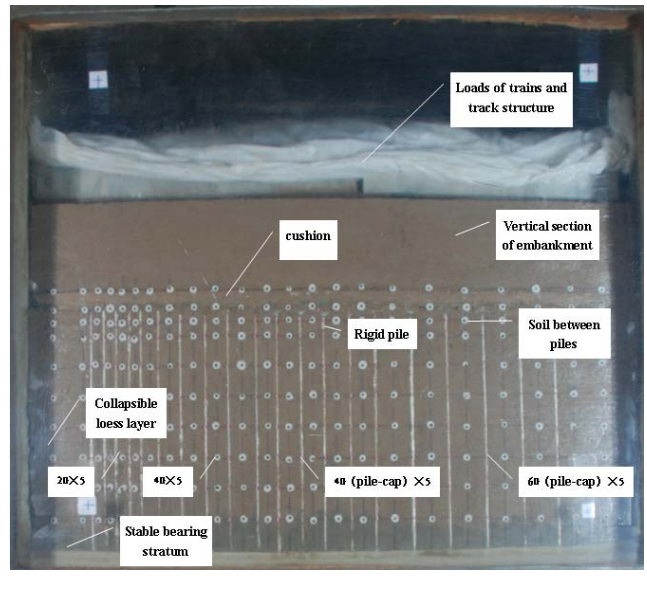
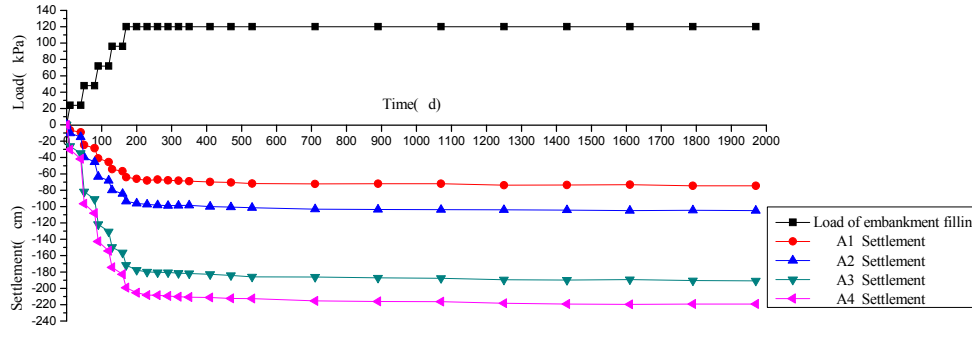
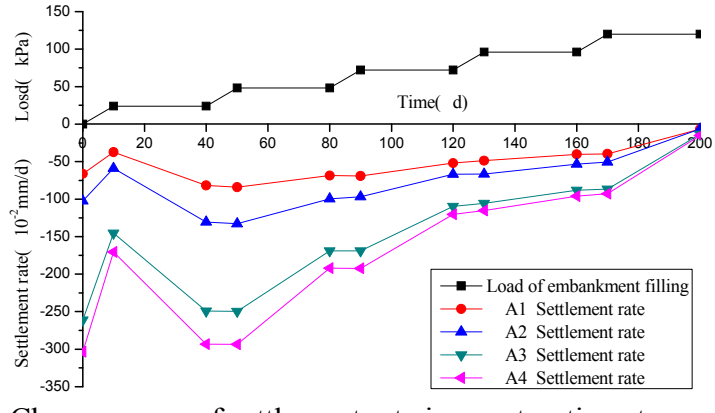


Figure 2. The model of test on piled embankment.

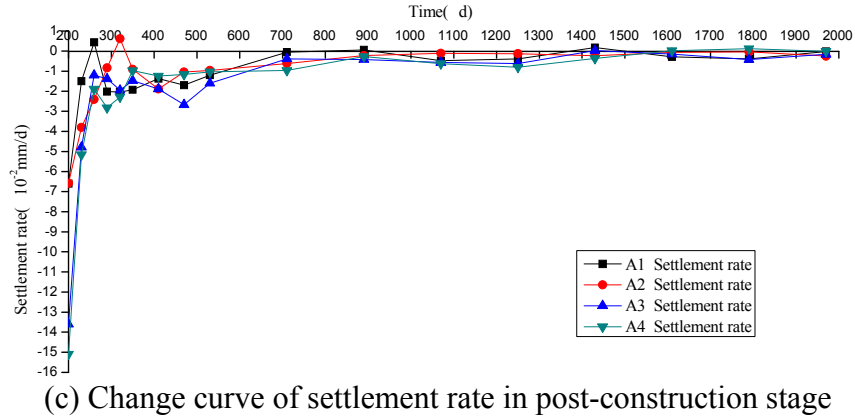
Test Data Acquisition. The boundary effect is binding with model from the wall model box. The vaseline is applied at wall of model boxes to reduce friction in these tests, to ensure the reliability and accuracy of test data. The plexiglass plate witch is removable and transparent is one side of the model box, plastic particles with pin as observation monuments can be set at this side of model according to size of test. The displacement of plastic particles is synchronous with the displacement of soil. Through fixed-point high-speed camera system make the shift imaging of observation monuments at different stages, the displacement and stability of soil can be evaluated and analyzed by computer graphic processing software.



(a) Change curve of settlement



(b) Change curve of settlement rate in construction stage



(c) Change curve of settlement rate in post-construction stage

Figure 3. Change curve of settlement on non-reinforced remold collapsible loess.

Test Process. The test model is processed in model box in terms of the design physical dimension and parameters of model. The model box is fixed in the centrifuge after arranging observation monuments, then installing and checking fixed-point high-speed camera system and video tracking system. The subgrade filling processing is simulated by adjusting the centrifugal acceleration. The soil deformation is observed at times on each 1.2 m (20g acceleration increment) subgrade filling, and lasting 1 month (30 days) before filling nest layer with stable load. The soil deformation is observed timed after centrifugal acceleration reaching 100g gradually. The processing of test last about 5 years (1970 days).

Test Results and Analysis

Analysis of Settlement on Non-Reinforced Remold Collapsible Loess. Change curve of settlement on non-reinforced remold collapsible loess of high speed railway is shown as Figure 3. The foundation deformation which are at 2 m outside toe of fill (A1), toe of fill (A2), under the centerline of road border (A3), under the centerline of subgrsde (A4) are observed respectively.

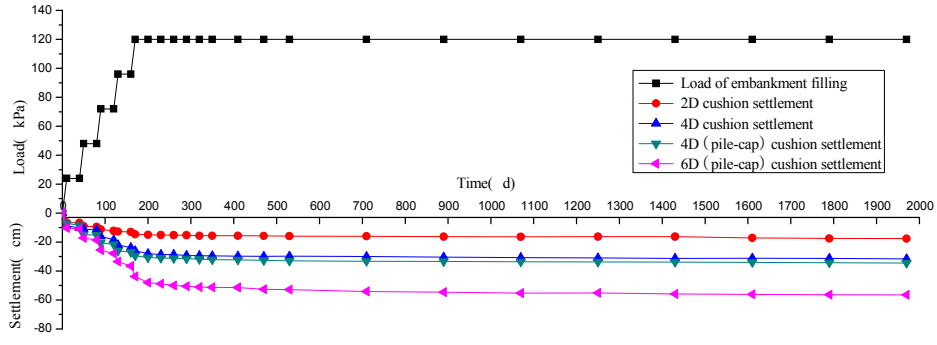
With the increased load of subgrade, the settlement of A1, A2, A3, A4 is 66.2 cm, 96.0 cm, 177.7 cm, 205.5 cm in construction stage. The settlement increase from 2 m outside toe of fill (A1) to the centerline of subgrsde (A4) gradually, and the largest settlement position is at the centerline of subgrsde (A4). The post-construction settlement in each observation monument is shown in Table 2. The post-construction settlement of the centerline of subgrsde (A4) is 13.6 cm, the largest settlement rate is 7.0 cm/year. The untreated ground which can not meet the request on settlement after construction in ballasted track of high speed railway need be reinforced.

Analysis of Settlement on Piled Embankment. The position of the largest total settlement and post-construction settlement is both at the centerline of subgrsde (A4). So vertical section at the centerline of subgrsde (A4) is observed mainly in piled embankment settlement observation, to analyze the influence of settlement with pile spacing changing.

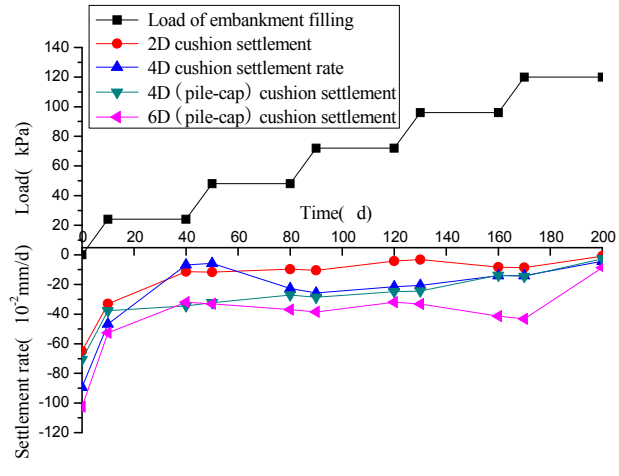
Table 2. Statistical table of post-construction settlement in each observation monument.

Observation monument	Post-construction settlement (cm)	1st year settlement rate (cm/year)	2nd year settlement rate (cm/ year)	3rd year settlement rate (cm/ year)
2 m outside toe of fill (A1)	8.4	5.1	1.9	0.8
Toe of fill (A2)	9.0	5.4	2.2	0.9
Under the centerline of road border (A3)	12.5	6.5	2.9	1.8
Under the centerline of subgrsde (A4)	13.6	7.0	3.5	2.2

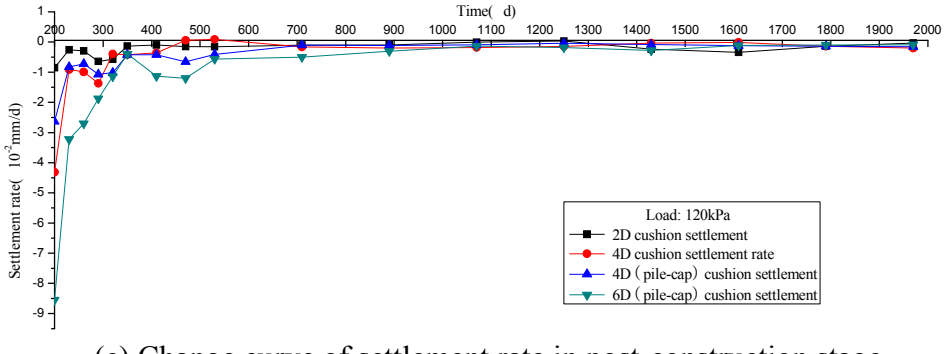
Change curve of cushion settlement in different pile spacing is shown as Fig.4. The settlement of cushion in 2D, 4D, 4D(pile-cap),6D(pile-cap) pile spacing is 15.0 cm, 28.2 cm, 30.8 cm, 48.2cm in construction stage. The settlement is 2.6 cm, 5.1 cm, 3.8 cm, 8.5 cm, and the settlement rate is 0.8 cm/year, 1.7 cm/year, 1.3 cm/year, 4.8 cm/year in post-construction stage respectively. With the pile spacing increased the total settlement, settlement and settlement rate in post-construction get larger significantly.



(a) Change curve of settlement



(b) Change curve of settlement rate in construction stage

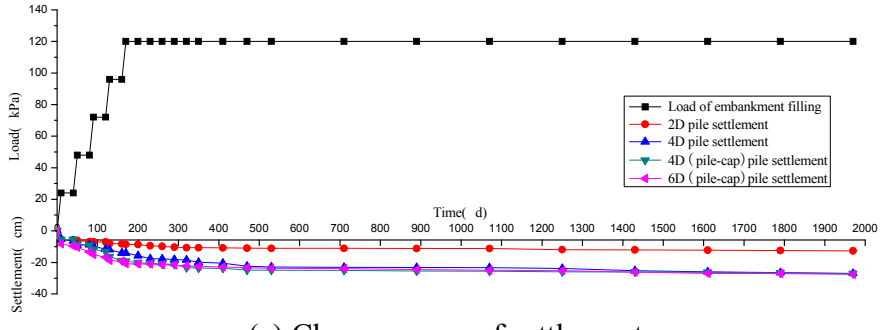


(c) Change curve of settlement rate in post-construction stage

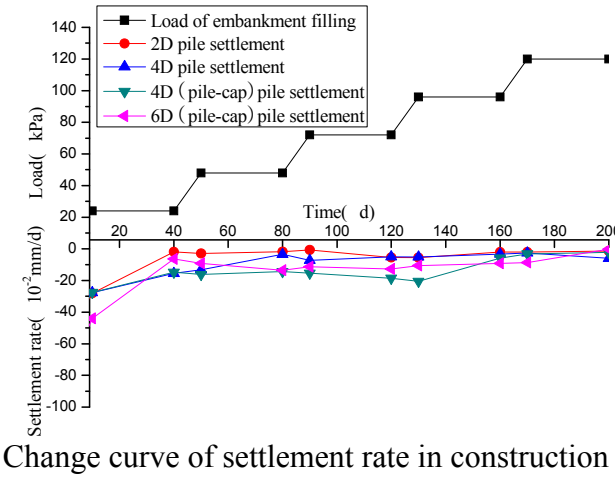
Figure 4. Change curve of cushion settlement in different pile spacing.

Change curve of pile settlement in different pile spacing is shown as Fig.5. The settlement of pile in 2D, 4D, 4D(pile-cap), 6D(pile-cap)pile spacing is 8.7 cm, 15.7 cm, 20.1 cm, 21.0cm in construction stage. The settlement is 4.0 cm, 11.2 cm, 6.9 cm, 6.4 cm in post-construction stage. With the pile spacing increased, the depth of pile piercing in cushion get larger and pile-soil load share ratio get lower.

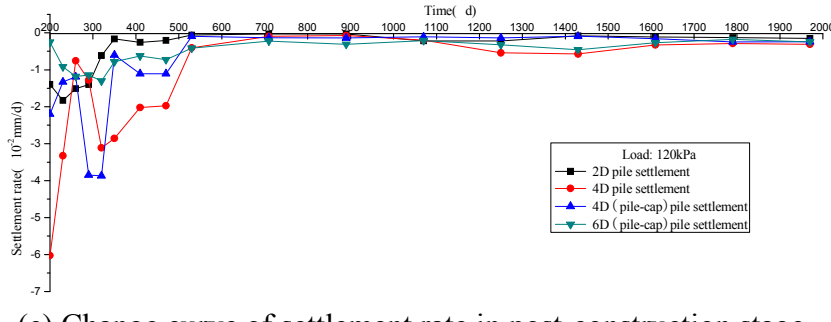
Change curve of soil between piles settlement in different pile spacing is shown as Fig.6. The settlement of soil between piles in 2D, 4D, 4D (pile-cap),6D (pile-cap) pile spacing is 14.7cm， 28.3cm， 28.7cm， 57.9cm in construction stage. The settlement is 3.0cm， 10.5cm， 31.2cm， 11.2cm in post-construction stage.



(a) Change curve of settlement

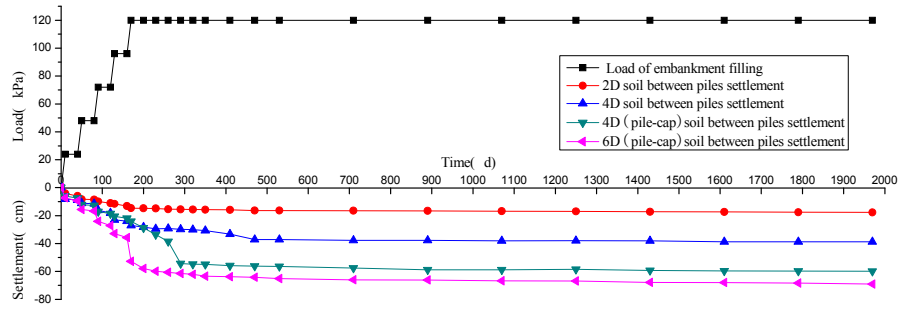


(b) Change curve of settlement rate in construction stage

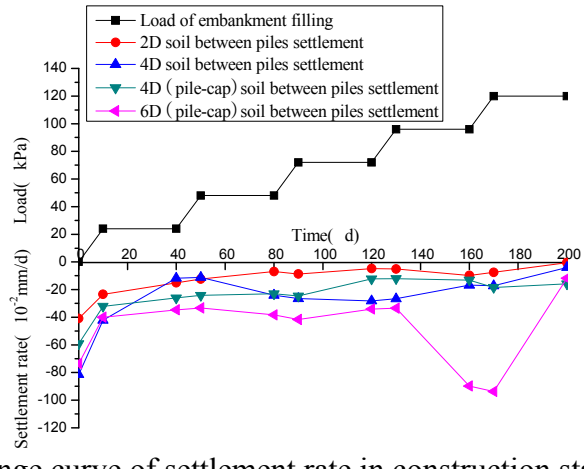


(c) Change curve of settlement rate in post-construction stage

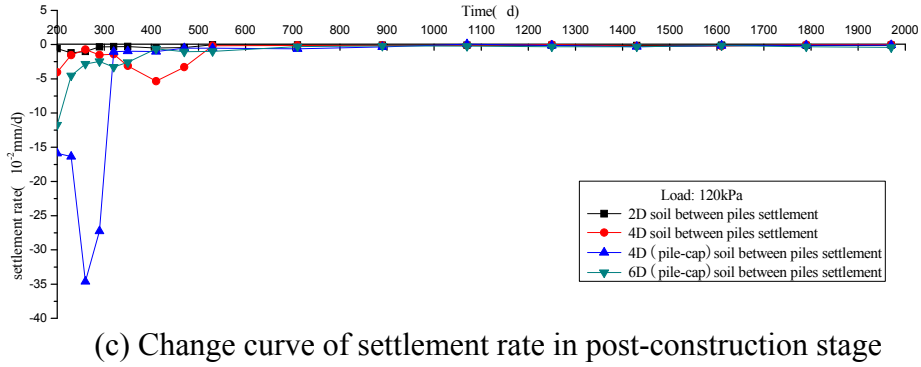
Figure 5. Change curve of pile settlement in different pile spacing.



(a) Change curve of settlement



(b) Change curve of settlement rate in construction stage



(c) Change curve of settlement rate in post-construction stage

Figure 6. Change curve of soil between piles settlement in different pile spacing.

Piled embankment which is a coordination of pile, cushion and soil between piles can bear load effectively. With the larger of pile spacing, the settlement of piled embankment increase and the incremental is larger. The settlement of soil between piles is faster that of pile, and negative skin friction of pile is larger. The pile-soil load share ratio is lower, and the load which is bore by soil between piles is increased.

Conclusions

Draw the conclusion as follows in terms of the centrifuge model tests of piled embankment in high speed railway: The method of centrifugal model test which uses

remold collapsible loess to research on cross section and vertical section of subgrade is practicable on the purpose of settlement study. Untreated ground which can not meet the request on settlement after construction in ballasted track of high speed railway need be reinforced. Piled embankment which is a coordination of pile, raft structure and soil between piles can bear load effectively. The total settlement, settlement after construction and settlement ratio of piled embankment get larger with the extending of pile spacing. With the larger of pile spacing, the settlement of soil is faster and more increasing than that of pile, and pile-soil load share ratio get lower and negative skin friction of pile get more increasing.

Acknowledgement

This research was funded by the Nation High Technology Research and Development Program of China (No. 2007AA11Z116).

References

Wang, B .L. (2007). *High-Speed Railway Subgrade Engineering*, Beijing: China Railway Publishing House.

Qian, H. J., and Luo, Y. S. (1985). *Collapsible Loess Foundation*, Beijing: China Building Industry Press.

Rao, W. G. (2004). *Theory and Practice of Pile-Net Composite Foundation*, Beijing: China waterpower press.

Zai, J. M. (2004). *Theory and Application of Composite Pile Foundation*, Beijing: China Waterpower Press.

Yan, M. L., and Zhang, D. G. (2001).*CFG Pile Composite Foundation Technology and Engineering Practice*, Beijing: China Waterpower Press.

Gong, X. N. (2002). *Theory and Engineering Application of Composite Foundation*, Beijing: China Building Industry Press.

Zeng, Y. J., and Zhang, W. M. (2003). "Reviews of state-of-art of modeling simulation on pile foundations." *Rock and Soil Mechanics*, 674-680.

Huang, Z. Q., and Wang, S. J. (1998). "A Study of application of centrifugal model testing in China". *Chinese Journal of Rock Mechanics and Engineering*, 199-203.

Lee, C. J., and Chen, C. Z. (2002). *Negative Skin Friction on Grouped Piles*.

Wang, C. D. (2008). *The Research on the Settlement Controlling and Geotechnical Centrifuge Tests of Piled Embankment in Collapsible Loess Ground of High-speed Railway*, Master Thesis of Lanzhou Jiaotong University.

Taylor, R N. (1995). *Geotechnical Centrifuge Technology*. Blackie Academic & Professional.

Bao, C. G., Rao, X. B. (1998). "Principle of the geotechnical centrifuge model test." *Journal of Yangtze River Scientific Research Institute*, 2–7.

Determining the Rational Completion Cable Forces Based on Influence Matrix Method United Minimum Bending Energy Method

Jun Yang[1]

[1] Department of Civil Engineering, School of Civil & Hydropower Engineering, China Three Gorges University; School of Civil Engineering, Wuhan University, China; PH: 15971646394; Email: wangjing750301@163.com

Abstract

There are many methods to determine a rational completion state. However, it is difficult to determine the optimal cable tension using just one method, besides, different methods must be adopted for different types of bridges. As a result, it is feasible to combine minimum bending energy with the influence matrix method. The basic approach is: the initial cable tension can be determined through the minimum bending energy method, then the cable tension is adjusted through influence matrix, whose constraint conditions are stresses of key sections and support reaction from the top of pier, at last, the optimal cable tension of the finished state is determined. This method is not only simple but also very clear, and it achieves a good effect in determining the rational completion state of this type of bridges.

Introduction

Cable-stayed bridges are a combination system bridge, composed of beam, tower and cable. Cable-stayed bridge is a high-order hyperstatic structure for the existence of cables and increasing complexity of cable-stayed bridge design calculation. Cable force affects structure stress quietly and is determined by limited tension. For limitations by construction equipment number, cable tension only can be performed

one-by-one or group-by-group. Cable tension affects cable-stayed bridge stress quietly and cable force is a key factor affecting cable-stayed bridge stress. How to determine rational cable force and rational completion status of a cable-stayed bridge is a key problem to design a cable-stayed bridge.

There are many methods to determine rational completion status of a cable-stayed bridge, and each has its merits and defects. On the basis of extensive reading of domestic and foreign documentation, the author summarizes the methods of determination of rational completion status of cable-stayed bridge and analyzes their merits and defects, supplies a method combined bending energy method with influence matrix method.

Requirements of the Rational Completion State and Optimization Principle

Requirements of the Rational Completion State. The purpose of completion state computation is to determine completion state, including the internal force state under dead load. The line shape of girder means the elevation of girder meets the design requirement under dead loads. This becomes a definite objective in the conceptual phase after longitudinal slope of deck and vertical curve are determined. Generally, the rational completion state of cable-stayed bridge should meet requirements as follows:the distribution of cable force is uniform, the bending moments of tower are not great, the stresses of girder should be controlled within the feasible zone, and the reaction forces of side and assist piers are reserved under dead loads. For concrete cable-stayed bridge, the stresses of girder are the most important.

Tthe Principle of Influence Matrix. According to the concept of the generalized influence matrix, multiple objective functions of cable force optimizations are expressed with uniform cable force variant and the generalized influence matrix. Then the purpose that multiple objective functions are optimized with just one method is realized. Because the cable influence matrix that the multiple objective functions correspond can be formed at a time, the computation is convenient through programs. Assuming the linear superposition principle is satisfied, and then the computation of the influence matrix can be given by:

$$[C]\{X\}=\{D\} \tag{1}$$

where $[C]$ is the influence matrix; $\{X\}$ is the active adjustment cable force vector; and $\{D\}$ is the passive adjustment cable force vector.

The passive adjustment cable force vector, active adjustment cable force vector, and influence matrix can be expressed in the following passage.

The passive adjustment cable force vector $\{D\}$ is a column vector made up of independent elements $D_i(i=1,2,\cdots,m)$ corresponding to m designated values of key sections, and it can be written:

$$\{D\}=\{D_1,D_2,\cdots,D_m\}^T \tag{2}$$

The active adjustment cable force vector $\{X\}$ is a column vector made up of independent elements $X_i(i=1,2,\cdots,n)$ corresponding to n designated values that can change the internal force and displacement of key sections, and it can be written:

$$\{X\}=\{X_1,X_2,\cdots X_n\}^T \tag{3}$$

The influence matrix $[C]$ is made up of variation values that unit alteration the n active adjustment cable force vectors lead to the m passive adjustment cable tension vectors, and it can be given by:

$$[C]=\begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{21} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ C_{m1} & C_{m2} & \cdots & C_{mn} \end{bmatrix} \tag{4}$$

The Principle of Minimum Bending Energy Method. This method takes the bending strain energy of a structure as an objective function. If no constraint condition is imposed, this method can change into once structure analysis, which can be performed as follows:the axial stiffiness of girder, tower, and cable takes large value, the bending stiffness of girder and tower is not changed, and all the dead load are imposed on the structure, then the obtained internal force state is the desired. Because this method accounts for the influence of both bending moment and displacement, the bending moment of completion state is relatively small, and the line shape is rational (Xu, 2002).

The cable-stayed bridge is made up of member structures, and the energy of system can be given by:

$$U=\int\frac{M^2ds}{2EI}+\int\frac{N^2ds}{2EA}+k\int\frac{Q^2ds}{2GA} \tag{5}$$

where *M*, *N*, and *Q* are bending moment, axial force, and shear force of members, respectively; *E* and *G* are elasticity modulus and shear modulus of material, respectively; I and A are inertia moment and area of member section, respectively; k is shape coefficient of section.

In Eq. 5. the potential energy caused by shear force Q is much smaller, so it can be neglected generally. Besides, the bending moment is the most important control requirement for girder and tower, and therefore, the Eq. (5). Can be reduced to the

expression of bending energy. For the discrete structure with m elements, assuming the section of each element is equivalent, and the elasticity modulus is invariable, then the expression of bending energy can be expressed as:

$$U = \sum_{i=1}^{m} \frac{l_1}{4E_1 I_1} \left(M_{Li}^2 + M_{Ri}^2 \right) \tag{8}$$

where l_i, E_i, and I_i are length, elastic module and inertia moment of element I respectively: M_{Li} and M_{Ri} are left and right bending moments of element i.

Eq. 7can be written as matrix form:

$$U = \{M\}^T [B] \{M\} \tag{9}$$

where $\{M\}$ is bending moment vector; $[B]$ is diagonal matrix, and the diagonal element is $b_{ii} = \frac{l_i}{4E_i I_i}$.

According to minimum potential principle, Eq. (7) can be written as:

$$\frac{\partial U}{\partial M} = 0 \tag{10}$$

Process of Determining the Rational Completion State

The Cable Force Influence Matrix. In order to determine the initial cable force with a minimum bending energy method, the relation between cable force and bending moment, which is called force influence matrix, needs to be determined first. After the completion state is determined, partial cable force is adjusted based on the control conditions of section stresses and reaction force. Then the cable force influence matrix on section stresses and reaction force needs to be determined, too. The computation process of the cable force influence matrix is as follows:

Build the finite element model of cable-stayed bridge with software ANSYS. Girder and tower are simulated by beam element, while cable is simulated by link element that is tensioned only. Rigid arm is used between cable and tower, cable and girder. Neglect the dead load, and the initial strain of cable is set to be zero, and then the unit alternation of cable force is performed through unit strain of link element. Based on the model that cables are not broken, ANSYS secondary development is carried on, then the influence vector on cables can be computed, subsequently, the influence matrix is obtained. Influence matrixes that bending moment influences cable, stress, and reaction force can be written as follows (Yun et al., 2000):

$$\{M\} = \{M_0\} + [C_M]\{T\} \tag{11}$$

$$\{\sigma\} = \{\sigma_0\} + [C_\sigma]\{T\} \tag{12}$$

$$\{F\} = \{F_0\} + [C_F]\{T\} \tag{13}$$

where $\{M_0\}$, $\{\sigma_0\}$, and $\{F_0\}$ are bending moment, stress, and reaction force vector, respectively; $[C_M]$, $[C_\sigma]$, and $[C_F]$ are influence matrixes that bending moment influences cable, stress, and reaction force, respectively; $\{T\}$ is the active adjustment cable force vector.

Determination of Preliminary Completion State with Minimum Bending Energy Method. Substituting Eqs. (7) for (9), and according to Eq. (8), then Eqs. (12) and (13) can be obtained:

$$U = \{M_0\}^T [B]\{M_0\} + 2\{M_0\}^T [B][C_M]\{T\} + \{T\}^T [C_M]^T [B][C_M]\{T\} \quad (14)$$

$$\frac{\partial U}{\partial T_i} = 0 \quad (15)$$

where T_i is the adjustment value of cable tension.

Substituting Eq. 13 for 12, then Eq. 14 can be obtained:

$$[C_M]^T [B][M_0] + [C_M]^T [B][C_M]\{T\} = 0 \quad (16)$$

According to the minimum bending energy method, the area of girder, tower and cable is magnified by 10^5 times, while the bending stiffness is invariable, then bending energy is left only among structure strain energy, the preliminary initial cable force of the bridge can be obtained under the action of dead load.

Partial Adjustment with Influence Matrix. Stresses of key section, reaction forces of side and assist pier can not be considered through minimum bending energy method only, so the obtained cable force is not rational, and partial cable force is non-uniform, even some cable force is negative. Therefore, the cable force needs to be adjusted furthermore. The influence of section stresses and reaction forces can be dealt with through influence matrix.

According to the design code, the maximum compression stress in the compressive region should satisfy:

$$\sigma_p \le 0.6 f_{ck} \quad (17)$$

where f_{ck} is the standard value of concrete's compression strength.

For full PC bridge, the concrete girder and tower should be compressive in the whole section, and positive reaction forces are necessary in the support.

Example Analysis

Bridge Introduction. A single main tower cable-stayed bridge is located in Changsha city. It is a concrete tower cable-stayed bridge, whose main span is 258.4 m. The main bridge is a PC cable-stayed bridge. Layout of cables is double cable

planes, and two cables are laid in the same location（numbers of side span cables are L01-L16 in turn from inside to outside, and numbers of the main span cables are R01-R16 in turn from inside to outside）. Height of the tower is 21.2 m. General layout of main bridge shows in Figure1 (Xiao).

Buildup of Finite Element Model. Because this bridge is double cable planes, the plane model of member system can be built. In order to analyze the problem in detail, girder and tower are simulated by beam 188, while cable is simulated by link 10.

Initial Cable Force under the Rational Completed State Based on APDL Language. The cable force influence matrixes on section moment, section stress, and reaction force can be computed based on AODL language. At the same time, mechanical state of bridge can be determined through combining ANSYS with APDL language. The process is as follows:Input basic parameters of the bridge, and select nonlinear static analysis; The initial strain of cable is set to be zero., analyze the mechanical state of bridge under the action of dead load, first-phase load and second-phase load; The initial strain of cable is set to be 0.001, and then compute the influence matrixes; Compute the adjusted value of cable force $\{T\}$ based on Eq. (14), and then obtain the finial cable force with influence matrix.

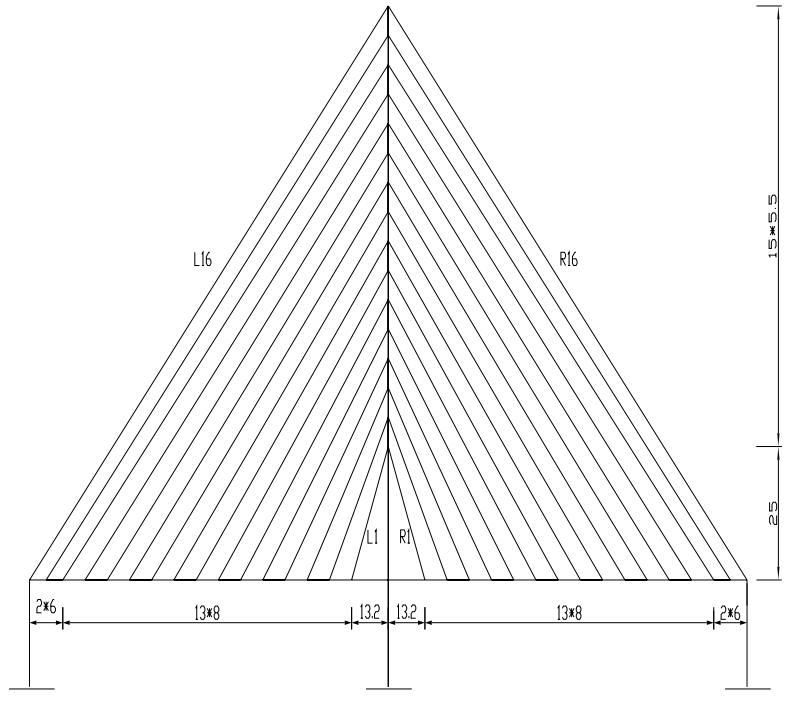


Figure 1. General layout of the bridge.

Table 1. Cable force computed with the method.

Cable number	Cable force（KN）	Cable number	Cable force（KN）
L1	5111.6	R1	5222.0
L2	4012.4	R2	3439.9
L3	3961.8	R3	3882.2
L4	4715.2	R4	3961.7
L5	4512.5	R5	3977.7
L6	4372.0	R6	4055.2
L7	4385.8	R7	4132.1
L8	4245.5	R8	4169.7
L9	4169.7	R9	4245.5
L10	4132.1	R10	4385.8
L11	4055.2	R11	4372.0
L12	3977.7	R12	4512.5
L13	3961.7	R13	4715.2
L14	3882.2	R14	3961.8
L15	3439.9	R15	4012.4
L16	5222.0	R16	5111.6

After the optimization computation, the cable force can be obtained through minimum bending energy method. It is compared with the design value, from these, it is found that some cable force is non-uniform. Besides, stresses of key section force are not considered. Thus, the cable force is not reasonable and it needs to be adjusted. The final results are shown in Table 1. From Table 1, the cable forces are distributed rationally, therefore, the present method is reasonable.

Conclusions

The determination of rational complete state is the precondition of building a cable-stayed bridge. Presently,the common methods are numerous, while the designers adjust the cable force by their experience, so it is a demanding job. For this unique bridge style, this paper presents the method that combines the minimum bending energy with influence matrix. This method is simple, the process is clear, and the cable force is rational. Besides, the ANSYS secondary development is performed in this method, and the cable force optimization efficiency is increased.

References

Xu, D. (2002). "Research on the design of big span steel-concrete composite beam bridge", *Master Degree Thesis of Dalian University of Technology*.

Yan, D., Li, X., Liu, G., and Yi, W.. (2000). "Adopting stress balance method to determine rational finished bridge state for main beam of cable-stayed bridge", *China Journal of Highway and Transpor,* (3), 462-466.

Xiao, R., and Xiang, H.. "Influence matrix method of optimized cable force of cable-stayed bridge", *Journal of Tongji University,* 26(3), 235-240.

Light Polymer Cement Mortar Overlay for Bridge Decks

Xuelian Li[1], Yuliang Chen[2] and Zhigang Zhou[3]

[1] Ph.D. and Lecturer, [3] Ph.D. and Professor, School of Communication and Transportation Engineering, Changsha University of Science and Technology, Chiling Road 45#, Changsha, Hunan, China, 410076; E-mail: lixuelianfj@yahoo.com.cn, zzgcs@yahoo.com.cn
[2] Lecturer, School of Engineering & Architecture, East China Jiaotong University, Nanchang, Jiangxi, China, 330013. PH: (86791) 704-6083; Email: chenyuliangjx@yahoo.com.cn

Abstract

A new kind of light polymer cement mortar (LPCM), cement-based intermediate layer modified by acrylic acid, was presented for a bridge deck, based on the advantages and great achievements of light concrete and polymer modified cement concrete in civil engineering. Also, a number of tests were carried out to investigate the performance of the LPCM, such as flexural strength, compressive strength, fatigue life, characteristics of stress-strain curves, and chloride ion penetration test. The results turn out that the density of mortar was about 1800 kg/m^3, which means it is a light cement mortar. Acrylic acid does help to decrease the water-cement ratio, improve paste working property, increase flexibility and crack resistance of the mortar, but there is a best dosage of the acrylic acid. Mechanism analysis shows that high-tensile strength of the polymer membrane is the main factor to significantly improve the mechanical properties of the LPCM. It means that light polymer mortar is an excellent flexible cement material. And it can be used at some special places where the thickness, weight, and flexibility are limited, just like bridge deck pavement.
Key Words: light concrete; polymer mortar; fatigue property; flexural toughness

Introduction

Most of the bridge deck pavement is asphalt concrete, however, early failure is very common in such concrete, for the stress and deformation of bridge decks is very complicated. Many researchers found the strength, bonding, and durability of cement

concrete could be improved after some organic macromolecule polymer was added while the cost could be reduced. The existing construction technology and process of cement concrete fits perfectly to polymer modified concrete (PMC), and it has become a new trend to improve the performance of concrete (Liang 1995; 0'Connor and Saiidi 1993). On the other side, Wang (2005) found light concrete (LC) has many advantages, such as high strength, high specific intensity, low modulus, good durability, and volume stability. Given the unique advantage and great achievements of PMC and LC in civil engineering, a new kind of light polymer cement mortar (LPCM), cement-based intermediate layer modified by acrylic acid, was presented for bridge decks, with reasonable selection and mixture design. And a number of tests were carried out to investigate the performance of the LPCM in this study.

Concrete Design

Tested materials **.**(1) Moderate-heat portland cement was provided by Hunan Shimen Special Cement Company.

(2) Anti-cracking agent was provided by Hunan Study and Design Institute of Building Material, to avoid cracking.

(3) The particle size of ceramisite agitator is from 0.6 mm to 2.36 mm, and its density is no more than 760 kg/m^3, light ceramisite agitator.

(4) Acrylic acid is a polymer material, milky white liquid, produced by the National Starch and Chemical Company. Table 1 summarizes the property of the acrylic acid. Defoaming agent is white water-soluble powder, and its defoaming effect can help the acrylic acid to play fully, provided by Shenzhen Haichuan Chemical Technology Company.

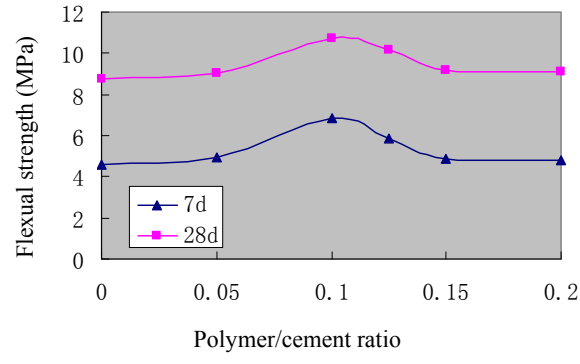
Mixture Design. If the dose of polymer is too small, the concrete cannot form a continuous polymer membrane. With an increased dose, the comprehensive performance of concrete improves. However, the compression strength of polymer modified concrete would decrease, the flexural tensile strength and splitting strength increases very slowly, and the cost will increase (GB/T 2419 2005). So the polymer cement ratio is from 0 to 0.2. The water content was regulated by cement mortar fluidity according to the flow table test (DL/T 5126 2001). The specific maintain method and vibration time is shown in Table 2.

Table 1. Property of the acrylic acid

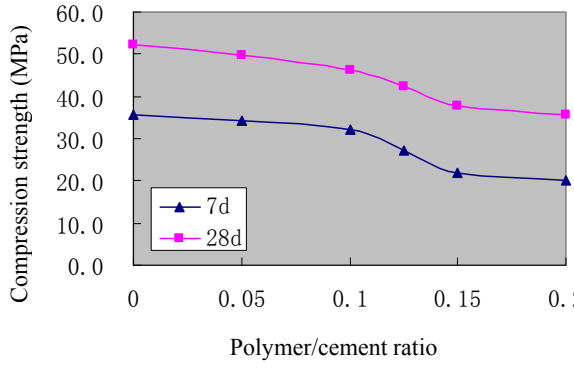
Solid content (%)	Particle size (μm)	Viscosity (Brookfield viscometer, mPa.s)	PH value	Film-forming temperature (°C)	Glass change temperature (°C)
50	0.1	1000-4000	7.0-9.0	16	25

Table 2. Manufacture of LPCM.

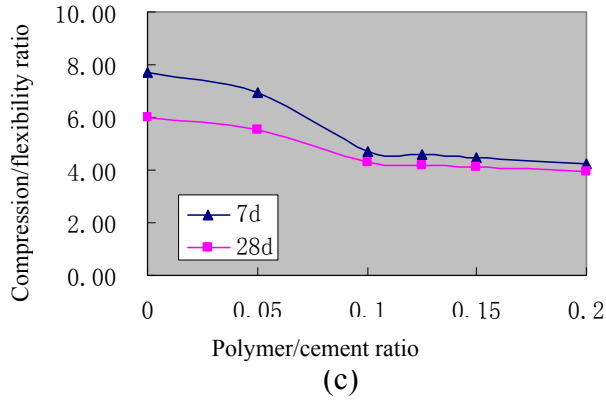
P/C	W/C	Cement: ceramisite agitator	Curing method	Defoaming agent (%)	Vibration time (s)
0	0.38	1:1.5	7d standard curing, 7d-28d dry curing	0	120
0.05	0.31			1	
0.1	0.28			1	
0.125	0.27			1	
0.15	0.26			1	
0.2	0.24			1	



(a)



(b)



(c)

Fugure 1. Correlation between mechanical index and polymer/cement ratio.

Table 2 shows that cement-sand ratio, 1:1.5, is so large as to improve the uniformity and avoid heat accumulation of the mortar. The polymer fluid helps to decrease the water-cement ratio. In addition, the density of 0.1-polymer/cement-ratio cement mortar is 1800 kg/m^3, detected by volumetric flask method, which means it is light concrete (You et al. 2002).

Pavement Performance

Flexural Tensile Strength And. The 160 (length) x 40 (width) x 40 (height) mm samples were tested at the ages of 7 and 28 days. The loading velocity was 50±5N/s, and the results are shown in Figure 1.

In Figure 1, it can be seen that: (1) the flexural strength and compression strength of LPCM increases with age. When polymer/cement ratio is larger than 0.1, the rate of increase of flexural strength of LPCM is higher than that of others. (2) When the dose of acrylic acid is less, its effect on mortar is indistinct. When it is raised to some quantum, the flexual strength and flexible property of the mortar is clearly higher. It is not better to have more acrylic acid. There is optimal mixing amount of acrylicratio rises, for the polymer made the density of mortar drops. (4) The compression/flexibility ratio monotonically decreases with age, which is the same with the common mortar.

Fatigue Property

The fatigue tests were carried out on the JC95-2 gravity flexural fatigue test machine. The test temperature was 20℃. The samples were kept in the environment box for 12 hours before the tests to make sure the internal and external temperatures were the same. The results are presented in Table 3.
Table 3 shows the fatigue lives of sample #3 of 0.1-polymer/cement ratio and sample #1 of 0.125-polymer/cement ratio are shorter, and there were more pores in the broken section of these two samples, which means there are more defects in these two samples. And the fatigue lives of the others are all more than 5000 000 times. Given the reason of cost, the ideal polymer/cement ratio is 0.1.

Modulus and Ultimate Bending Extensibility Stress-strain curve of 0.1-polymer/cement ratio was tested on the hydraulic universal testing machine and strain gauge, and the resolution rate was 5×10^{-6}. The size of the sample was 230 (length) x 50 (width) x 30 (height) mm, and the effective span length was 180 mm. The load was added at the midspan, and the loading rate was 100 N/min.

Table 3. Results of fatigue tests (0.6 stress level).

P/C		0.1	0.125	0.15
Fatigue life ($\times10^6$ times)	1#	5.12	1.43	5.11
	2#	5.13	-	-
	3#	1.57	-	-

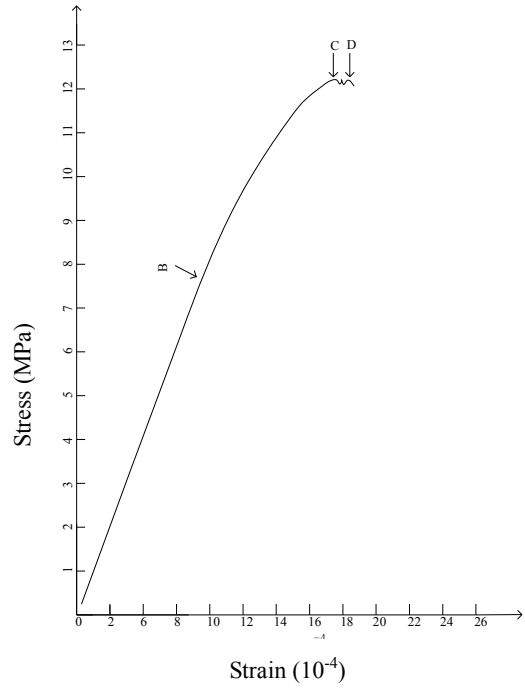


Figure 2. Stress-strain curve.

Table 4. Characterestic data of the stress-strain curve.

Modulus (GPa)	Stress scope in elastic stage (MPa)	Strain scope in elastic stage (με)	Flexural strength (MPa)	Ultimate flexural strain limit (με)
9.61	0-7.4	0-770	11.9	1625

Figure 2 and Table 4 show that the stress-strain curve is different from the common mortar: (1) the curve segment is longer and the curvature is bigger. (2) There is yielding segment, just like metal material. (3) The modulus, 9.61 GPa, is far less than the common one. (4) The ultimate flexural strain limit and flexural strength is larger than the common mortar. All these mean the flexural property of LPCM is better.

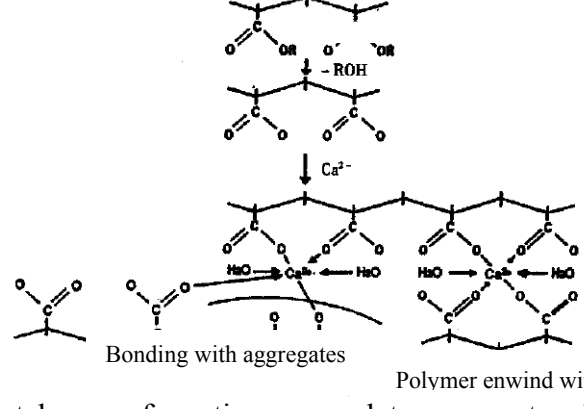


Figure 3. Sketch map of reaction among latex, cement and aggregate.

Mechanism Analysis

The stickiness and flocculation of the acrylic acid is very good. The side group on the molecular chain length is lively amido, which can react chemically. Then the acrylic acid latex is added into the light mortar, fully stirred, the latex can spread around uniformly, being a cement-based compound material. The lively group in the latex can react mutually with $Ca2^{+}$, $Al3^{+}$in the hydration products of cement, which means that polymer and hydration products of cement enwind with each other and wrap the aggregate. Then the condition of the interface transition area is changed radically (Figure 3), and the flexility and the fatigue property of LPCM were greatly improved.

Conclusions

The acrylic acid modified mortar presented in this study is light concrete, for the density was 1800 kg/m^3. The optimal amount of acrylic acid latex is 0.1, and the latex in the modified mortar helps to raise the flexibility and the anti-cracking ability of the mortar. The mechanism analysis shows that the polymer menbrane in the LPCM is the fundamental reason for improved performance.

References

0'Connor, D. N., and Saiidi, M. (1993). “Polyester concrete for bridge deck overlays.” *Concrete international*, 36-39.

Beeldens, A. (2002). “Influence of polymer modification on the behavior of concrete under severe conditions.” *Ph.D. dissertation*, Faculty of Engineering, Katholieke Universiteit Leuven, 248.

DL/T 5126-2001. “Test specifications of polymer modified cement mortar.” Beijing: *Standards Press of China*.

GB/T 2419-2000. “Test method of fluidity of the cement mortar.” Beijing: *Standards Press of China*.

Liang, N. (1995). “Polymer modified cement concrete.” Beijing: *People's Transportation Press*.

Mirza, J., Mirza, M.S., and Lapointe, R. (2002). “Laboratory and field performance of polymer-modified cement based repair mortars in cold climates.” *Construction and Building Materials*, 16, 365-374.

Ohama, Y. (1995). “Handbook of polymer-modified concrete and mortars, properties and process technology.” *Noyes Publications*, 236.

Ru, K., Zhang, D., and Song, J. M. (2002). “Properties of polymer modified

cement mortar using pre-enveloping method." *Cement and Concrete research*, 32, 425-429.

Toutou, Z., Lanos, C., and Mélinge, Y. (2004). "The use of squeezing test-extrusion flow similitude to identify cement based materials extrudability." *Proc XIVth Int Congr on Rheology*, Seoul.

Wang, W. (2005). "Test study on granule comcrete and steel bar bonding and anchoring performance." Nanning: *Guangxi University*: 1-3.

You, Y., Liao, C., and Liu, J. (2002). "Study on JM2000 new concrete aAir-entraining agent (40)." *Chemical Building Material*: 39-41.

Numerical Simulation and Experimental Research of Bridge Damage Identification

Alan Jiang[1]

[1] Associate Professor, School of Civil and Safety Engineering,Dalian Jiaotong University,Dalian,116028,P.R.China, E-mail:jiangalan@163.com

Abstract

The paper mainly studies Yin He bridges' damage detection.This bridge was completed in 2000.In the morning of December 13 2001, a truck are bumped against #8 suspension rod of the bridge.Because the main loading consists of a rib arch and suspension rod, the paper analyzes L/2 across, L/4 across of the rib arch, and bumped #8 suspension rod different damage degree's dynamic characteristics. It concludes that the changing regularization of the whole structure's dynamic characteristic made it possible forthe damage detection to the rib arch according to the dynamic characteristic changing and strain modal.

Key Words:bridges; dynamic characteristics; damage detection; finite element method; modal analysis; numerical simulation

Introduction

Yin He Bridge is located in Chifeng City. The bridge was constructed of a reinforced concrete arch rib-bearing box by two side span and a mid-span. Across the bridge that is arranged as 70 +160 +70 (m), its main load-bearing components are two separate ribs.Using C30-class waterproof concrete arch ribs, the main axis of arch-type for the variable cross-section arch catenary, arch-axis coefficient of m=1.543, and arch-thick coefficient n=0.8. Geometric characteristics of the value of the main arch rib are: net span l_0 = 160 m,calculations span l = 163.214 m,the net vector of high f0 = 4 0m,calculation of high vector f = 40.296 m,the main arch rib width b = 2.00 m, vault thickness ads = 3.60 m, arch foot thickness dk = 4.403 m,and arch axis length s = 187.9 m. In order to facilitate construction of the rib cavity, dimensions remain unchanged, with 1.50 × 3.00 m, rib angle 0.5 × 0.5 m, box wall thickness of 0.25 m, tank top, and base plates for variable thickness.

Carriageway between the arches are deployed in Liang Lei and arranged on the outside of the sidewalk suspension rod. The rod using high-strength steel wire produced 109φ5, $\Delta l = 6.800554$ m, the side arch $\Delta l' = 5.977$ m. Width: overall width 24.5 m, or 14 m (vehicle road) +2 × 1.8 m (non-motor vehicle road)+2×0.2 m (trabecula) +2 × 3.25 m (sidewalk).Figure 1 shows the overall layout of the bridge.

Yin He Bridge Structural Dynamic Checking

Since the main arch span is a relatively large force, and the structure is more complicated, this article analyzes only the main arch. In order to accurately solve the dynamic characteristics of the Yin He Bridge, a three-dimensional model of the structure was simulated. A computing model simulation focused on the structure, stiffness, quality, and boundary conditions simulation and should be as much as possible consistent with the actual structure. Calculating arch ribs, beams, cross supports, arches on the columns with the three-dimensional physical space beam element (beam 4 unit) helps to simulate the suspension rod using a two-node bar element (Link 10 unit) to simulate the deck system using a four node shell element (shell 63 unit) to simulate. Table 1 shows the natural vibration frequencies and mode shapes description.

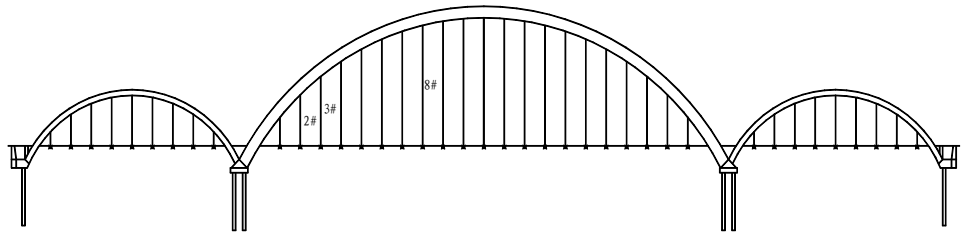


Figure 1. Yin He Bridge Overall layout.

Table 1. Yin He Bridge dynamic characteristics.

Order times	Frequency (H_z)	Direction	Modal description
1	0.712	Horizontal	Rib-plane bending vibration, the two arch ribs simultaneously.
2	0.844	Horizontal	Rib-plane bending vibration, two rib asynchronous.
3	1.354	Vertical	Arch rib, deck-plane bending vibration, two rib synchronization.
4	1.397	Horizontal	Rib-plane bending vibration, the two arch ribs simultaneously.
5	1.854	Vertical	Deck-plane bending vibration.

Yin He Bridge Dynamic Load Test

The dynamic load test of the bridge is shown in Tables 2 and 3.

Table 2. Yin He bridges' a first-order natural frequency and damping ratio.

Item		A first-order natural frequency	Damping ratio
Cross the main arch	Vertical	1.453Hz	0.048
	Horizontal	1.531Hz	0.026
The main arch L / 4 section	Vertical	1.453Hz	0.035
Arches across the side	Vertical	2.718Hz	0.045

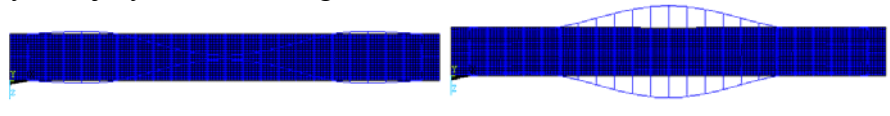
Table 3. Impact factor.

Speed v (km/h)	5	10	15	20	30	35	40
Impact factor of the main arch $1+\mu$	1.005	1.101	1.135	1.125	1.163	1.062	1.056

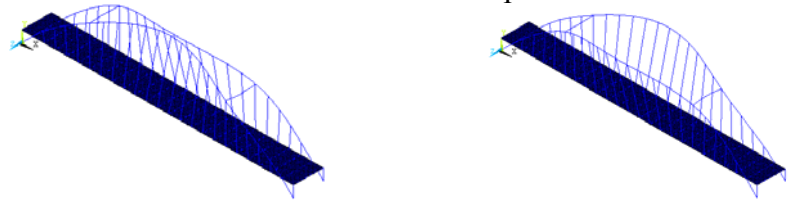
The Dynamic Characteristics of Bridge Structures Based on Damage identification Simulation

In order to obtain information about damage on the bridge structure dynamic characteristics of the impact of variable-stiffness method in this section, namely, by changing the stiffness after the units to the calculation, together with other normal cells. Using the Saint-Venant principle, the damage will only affect the appearance of cells near the part, while the distance can be neglected, due to the Yin He Bridge. This article only simulates the rib, respectively, and suspension rod damage, rib and in selected cross-L/4 cross-sections suspension rod is selected for this bridge had been hit #8 suspension rod to analysis.

Rib Cross-sectional Analysis of the Damage. Simulation of cross-sectional element stiffness damage in 10% and 20% of the two conditions shows that the stiffness (EI) is reduced by 10% and 20%. Use of the large general-purpose finite element program ANSYS5.7 calculates structural damage and injury and is not a self-vibration frequency of injury. Shown in Figure 2.



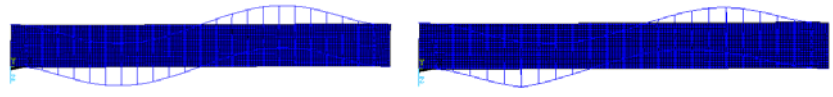
(a) No damage, a single mid-span arch rib injury in 10% of second-order vibration mode vertical view comparison



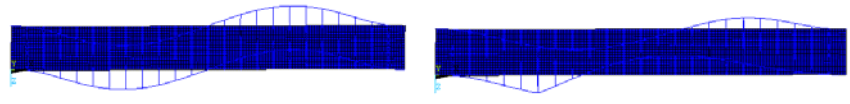
(b) No damage, a single mid-span arch rib injury in 20% of second-order vibration mode three-dimensional map comparison

Figure 2. No damage, a single mid-span arch rib injury in vibration-based map comparison.

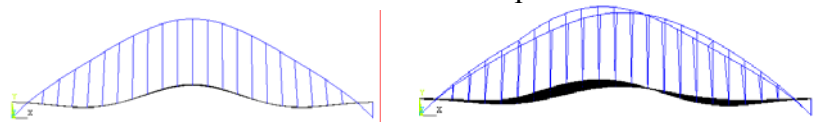
Rib L / 4 Section of the Damage Analysis. Simulation of rib L / 4 section element stiffness damage in 30% and 50% of the two conditions shows that the stiffness (EI) is reduced by 30% and 50%. Computational structural damage and injury is not a self-vibration frequency and did not change significantly. The corresponding change in mode shape is shown in Figure 3. (Space is limited, so this article only shows part of the graph.)



(a) No damage, a single rib L / 4 section of injury 30% of a first-order vibration mode vertical view comparison



(b) No damage, a single rib L / 4 section of injury 50% of a first-order vibration mode vertical view comparison



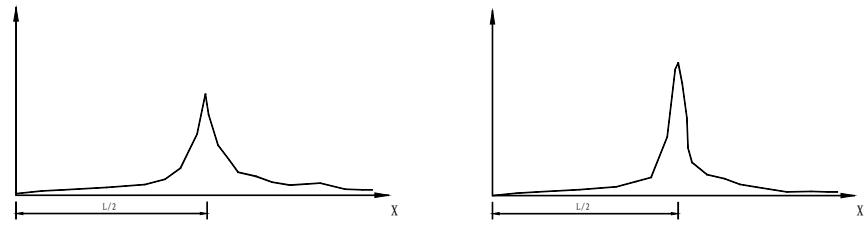
(c) No damage, a single rib L / 4 section of injury 50% of third-order vibration mode front view comparison

Figure 3. No damage, a single rib L/4 section injury in vibration-based map comparison.

With the Strain Modal to Rib for Damage Identification. Rib arch bridge is the key of the whole force area, and its damage will have a direct bearing on the stability and security of the entire arch bridge. The above analysis shows that the decline in local stiffness of the arch ribs will not make the whole arch of the natural frequency of significant changes take place, while the strain mode can detect the structure of localized damage, particularly those for structural vibration parameters is not sensitive. The structure of local damage and the damaged region internal force redistribution occurs, the corresponding stress and strain can be clearly reflected by this change, and the closer from the injury site, the greater the change.

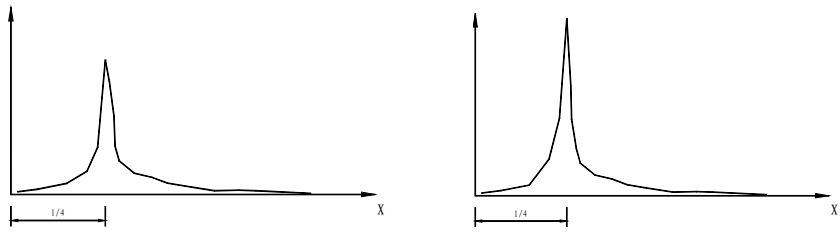
Based on the unit to calculate the bending strain mode, analog unit injury, still does not take into account changes in the quality unit will be attributed to the degradation of stiffness modulus of elasticity decrease. The results are shown in Figures 4 and 5.

From Figure 4, Figure 5 shows that structural damage before and after a first-order strain mode of the relative change in the volume of injuries occurring at the "peak" and a positive correlation with the degree of injury can be used to rib damage location and extent of recognition.



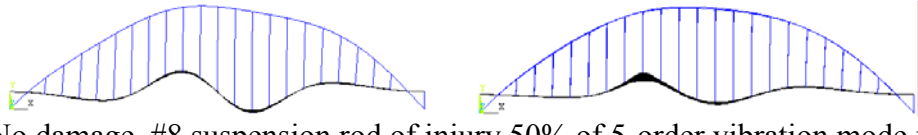
(a) Element stiffness damage 10% (b) Element stiffness damage 20%

Figure 4. Cross-sectional rib injury before and after the first-order change in the relative amount of strain mode.

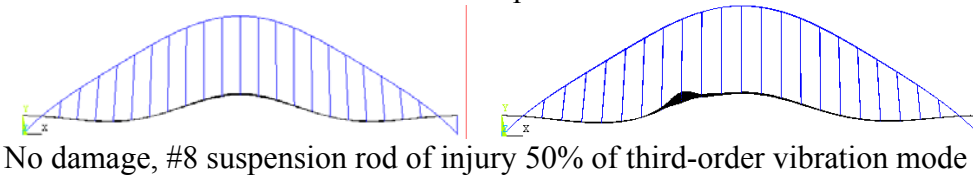


(a) Element stiffness damage 30% (b) Element stiffness damage 50%

Figure 5. Rib L / 4 section injury before and after the first-order change in the relative amount of strain mode.



(a) No damage, #8 suspension rod of injury 50% of 5-order vibration mode front view comparison



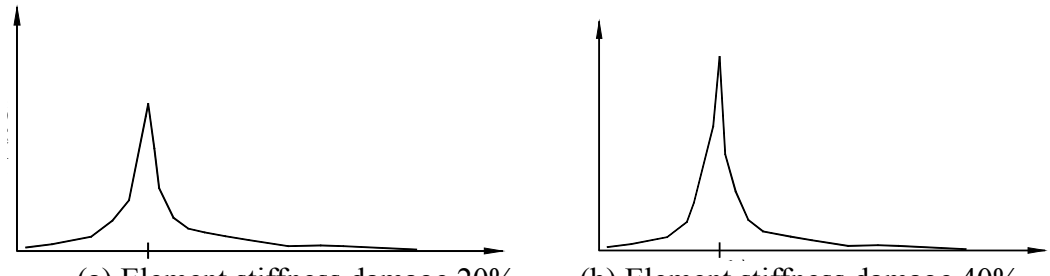
(b) No damage, #8 suspension rod of injury 50% of third-order vibration mode front view comparison

Figure 6. No damage, #8 suspension rod a single rib L / 4 section injury in vibration-based map comparison.

#8 Suspension Rod of the Damage Analysis. Analog #8 suspension rod stiffness (EI) reduced by 20% and 40%. Self-vibration frequency is still no significant change, the corresponding changes in mode shape shown in Figure 6.

Difference with the Curvature of the Suspension Rod Mode for Damage Identification. Half-through the arch bridge rib the root system can be seen to anchor the suspension rod in a multi-span continuous beam. Suspension rod of the damage can be seen as a decline in the bearing stiffness, thus effectively changing the bridge

vibration system node type. Structural units after injury change in vibration mode to reflect the local changes in the structure, which is shown in Figure 6. Modal curvature difference analysis with the department of injury before and after the bridge vibration mode calculation results shown in Figure 7, found that damage cells at the margin normalized curvature mode obviously a "peak," can be used to suspend rod damage location and extent of recognition.



(a) Element stiffness damage 20% (b) Element stiffness damage 40%

Figure 7. #8 suspension rod injury before and after the first-order modal curvature difference.

Conclusions

- The finite element method for a large-span arch bridge in the rib-bearing properties to conduct dynamic analysis of calculation results coincide with the measured data, show that the established theoretical model is correct. Yin He Bridge in-plane vibrational frequencies have a larger in-plane vibrational frequency. The bridge is a more flexible structure.
- Numerical simulation of bridge damage identification study found the following two points:1) rib modal frequency shift in the changes is apparent in many locations and different injury does not exist between the same order of the changes in vibration mode, so changes in vibration mode can be used to determine the damage location; injury before and after a first-order strain. The relative changes in modal volume of injuries occurring at the "peak" and a positive correlation with the degree of injury can be used to locate rib damage and the extent of recognition. 2) Modal curvature difference analyzess using the bridge mode shape before and after the department of injury and found that cell damage is also apparent at "peak" and can be used to suspend rod damage location and extent of recognition.

References

Hassiotis. (2000). "Identification of damage using natural frequencies and Markoy parameters," *Computers and Structures*, 365-373.

Ratcliffe, C. P. .A. (2000). "Frequency and curvature based experimental method for locating damage in structures," *Journal of Vibration and Acoustics*, 324-329.

Abdalla, M. O., and Grigoriadis, K. M. (2000). "Structural damage detection using

linear matrix inequality methods," *Journal of Vibration and Acoustics,* 448-453.

Fryba, L., and Pirner, M. (2001). "Load tests and modal analysis of bridges," *Engineering Structures,* 102-109.

Liu, C., Huang, D., and Wang, T. L. (2002). "Analytical dynamic impact study based on correlated road roughness," *Computers and Structures,* 1639-1650

Song, H. W., You, D. W., Byun, K.-J., and Maekawa, K. (2002). "Finite element failure analysis of reinforced concrete T-girder bridges," *Engineering Structures,* 151-162.

Li, Z. X., Chan, T. H. T., and Ko, J. M. (2001). "Fatigue damage model for bridge under traffic loading: application made to Tsing Ma bridge," *Theoretical and Applied Fracture Mechanics,* 81-91.

Zhu, X. Q., and Law, S. S. (2002). "Practical aspects in moving load identification," *Journal of Sound and Vibration,* 123-146.

Estimation of Average Vehicle Time Headway for Light Traffic

Baibing Li[1]

[1] Business School, Loughborough University, Loughborough LE11 3TU, United Kingdom, Email: b.li2@lboro.ac.uk

Abstract

Vehicle time headway is one of the most fundamental microscopic characteristics of traffic flow. It affects roadway safety, capacity, and level of service. An estimated average time headway can provide an input to intelligent transportation systems for the online monitoring of roadway safety. Single inductive loop detectors are widely installed devices in road networks to regularly measure traffic volume and occupation. The purpose of this paper is to investigate the recursive estimation of average headway using the information from a single loop detector in the scenario of light traffic (e.g. during early morning or late evening period for most of the sections of motorways). A set of recursive estimation equations is developed with a nice and neat form using Bayesian analysis. The incurred computational cost for updating the estimate of average headway is kept to be a minimum. The developed method is illustrated using real traffic data.

Key Words: intelligent transportation system; microscopic flow characteristic; single loop detector; vehicle time headway

Introduction

Estimation of time headway has been an important research topic in traffic studies. Due to the difficulty of data collection, however, most existing studies on vehicle time headway have focused on static analysis. Piao and McDonald (2003) and Marsden et al. (2003) have recently designed a new mobile device for data collection so that differences in driving behavior can be assessed. However, because of the cost of both the device and manpower, this approach cannot be widely used for routine data collection in road networks. Zhang et al. (2007) used the data collected from a dual inductive loop detector to analyze vehicle time headway. Dual loop detectors can provide information on traffic flow on a vehicle-by-vehicle basis, so that vehicle time headway can be estimated with a high accuracy and a minimum manpower cost.

However, this approach has a limitation: the devices for data collection, i.e., dual loop detectors, are neither mobile nor widely installed in road networks so the information on vehicle time headway is available only for the locations where dual loop detectors are installed. Recently, Dailey and Wall (2005) in their pioneering research have suggested estimating the vehicle time headway using single inductive loop detector (ILD) data. Because single ILDs are installed in most sections of strategic motorway networks, this approach makes the estimation of time headway widely feasible.

This paper focuses on the scenario of light traffic which is typically the case during early morning or late evening period for most of the sections of motorways. The aim of this paper is to improve on the estimator developed by Dailey and Wall (2005) by addressing the following issues. First, a recursive approach for light traffic scenario will be developed to accommodate the nature of traffic data that are collected online via a single ILD. In contrast to the estimator of Dailey and Wall (2005) that depends only on the single observation collected in the current time interval, the estimate of the average headway parameter in this paper is updated recursively through a Bayesian analysis, where the current estimate is a weighted average of the previous estimate and the current observation. Since much more information is used in the estimation, the quality of the estimate is greatly improved. Secondly, as well as a point estimate of average headway, we also provide a measure of uncertainty by calculating the associated Bayesian credible interval and the posterior variance.

The paper is structured as follows. Next section is devoted to problem formulation. Then a set of recursive formulae is derived via Bayesian analysis. Next, the developed method is illustrated using real traffic data. Finally, concluding remarks are offered.

Problem formulation

Time headway is the elapsed time between the front of the lead vehicle passing a point on the roadway and the front of the following vehicle passing the same point. In this section we consider estimation of time headway using data from a single ILD.

Data from a single ILD. Consider a single ILD that measures traffic flow during a time period that is split into a number of successive time intervals, each having a duration of T (typically 20 to 30 seconds). During each time interval k, data measured by the single ILD include (see, e.g., Dailey, 1999; Li, 2009): (a) traffic volume m_k defined to be the number of vehicles entering the front edge of the ILD in time interval k; and (b) occupancy O_k defined to be the percentage of time that the ILD is occupied in time interval k. Because of the nature of single ILDs, data provided by a single loop detector are not available on a vehicle-by-vehicle basis. Rather, measurements are available only at an aggregate level of each time interval.

In this paper, rather than investigating vehicle time headway using occupancy data O_k as Dailey and Wall (2005) did, we focus on traffic volume m_k. In traffic engineering, much attention has been paid to traffic volume and as we shall see later, this enables us to investigate time headway more straightforwardly.

A statistical model. Since ILD data are available at an aggregate level, the statistical analysis in this paper will be carried out at the aggregate level as well and the parameter of interest is the average time headway at this aggregate level. Consequently it is no longer possible to detect if vehicles under investigation are in a free-flowing scenario or in a car-following scenario.

When traffic is light and there are no disturbing factors such as traffic signals, the behavior of traffic is essentially random. Under this assumption, the most commonly used model for traffic volume is a Poisson distribution (see, e.g., Gerlough and Huber, 1975):

$$m_k \mid \lambda \sim Poisson(\lambda T) \qquad \text{with} \quad \lambda > 0 \tag{1}$$

where T is the duration of each time interval of a single ILD and λ is the average rate of arrival. $Poisson(\lambda T)$ denotes a Poisson distribution with probability mass function $p(m_k;\lambda) = (\lambda T)^{m_k} \exp(-\lambda T)/m_k!$. The reciprocal of the average rate of arrival λ, $\tau = 1/\lambda$, is the mean time headway, which is the parameter of interest in this paper.

Recursive estimation

In this section, we consider statistical inference for the time headway parameter when traffic is light. It is assumed that there is no disturbing factor so that the traffic volume follows a Poisson distribution (1) with probability mass function $p(m_k;\tau^{-1}) = (T/\tau)^{m_k} \exp(-T/\tau)/m_k!$. We shall perform Bayesian analysis for the vehicle time headway parameter τ to pool the current observation on traffic volume with those obtained in previous time intervals.

Bayesian inference. We start with the first time interval $k = 1$. It is assumed that during this time interval, there is no any prior knowledge about the time headway parameter τ. Hence, a non-informative prior is used so that the subsequent inference is unaffected by information external to the current data. The non-informative prior is chosen using Jeffrey's principle. For the Poisson distribution $p(m_k;\tau^{-1})$, Jeffrey's principle leads to the following non-informative prior:

$$p(\tau) \propto \left|I(\tau)\right|^{1/2} = \tau^{-3/2}$$

where $I(\tau)$ is the Fisher information for τ. When other commonly used non-informative priors are used, we may have slightly different numerical results but the difference will almost always be negligible from a practical perspective.

As shown in the numerical analysis in the next section, the chosen non-informative prior will result in a great deal of uncertainty in the first a couple of time intervals during which far too few observations are collected. When more data become available, the degree of the uncertainty associated with the estimate will be rapidly reduced. This is similar to most of the recursive methods (see, e.g. Ljung and Soderstrom, 1987).

Now applying Bayes' rule to combine the above prior with likelihood $p(m_k;\tau^{-1})$ in equation (1), we obtain the following posterior distribution of τ in the first time interval:

$$p(\tau \mid m_1) \propto p(\tau)p(m_1;\tau^{-1}) \propto (T/\tau)^{(m_1+3/2)} \exp\{-T/\tau\}$$

Hence, the posterior distribution of τ has an inverse gamma distribution $inv\Gamma(m_1+1/2,T)$ with a posterior mean equal to $\mu_1 = T/(m_1-1/2)$ and a posterior variance equal to $\mu_1^2/(m_1-3/2)$, where $inv\Gamma(\alpha,\beta)$ denotes an inverse gamma distribution having probability density function $g(t) = \{\beta^\alpha/\Gamma(\alpha)\}t^{-(\alpha+1)}\exp(-\beta/t)$, and $\Gamma(\alpha)$ is the gamma function.

Let $\alpha_2 = m_1 + 1/2$. Then the posterior distribution of τ in time interval $k=1$ can be rewritten as $inv\Gamma(\alpha_2,(\alpha_2-1)\mu_1)$.

Now turning to the next time interval *k*, we specify the prior in time interval *k* as the posterior obtained in the previous time interval $k-1$:

$$\lambda \sim inv\Gamma(\alpha_k,(\alpha_k-1)\mu_{k-1}), \tag{2}$$

where the prior mean is equal to μ_{k-1} and the prior variance is equal to $\mu_{k-1}^2/(\alpha_k-2)$. The prior mean represents an estimate of the time headway parameter τ obtained *a priori*. The hyper-parameter α_k is associated with the accuracy of the prior information about τ. Note that this prior distribution reflects prior knowledge on the time headway parameter for the 'local' time interval *k* only.

Now we apply Bayes' rule to combine likelihood $p(m_k;\tau^{-1})$ in equation (1) with the prior specified in equation (2). It can be shown by some algebra that the posterior distribution of τ is

$$\tau \mid m_k \sim inv\Gamma(\alpha_k+m_k,(\alpha_k+m_k-1)\mu_k), \tag{3}$$

with the posterior mean

$$\mu_k = \omega_k\mu_{k-1} + (1-\omega_k)T/m_k \tag{4}$$

and the posterior variance

$$\text{var}(\tau \mid m_k) = \mu_k^2 /(\alpha_k + m_k - 2),$$

where $\omega_k = (\alpha_k - 1)/(\alpha_k + m_k - 1)$ is a weight. Note that the currently observed time headway in time interval *k* is T/m_k which may be considered as a crude estimate of the vehicle time headway parameter τ. Throughout this paper, however, we use a posterior mean as an estimate of the vehicle time headway parameter. From equation (4), the current estimate of the time headway parameter τ is a weighted average of the previous estimate μ_{k-1} and the current crude estimate T/m_k. Since more information is used, the quality of the estimate (4) is better than the crude estimate T/m_k.

From the posterior distribution of τ in equation (3), a 95% credible interval for τ can be obtained:

$$(2(\alpha_k + m_k - 1)\mu_k / \chi^2_{0.025}(2(\alpha_k + m_k)), 2(\alpha_k + m_k - 1)\mu_k / \chi^2_{0.975}(2(\alpha_k + m_k))),$$

where $\chi^2_\alpha(df)$ is the value for the chi-squared distribution with *df* degrees of freedom that provides a probability of α to the right of the $\chi^2_\alpha(df)$ value.

Now moving to the next time interval, we treat the posterior distribution (3) obtained in time interval *k* as the prior distribution in time interval $k+1$:

$$\tau \sim inv\Gamma(\alpha_{k+1}, (\alpha_{k+1} - 1)\mu_k)$$

where the diffusion parameter is updated as $\alpha_{k+1} = \alpha_k + m_k$. Hence, when a new observation m_{k+1} becomes available in time interval $k+1$, the estimate of τ can be updated as outlined earlier.

Recursive estimation. In practice, the headway time parameter τ evolves slowly over time. To take this into account, we follow Li's (2009) approach and define a 'forgetting' factor so that observations collected at different times are weighted differently, with the latest observations given the largest weights. In practice, the forgetting factor can be treated as a tuning parameter and determined experimentally. Specifically, the prior distribution is now specified as

$$\tau \sim inv\Gamma(\delta\alpha_{k+1}, (\delta\alpha_{k+1} - 1)\mu_k)$$

where δ is a forgetting factor lying in the interval (0, 1). This factor does not affect the prior mean of τ but its prior variance is inflated, reflecting the fact that *a priori*

we are less sure about the current value of τ because of its evolution. This results in the following algorithm for the recursive estimation of τ :

ALGORITHM. Given: length of time interval *T*; forgetting factor δ ; initial prior parameters $\alpha_1 = 1/2$ and $\mu_0 = 0$.
For *k*=1:*K*
Step 1. Collect the current traffic volume m_k ;
Step 2. Calculate weight $\omega_k = (\alpha_k - 1)/(\alpha_k + m_k - 1)$;
Step 3. Estimate the time headway parameter τ by $\mu_k = \omega_k \mu_{k-1} + (1-\omega_k)T/m_k$;
Step 4. Calculate the posterior variance $\mu_k^2/(\alpha_k + m_k - 2)$;
Step 5. Update the diffusion parameter: $\alpha_{k+1} = \delta(\alpha_k + m_k)$;
End.

If there is no vehicle passing through the ILD in a time interval, then according to Bayesian theory the posterior distribution remains the same as the prior distribution, i.e., $\mu_k = \mu_{k-1}$ and $\alpha_{k+1} = \max(\alpha_1, \delta\alpha_k)$. In addition, it is clear that when the forgetting factor δ becomes larger (smaller), the estimate of vehicle time headway becomes smoother (rougher). In particular, when δ is taken to be 0, all information collected previously is discarded and the estimate is based solely on the current observation T/m_k .

The optimal value for the forgetting factor δ depends on many traffic related factors. Technically it can be treated as a tuning parameter so that it is determined experimentally in practice to achieve the best performance. Consequently, by suitably choosing the value of δ , the developed recursive method can always outperform that of the crude estimate T/m_k as the latter is simply its special case with δ =0.

A practical example

To illustrate the method, we consider a real traffic example. A single ILD was chosen for analysis. It is located in the north of Seattle in Interstate-5 with three lanes in both the northbound and southbound directions. Data measured via this single ILD were downloaded at http://www.its.washington.edu/tdad. The downloaded data included measurements on volume and occupancy during each time interval of duration 20 seconds in the earlier morning period from midnight to 6 a.m. on a normal working day in July 2006.

For illustration purposes, we consider the traffic in northbound lane 2 only. For simplicity, in the following analysis the forgetting factor was fixed at $\delta = 0.95$. Figure 1 displays the estimated vehicle time headway (real line) and the envelope of a nominal 95% credible interval (dotted lines).

It can be seen that during the early morning period, the estimated time headway was usually between 10 and 15 seconds. After 4:30 a.m., it reduced rapidly to a level below 5 seconds. These results are in line with those obtained using dual loop detector data in Zhang et al. (2007).

At the beginning of the time period of interest, i.e. the time just after the midnight, the upper limit of the Bayesian credible interval was very high due to the diffusive prior at the time interval k=1. However, when more information was collected and thus a more informative prior was incorporated in each time interval, it came back rapidly to the normal level.

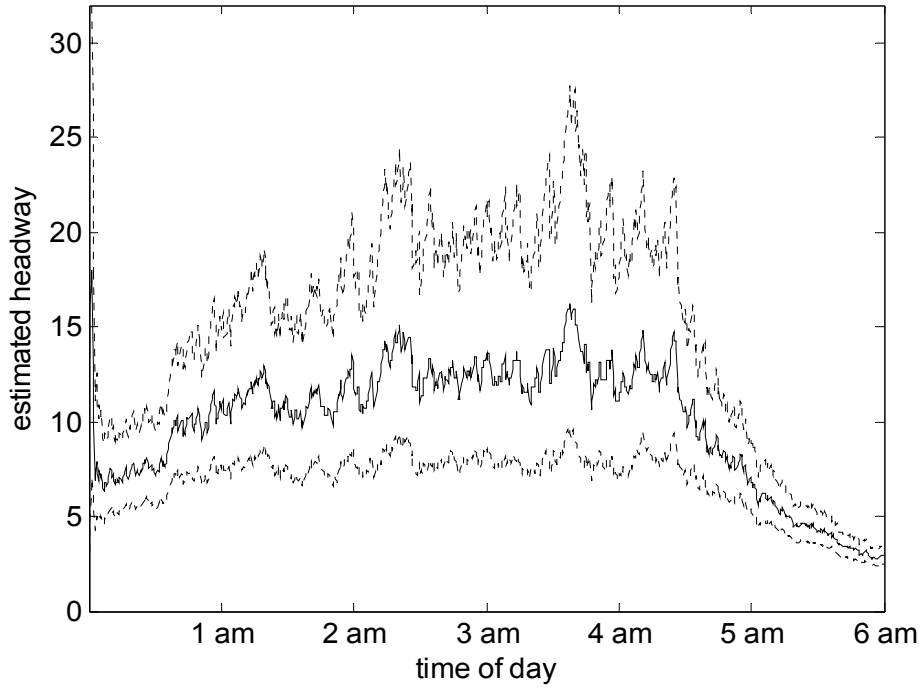


Figure 1. The estimated time headway in the early morning (real line) and the associated envelope of a nominal 95% credible interval (dotted lines).

Although the estimated time headway was mostly above the level of 10 seconds between 1:00 a.m. and 4:30 a.m., the envelope of the 95% credible interval was much wider. The upper limit could reach to 20 seconds or occasionally over 25 seconds. This indicates a substantial degree of uncertainty associated with the estimate. The wider credible interval is a consequence of the larger variability of the traffic in this early morning period, and/or the fewer observations in each time interval.

Concluding remarks

Under the light traffic condition, time headway analysis has been carried out using the data collected from a single ILD. As a result, a new approach has been developed

for the recursive estimation of time headway. Compared with static analysis for vehicle time headway, the recursive estimation developed in this paper has utilized the information collected in the previous time intervals and can reflect the evolution of vehicle time headway over time. The quality of the estimate is assessed in terms of Bayesian credible intervals and posterior variances. Since single ILDs are installed throughout strategic arterial roads, this recursive method is applicable in wide geographical areas for online monitoring of traffic flow in all road networks where ILD data are available.

References

Dailey, D. J. (1999). "A statistical algorithm for estimating speed from single loop volume and occupancy measurements." *Transportation Research Part B*, 313-322.

Dailey, D. J., and Wall, Z. R. (2005). "A general simulation model for use with real freeway data to perform congestion prediction - Phase 3." Washington State Transportation Center - TRAC/WSDOT, Final Research Report WA-RD 635.1.

Gerlough, D. L., and Huber, M. J. (1975). *Traffic Flow Theory: A Monograph*. Transportation Research Board, Washington, D. C.

Li, B. (2009). "On the recursive estimation of vehicular speed using data from a single inductance loop detector: a Bayesian approach." *Transportation Research Part B*, 391-402.

Marsden, G. R., McDonald, M., and Brackstone, M. (2003). "A comparative assessment of driving behaviours at three sites." *European Journal of Transport and Research*, 5-20.

Piao, J., and McDonald, M. (2003). "Low speed car following behaviour from floating vehicle data." *Proc. IEEE Intelligent Vehicles Symposium*, Columbus, Ohio, USA, June 2003.

Ljung, L., and Soderstrom, T. (1987). *Theory and Practice of Recursive Identification*. The MIT Press, Cambridge.

Zhang, G., Wang, Y., Wei, H., and Chen, Y. (2007). "Examining headway distribution models with urban freeway loop event data." *Transportation Research Record* 1999, 141-149.

Trading Facilitation in the Black Sea Area

Radu Hanzu-Pazara[1], Paulica Arsenie[2], Anastasia Varsami[3], Corina Popescu[4], Georgiana Sivriu[5] and Ramona Bejan[6]

[1] Lecturer, [2] Associate Professor, [3] Assistant Professor, [4] Assistant Professor, [5] Assistant Professor, [6] Assistant Professor, Navigation and Maritime Transport Department, Constanta Maritime University, Mircea cel Batran Street, No.104, Constanta 900663, Romania; Email: raduhanzu@yahoo.com

Abstract

Trading facilitation means all the steps that can be taken for obtaining a smooth flow of trade in general. This term is used to illustrate all kind of non-tax barriers. From the point of view of the World Commerce Organization, commerce facilitation is limited only to simplifying and harmonizing the international commercial procedures, covering all activities, practices and formalities involved in the collection, presentation, communication and processing of the necessary data for goods' circulation in international trade relations. The main objective of this paper is to show how trade facilitation is meant to reduce the transaction costs and complexity of international trade in the Black Sea Area and to improve the trading environment in this area, while maintaining efficient and effective levels of government control. This can be done easily by using the simplification of procedures and formalities, the standardization and improvement of infrastructures and facilities, the harmonization of rules and laws applied. Therefore, the main goal of trading facilitation around the Black Sea is going to be emphasized by describing the reduction of transition expenses and the international trade complexity for business meant to improve the changing background of the Black Sea, keeping it efficient and profitable at the same time.

Key Words: trading facility; regional development; transport; security

Introduction

For an optimum development in regional economical areas, having a program for trading facilitation proved to be necessary. Trade means taxes, which means funds

for government and local authorities, funds that can be used for improving roads and railways infrastructure.

Trading facilitation is based on easy procedures regarding custom formalities and transport operations. Once the custom formalities cleared of any useless paper work, transport operations will become faster and therefore less expensive for both parties involved. A clear legislation regarding these operations, might avoid any possible contention which would delay or even obstruct trading operations.

Harbours' safety is also an important factor for providing safe trade transactions, and that is why "safety" became a very important target in every domain nowadays. Economical accords with neighbouring countries, European Union countries involved in the facilitation process of commercial trade, would also contribute to trade facilitation, by providing the necessary support of any trade operations not only in our country but also on the neighbouring territories.

This paper is going to analyze the solutions and initiatives in Romania, especially in Constanta area, which are meant to create a proper environment for the development of trading relations, the trading facilitation.

Why is Trading Facilitation Important?

The importance of procedures to be followed during any import or export operations has increased during the last years and it also led to a higher cost of trade for governments and companies too.

That is why trading facility reasons became important elements in the economical development direction. One of the main reasons is the increasing trading volume due to progressive discharge of commerce. It should also be acknowledged that trading volume increase became reality due to modern technologies that made all the work easier to be done and the large complexity of trade relations had an important contribution by extending trading knowledge after sharing each party's experience.

Meanwhile, the government's role to provide trading facilities in the same time with the state assumption of obligations became very important in actual international security medium. Having governments' support, when dealing with such an important economical branch, proves to be very useful for the country's financial equilibrium.

For a product to be competitive on foreign markets, trade transactions' costs must remain lower. In order to do this, the managerial tools of the supplying chain must become an essential part of all activities involved in trading goods. But this is not enough, and that is why they should be doubled by efficient procedures at national borders meant to avoid unexpected delays or additional costs.
Therefore, trading facility procedures should cover measures concerning formalities' reduction, procedures' accessibility, and documents' content. These measures should

also impose the use of standard messages in commercial transactions in order to avoid any misunderstandings. Other measures which may be necessary to improve the physical circulation of goods are as follows:

- better services offer (transparent, predictable and uniform): appropriate infrastructure for communication and transport, healthy legal environment;
- use of modern technology for communication and information by services providers and end-users;
- increase of users' knowledge about benefits obtained through trading facility.

Trading facility has direct benefits for government and business community too. Governments have benefits due to incomings, improving economy's efficiency, security and social protection. Benefits for traders mean faster delivery of goods and reduced costs for transactions. These benefits are considerable especially for small companies for which the costs resulting from respecting procedures are higher: in this case, trading facilitation increase the capacity of small and medium companies to be part of the international trade.

An OECD study realized in 2001 showed a trade cost between 2% and 15 % from good value. Custom transactions involve 20-30 different entities, 40 documents, 200 data (30 of these are repeated for 30 times at least) according to UNCTAD rapport. So, reducing bureaucratic costs may lead to saving 300 million Euros per year.

Benefits for government are:

- higher efficiency of control methods;
- increasing incomes;
- higher efficient use of resources;
- better economic accords;
- great economical development;
- foreign investment increase.

Benefits for traders are:

- reducing costs;
- no delays in goods' delivery;
- simple commercial base;
- reducing customs formalities.

Trading facilitation methods can help government and business community to improve international trade through simplifying formalities and procedures. These methods would also facilitate standardizations and therefore avoid any doubts. Laws' harmonization may lead to a sort of unification of trading actions around the world and applying the same regulations would provide unprecedented harmony in the trade area. Already existent international suggestions and methods include:

- UN/CEFACT: Recommendations for Trade Facility Compendium;
- UN/EDIFACT;
- ICC regulations: International Customs Guidelines; Incoterms 2000; Payment instruments;

- World Bank Trade and Transport Facilitation Toolkit;
- WCO Accords; IATA Conventions; ISO Standards; FIATA general lines.

According to international trade organizations, implementation of trading facilitation systems can be done in different ways.

One of these methods could be through government strong implication, resulting in an excessive control and excessive regulations. A better solution will be to establish a policy based on cooperation with the business community. The second option is recommended, because the government and companies have common interests for improving the commercial environment.

The facilitation never stops governments to exercise the necessary control. Nevertheless, consulting the business community about the necessary improvement at the institutional and legal level will facilitate agreements to the companies' level. In that way, governments can start facilitation of trading procedures through open dialogue with the business community and by encouraging traders to discuss about their problems.

Next step can be represented by the national trade analyze patterns and the starting points of the facilitation process, including:

- taking in consideration a regional or global solution;
- using international standards and procedures already in use;
- including trading facilitation in customs, transportation, banking and trade finance policy;
- training people involved in international trade at governmental and business community level;

Facilitation of trade is a continuous process. Once the government and business community nominate and start the mechanisms for trade facilitation, these, altogether, must continue monitoring the improvement of trading, at local level and international level, adapting the existing rules and nominating new adequate measures for a continuously changing system.

Evaluation of National Situation

Every evaluation should always take into consideration both the strong and the weak points in a situation, concluding by trying to discover the best opportunities to improve that particular situation in the near future. Therefore we will start by discovering one by one, the strong points our country possesses concerning trade facilitation.

First of all we should mention that our territory enjoys a favourable geographical position for developing trade infrastructures. Given the fact that Romania is already a member of the European Union, our legislation regarding trade facilitation is already conceived according to European Union requests, so the necessary preparations in

infrastructure, cross borders points or even the necessary training courses will be paid by using European funds.

Our country's active participation in international organizations involved in trade facilitation (ICC, WTO, ITC, UNECE) provides us with a strong position also in European and regional business organizations (Euro Chambers, Economic Cooperation in the Black Sea Region, Association of Balkan CCIs); and therefore our trading development enjoys the proper basis and support of the most experienced agencies in Europe.

Another important contribution is brought by the existence of the Export Council as a public/private partnership, activating in facilitation trade problems area; and also by the existence of free change accords with Bulgaria, Moldova, Croatia, Albania, Macedonia, Serbia and Montenegro, Bosnia and Herzegovina, meaning that our country's relations with our neighbours is not only a good one but also a beneficial one from the trading facilitation point of view.

We should of course admit that there are also some weak points to be taken into consideration without omitting the fact that solutions are always to be looked for and applied when needed. When it comes to the weak development of roads and railways infrastructure our government has already taken a series of decisions meant to improve and even complete the necessary steps for mending this situation.

The high level of bureaucracy at cross border points may also be seen as an obstacle for the trade facilitation demarche, but becoming a member of the European Union is already having a beneficial effect over this problem, as long as the necessary paper work, for passing borders in countries which are members of the Union, is already considerably reduced. Our public sector also fights with the fact that there are enough public workers with a poor training level in various sectors of trading facilitation. Yet, as long as European funds are meant to be used also for training courses in every activity domain, it is expected in the near future for this problem to be almost completely solved.

The absence of consultation with the national business community about trading facilitation may also be considered a problem for trade facilitation and that is why various forms of partnerships between private and public business authorities in the branch have already been established. Of course there is always the risk of no abilities from government structures to use public-private partnerships to increase trading facilitation, which brings out the necessity to provide better communication instruments in the future.

All these solutions rely on too many changes of laws regarding trading facilitation and transportation which may be difficult to implement due to a long time use of custom formalities. That is why every change needs time to be really integrated into the old background, but as long as all the parties involved in the trade facilitation

process agree on the necessity of an immediate change, all the correspondent efforts will be put to work for the only purpose of reaching the final target of their projects.

One should also take into consideration the opportunities resulting from the steps which are to be followed in the trading facilitation domain. Building an Export National Strategy offers for the first time the possibility to articulate a facilitation of the trade policy with general objectives for export promotion. A first look towards possible importers from our country is given towards our neighbouring countries, and that is why a regional strategy for Balkans, encouraging entrance of these countries in EU till 2014, might be one of the most prolific opportunities.

Another step towards trade facilitation lies on the initiative to include all free change accords with southeast Europe countries in one accord, resulting in a considerable improved base for trading facilitation. The implementation of the New Community Transit System is another advantage our country enjoys by being a member of the European Union and therefore being able to use EU programs for crossing borders to increase trading facilitation. There should also be given the appropriate considerate to the possibility to receive FIATA accreditation for transporters training system.

After analyzing strong and weak points while discovering possible opportunities to achieve trade facilitation, we should also take a look on the inevitable threats hanging over the achievement of future strategies. There is a risk for the authorities to find themselves in the impossibility to create a medium and long term plan to use the economical potential offered by trading facilitation, until the obvious communication problem between all interested parties won't be solved. As long as there is no actual plan, there comes the impossibility to use actual resources to create a modern mechanism for trading facilitation, which is also caused by the lack of such resources or the impossibility to reach them when needed.

Our country is well known for the political changes of local public entities which therefore delay the implementation process for few initiatives, as every new agency needs to be all over again convinced of the advantages brought about by any initiative towards trade facilitation. An excess of bureaucracy also appears because of the poor communication between public entities involved in trade facilitation.

Vision

To realize this strategy, the vision defines trading facilitation as one of the essential elements of commercial competition and economical prosperity of Romania. Keenly recognizing the indispensable role ports play in the world maritime commerce, local authorities strongly believe in and promote the fact that a free and smooth flow of maritime cargoes contributes to the healthy and sound development of national trade and economy.

With this backdrop, our country will continue to work together with international organizations concerned with the simplification and harmonization of international

trade procedures, such as WCO (World Customs Organization) in the context of improving Customs procedures and UNCEFACT (United Nations Center for Trade Facilitation and Electronic Business) in the context of promoting Information Technology (IT) applications and electronic commerce in maritime commerce.

Therefore, trade facilitation reform is directed primarily at reducing, if not eliminating, unnecessary trade transaction costs, in particular those that impair the efficiency of international trade and commerce. In broad, conceptual terms, for governments trade facilitation can mean increased and more efficient revenue collections, and more effective detection of fraud, evasion and other non-compliance practices.

For the trading community, both exporters and importers, and those providing them with services, such as the transport industry, trade facilitation holds out the prospect of faster and more predictable movements of goods and services, and more efficient transit procedures.

And, for nations as a whole, including their citizens, trade facilitation reform means enhanced trade competitiveness, better economic performance, greater foreign direct investment and, through these, expanded employment opportunities and higher incomes. As would be expected, the greatest potential for, and hence likely beneficiaries of, trade facilitation reform are likely to be developing countries, such as our country, and those developed countries which lag behind international best practice.

Strategic objectives

For this vision the strategic objectives are:

- improving legislation for trading facilitation;
- increasing companies' competence in trading facilitation and transport area;
- increasing the efficiency of commercial transactions;
- identification of needs and demands of foreign markets;
- contribution to reduce custom formalities;
- increasing the security for Constanta Harbour to facilitate containers trade;
- reducing crossing border time for goods and people in southeast Europe;
- modernizing Romanian infrastructure;
- improving custom logistics.

Initiatives

One of the main initiatives to achieve these objectives include modernizing ROMPRO legislation as the only public-private organization specialized in trade facilitation, this process implying first of all a reduction of any unnecessary paper work. Another step would be the organization of special training programs for companies involved in trading facilitation, based on the experience gained in the

TTFSE program, reducing in this way if not excluding one of the weak points referring to the poor training level of public workers in various sectors of trade facilitation.

Increasing access of Romanian business people to global or international events axed on trading facilitation problems would provide an actualization of the local experience concerning trade reform. There should also exist an initiative towards the improving the access of Romanian companies to information regarding requests on products market, facilitation this way the local economical growth. Information about actual free change accords should be widely spread and local agents should be encouraged to use them, along with spreading the information about closer free change accord in southeast Europe area and its use.

To find solutions to relieve visa accordance procedures for business people and professional drivers in southeast Europe is an initiative which would facilitate any type of transit involved in trading facilitation. Local authorities should try to improve implementation of the "single window" concept and encourage meanwhile the use of electronic commerce procedures to increase transparency and to limit corruption. Removing National Roads Company infrastructure from border area to inside points may prevent the influence of one of the weak points referring to the many changes of laws regarding trading facilitation and transportation. The number of authorities in cross border points should be reduced along with the number of taxes applied in customs to transporters and the number of local taxes applied for transit. In order to bring the local general situation up to date initiatives such as modernizing roads and railways infrastructure should be taken along with the modernization of customs facilities.

Resources

Main resources identified for achieving these objectives are financial, technical and human.

Financial resources are mainly allocated by:

- state founds;
- European Union founds allocated to Romania especially after integration in European Union (structural founds);
- pre-adhere European Union programs, especially for training;
- World Bank programs.

The technical resources taken into consideration are those which come from shifting experiences and practices offered as a collaboration form in the strategy plan area.
Human resources for implementation of these programs and initiatives are:

- Romanian experts from Romanian Centre for Research, Education Ministry, Export Council and ROMPRO Organization;
- foreign experts from European Union Commission, TAIEX, UNECE and other European Union countries;

- Romanian experts trained in previous programs as TTFSE I or other programs provided by SECIPRO.

A very important element regarding human resources is changing the mentality, which must be induced to Romanian authorities and to operators for a successful implementation of a modern and flexible mechanism of trading facilitation.

Conclusions

Regional development will always depend on the increasing quota of commercial transactions made in the area. In order to improve commercial relationships it is necessary to stimulate them by creating and according facilities.

Trading facilitation can be made by governments of states in the region, but also by the implication of business environment.

Simplification or even exclusion of some custom formalities, elimination of bureaucracy in commercial relationships, creating a safe environment for commercial trade seem to be the main targets that can lead to a strong and viable regional development.

The application of commercial facilities in the Black Sea area by the governments of states in the region and by the local business communities will create a prosper business environment and will make possible the extension of trading relationships worldwide.

References

Branch, A. E. (1998). *Maritime Economics, Management and Marketing.*"Stanley Thornes Publishers, London.

Constantin D. (2002). *Probleme actuale ale dezvoltarii regionale in Romania***.** Oscar Press Publishing House, Bucuresti.

Dollar, D., M. Hallward-Driemeier and Mengistae, T. (2003). "Investment Climate, nfrastructure and Trade: A Comparison of Latin America and Asia." *Stanford Center for International Development*.

Gakunu, P. (2003). "Trade Facilitation: The Challenge for Growth and Development." *United Nations Economic Commission for Europe (UNECE)*.

Gibson, S. (2006). "Ports & Terminals Guide 2007-2008 – Volume 1." *Fairplay Publications Inc, London.*

Goorman, A. (2004). "Customs Modernisation Initiatives: Case Studies." *Washington, D.C.: World Bank*.

Hummels, D. (2000). "Have International Transportation Costs Declined*?." Purdue University*.

Ianos, I**.** (2000). "Sisteme teritoriale, o abordare geografica**.**" *Tehnica Publishing House, Bucharest.*

Messerlin, P.A. and Zarrouk J. (1999). "Trade facilitation : technical regulations and customs procedures." *Geneva: Trade and Development Centre*.

Nordas, H., Grosso K. and Pinali E. (2006). "Logistics and Time as a Trade Barrier." *OECD Trade Directorate, Paris*.

Persson, M. (2007). "Trade Facilitation and the EU-ACP Economic Partnership Agreements: Who Has the Most to Gain?" *Department of Economics, Lund University, Sweden*.

Porter, J. (2009). "Black Sea and East Mediterranean weakness knocks box cargoes." *Lloyd's List*.

Staples, B.R. (2002). "Trade Facilitation: Improving the Invisible Infrastructure." Washington, DC.

Wilson, J.S., Mann C.L. and Otsuki T. (2003). "Assessing the Potential Benefit of Trade Facilitation: A Global Perspective." *World Bank Policy Research Working Paper*.

Determination of Crash Surrogates Using Tire Skid Marks

Hongyan Gao[1] and Zong Z. Tian[2]

[1] Department of Civil and Environmental Engineering, University of Nevada, Reno, USA; PH: 775-784-6195; gaoh@unr.nevada.edu
[2] Department of Civil and Environmental Engineering, University of Nevada, Reno, USA; PH: 775-784-1232; zongt@unr.edu

Abstract

The objective of this research was to seek valid surrogates for crashes, so that more reliable decisions and proactive measures could be made to improve traffic safety. Inspired by the FHWA's Road Safety Audit program, this paper focused on tire skid marks. A case study was conducted on tire skid marks collected from selected freeway and arterial segments in the Reno-Sparks area in northern Nevada, USA. Parameters were identified to provide a possible guideline for skid mark classifications. Furthermore, linear regression was applied to determine whether skid mark frequency was a suitable predictor of crashes. The regression analyses indicated that there was a relatively strong linear correlation between skid marks and crashes on freeway segments but a weaker correlation on arterials. The frequency of tire skid marks on freeway was found to be statistically significant (at a confidence level of 95%) in being a predictor for crashes.

Introduction

Since the 1960s, crashes have been widely used as the key performance measure in traffic safety studies. Crash statistics, such as crash frequency, crash rate, and crash severity are usually used to evaluate the performance of traffic facilities, and to influence safety improvement decisions. However, considering the characteristics of scarcity, inaccuracy, and incompleteness of crash records, one should be skeptical about the reliability of the conclusions and recommendations drawn based on studies which are entirely based on crashes. Therefore, it is needed for researchers to explore other alternatives to the crash data, which could provide more information and occur more frequently. These alternatives are referred to as surrogate measures of safety (or crash surrogates). Another advantage of using crash surrogates is that proactive safety

measures can be implemented before crashes occur. The primary objective of this study is to identify valid crash surrogates for roadway arterials and freeways. By referring to these surrogates, dangerous locations could be determined and mitigation measures implemented, such that crashes could be prevented before loss occurs.

Previous Studies on Crash Surrogates

According to previous studies, intersections and freeways are the two traffic facilities which have been given the most attention. For intersections, Traffic Conflicts (Amundsen et al., 1977; Zimmerman et al., 1977), Time-to-Accident (Hyden, 1977 and 1987), Time-to-Collision (Hayward, 1972; Kruysse, 1991), Post-Encroachment Time (Allen et al., 1978), and Deceleration Rate (Allen et al., 1978) are the most widely used measures to assess intersection safety. Also, volume-based variables were studied as predictors for crashes (Hauer et al., 1988; Box et al., 1976). However, despite the extensive studies on these measures, no apparent agreement has been reached regarding which surrogate measures are the most valid and reliable. For highway/freeway segments, measures from four critical aspects are used to evaluate traffic safety. They are: 1) geometrics, which include curve radius (Glennon et al., 1985), grade (Hauer, 1999), and width of lanes, median, and shoulders (Belmont, 1954; Perkin, 1957), 2) weather and environment (Madanat, 1997; Kopelias, 2007), 3) speed-related variables, for example, speed limit and speed variation (Carlsson, 1996; Joksch, 1993), and 4) traffic flow (Amundsen et al., 1977; Lee et al., 2003). Again, based on the studies reviewed, even though certain relationships between these measures and crashes were reported by previous researchers, conclusion varies on which measures could be considered as effective crash surrogates.

The Road Safety Audit program (*FWHA Road Safety Audit Guidelines*, 2006; FHWA Road Safety Audits website, 2009) launched recently by the FHWA has been implemented in some States' Department of Transportation. This program provides new ways of approaching traffic safety issues and identifying promising surrogate measures which have not received much attention up to now. These potential surrogates include citizen complaints and physical evidence of past crashes and off-road excursions, such as curb damage, tire skid marks and evidence of broken auto parts. Locations where these evidences concentrate are believed to be hazardous spots, thus corresponding mitigation measures need to be decided for these places to prevent more severe crashes in the future. Even though they are considered as potential indicators for crashes, to the authors' best knowledge, there is no literature documenting their validity of being a crash surrogate. For this reason, this study was devoted to investigating whether the tire skid mark on the pavement is a valid crash surrogate.

Case Study

Data Collection. The study sites included both arterial segments and freeways in the Reno-Sparks metropolitan area in northern Nevada (see Figure 1). Two arterial segments on McCarran Boulevard were selected for the study: a segment between

Plumas Street and Neil Road on South McCarran Boulevard for a distance of 1.6 miles with 7 signalized intersections and a segment between Sutro Street and Fourth Street on North McCarran Boulevard for a distance of 2.7 miles with 11 signalized intersections. For the freeway segments, the study sections include a west-east stretch on Interstate 80 and a north-south stretch on U.S. Highway 395. The first study section includes approximately 10.5 miles of Interstate 80 between Robb Drive (Exit 9) and Vista Boulevard (Exit 19) interchanges. The second stretch of freeway runs along U.S. Highway 395 between the Mount Rose Highway/SR 431 (Exit 56) and Panther Valley (Exit 72) interchanges for a distance of approximately 14.8 miles. Both directions of travel were evaluated for each selected segment, resulting in four freeway segments and four arterial segments analyzed.

Per the study objective, to identify whether skid marks are a valid crash surrogate, there is a need to know if crashes took place at locations where skid marks were observed. To acquire information on their location and appearance, videos were used to collect image records of skid marks on the pavement. The data collection was conducted in August, 2007 for freeways and October, 2007 for arterials. It should be noted that the collected skid marks are only good representations of what had happened over the past few months of the year. Considering the frequent precipitation in winter season in the study area, the tire skid marks left from the past year were normally washed away entirely. The area's generally dry weather condition in spring and summer allow a long-lasting presence of new tire skid marks. Thus, by the time of the data collection, it is assumed that about three to five months of newly generated tire skid marks had accumulated. Therefore, to correlate the collected skid marks with crashes, crash records for the study sections were obtained by courtesy of the NDOT Safety Division for the common time period. However, due to the scarcity of crash records, a three-year period, starting from September 1, 2004 to September 1, 2007, was used for arterials while a one-year period, ranging from September 1, 2006 to September 1, 2007, was used for freeway segments.

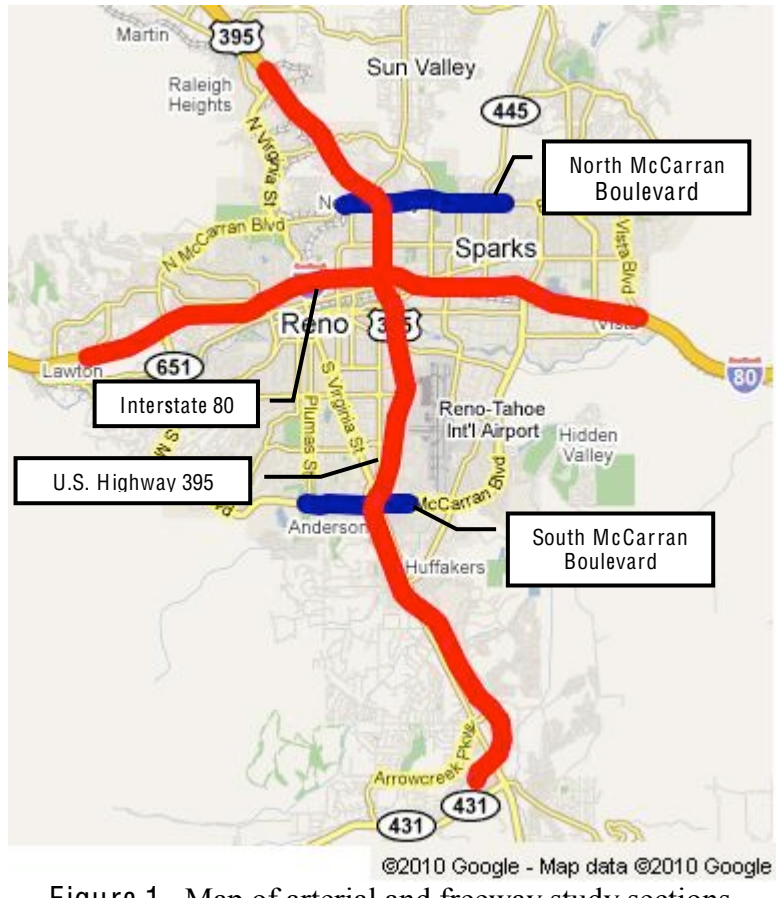


Figure 1. Map of arterial and freeway study sections.

Characteristics of Skid Marks. Tire skid marks of different characteristics are caused by different situations, thus different types of marks could be used to describe different incidents. Upon reviewing the skid mark videos, the authors observed many unique skid marks on the arterial and freeway sections. Variations were noted in length, darkness, position, apparent course, and other factors. Nine parameters were eventually identified and given pertinent definitions to document these skid marks. Meanwhile, corresponding photographs from the study segments were taken to provide pictorial demonstration.

They are: (1) Location, which describes the position of the end of a skid mark. Considering, in most instances, vehicles leaving a skid mark on the roadway are undergoing a sudden deceleration in a preventative or evasive action to avoid a suddenly stopped car or other obstacle, it would be reasonable to assume that a collision involving a vehicle which left a skid mark occurred towards the end of the mark. (2) Lane position, which defined as the lane of the roadway in which the skid mark occurred. (3) Wheel, which indicates whether a vehicle left behind skid marks on one or both sides of the vehicle. The skid mark shown in Figure 2(a) appears to have been made by both the left and right wheels of a heavy vehicle, while the mark in Figure 2(b) appears in the middle of the lane, thus its wheel parameter is uncertain for this case. Figure 2(c) and 2(d) show the skid marks left by left wheel and right

wheel respectively. (4) Heavy vehicle, which indicates whether or not a skid mark appears to have been caused by the actions of a heavy vehicle, including a truck, bus, trailer, or other commercial vehicle which typically features four or more wheels on its rear axle. This type of skid marks was featured with closely-spaced on one side of a lane, beginning and ending at approximately the same point, and following one distinct path. Figure 2 (a), 2 (c), and 2 (d) all show skid marks left by the heavy vehicles. (5) Direction, which indicates the approximate direction the vehicle was heading when it formed the skid mark. It can be mainly described by the following three formats, which are straight, veered, (deviated less than approximately 30° from the centerline of the original lane), and swerved (deviate more than 30° from the centerline of the original lane). The skid mark in Figure 2(e) slightly veered towards the other lane on its left side, while the one in Figure 2(f) veered to the freeway shoulder. Figure 2(g) and 2(h) show drastic swerving marks, which were featured by the tread marks appearing to continue in a straight path while the overall mark curved to the side. This is an indication that the vehicle likely lost control. (6) Impact, which is used to indicate whether the vehicle that caused the skid mark appears likely to have impacted the barrier (dividing wall, guardrail, median, etc.) on any side of the road. Figure 2(g) shows the mark on Interstate 80 Eastbound which appears to have swerved left and impacted the left guardrail. (7) Intensity, which is a relative measure of the darkness of skid marks. Three basic severity levels were defined based upon the contrast of the color of the skid mark versus the color of the pavement surface. They are 1) faint skid marks which were difficult to see, since the predominant visible color was the color of the pavement surface, 2) average skid marks whose tire tread patterns were generally fairly easy to see on the pavement when viewing the video, and 3) dark skid marks which generally appeared to be a solid black line, thus tread marks could not be easily discerned on the pavement. The skid mark in Figure 2(i) was classified “faint”. Figure 2(j) shows the “average” marks. In Figure 2(k), the outer tires of this heavy vehicle skid mark are considered “dark.” (8) Skip, which indicates whether a skid mark appeared as a dotted line on the pavement. It was most likely to occur on marks that were also classified as heavy vehicles. Figure 2(l) shows a pair of skipping marks on Interstate 80 Eastbound. (9) Width, which indicates the width change of skid marks. In cases when one skid mark has multiple widths longitudinally along the pavement, a multiple letter input was given to this parameter. In most cases, skid marks that changed width also veered or swerved from the originating lane. Figure 2(m) shows two pairs of swerving marks that changes their width from narrow to wide and back to narrow again.

It was theorized that there might be interesting correlations between different skid mark characteristics and different crash types. Such associations are somewhat typical to investigators in reconstructing an accident scenario, but could also provide some insight into accident prediction. The study on parameters presented here was intended to provide a possible guideline of skid mark classification and show how such analysis should be conducted in a future study. However, since we mainly focused on using their locations to investigate crashes in this study, the characteristics of skid marks were not further explored.

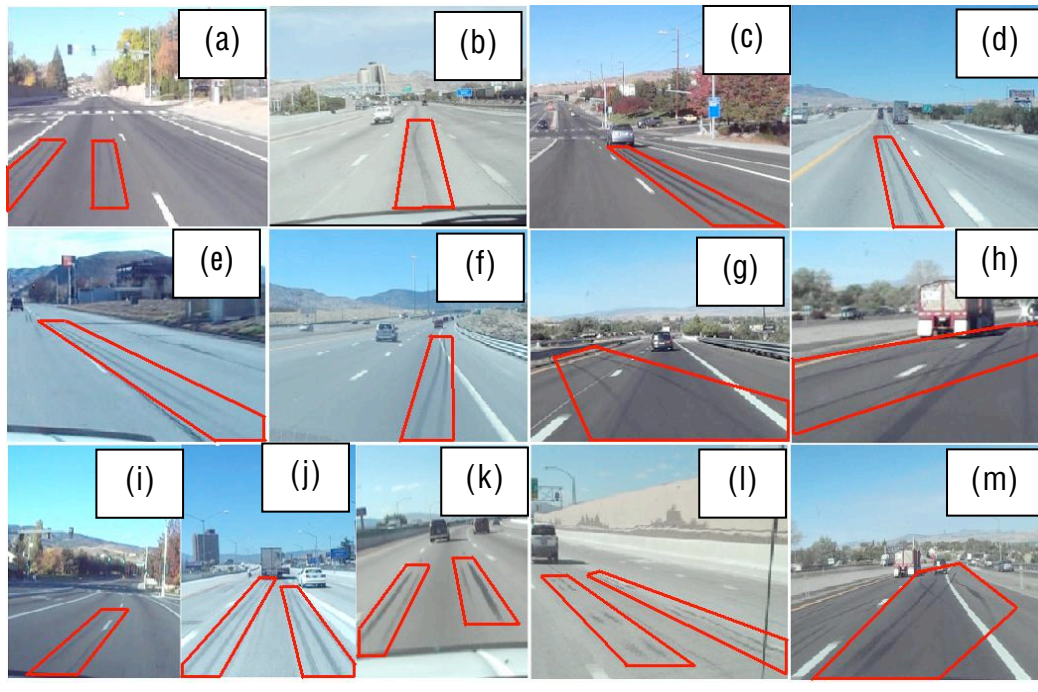


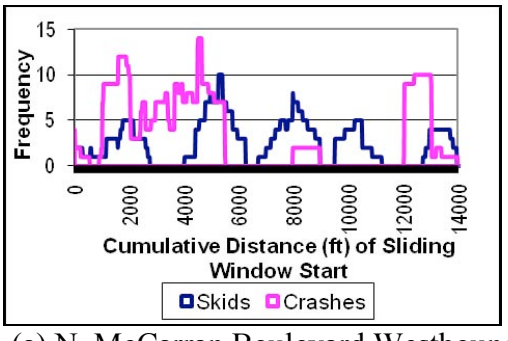
Figure 2. Examples of different types of skid marks in study segments.

Correlation Analyses of Skid Marks and Crashes. Similar to many surrogate studies in the past, statistical correlation was largely utilized in this study. To use linear regression to demonstrate if the skid mark is a valid variable to predict crashes in each study segment, the numbers of crashes and skid marks were collected within consecutive sections. These sections were divided by reference points, such as intersections on arterials and mileposts on freeways.
Furthermore, to study the sections which contain both upstream and downstream sides of the reference point, sliding-window analysis was applied. In a sliding window, a window of a user-determined length "slides" forward by a small but appropriately-sized user-specified step, and reports all relevant information within the window for each step. Here, it is used to report the amounts of crashes or skid marks and the locations with high frequencies of skid marks and crashes.

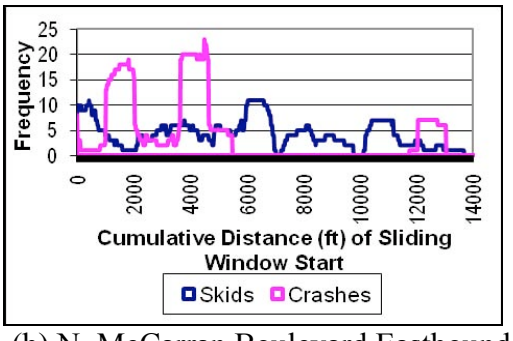
Skid Marks on Arterial Segments. To identify the statistical relationship, linear regression was conducted using skid mark and crash frequencies in intersection-divided sections, frequencies per 1000 feet (304.80 meter) in every intersection-divided section, and frequencies in sliding windows. The last two segmentation scales eliminated the impact of section distance, considering the fact that a longer section can increase the probability of observing more crashes.

Three different scenarios of window length and forward-step increment were initially studied, including a 1000-foot window with a 100-foot (30.48 meter) step, a 1000-foot window with a 20-foot (6.096 meter) step, and a 500-foot (152.40 meter) window with a 50-foot (15.24 meter) step. After regression testing, the scenario of a 1000-foot window with a 20-foot step was selected for further study, since it showed

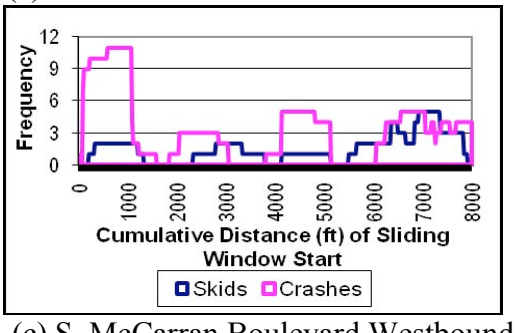
stronger correlations. Based upon the sliding-window output, frequencies of crashes and skid marks were plotted in Figure 3.



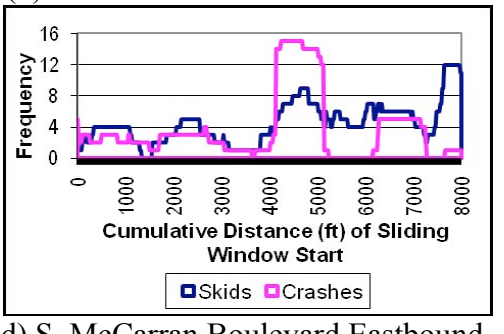
(a).N. McCarran Boulevard Westbound



(b).N. McCarran Boulevard Eastbound



(c).S. McCarran Boulevard Westbound



(d).S. McCarran Boulevard Eastbound

Figure 3. Frequencies of crashes and skid marks based on sliding windows

Table 1. Regression outputs for arterials.

Arterial Stretch	Segmentation Scale	Direction	Regression Equation		P-value of Slope	R-square
			Slope	Intercept		
North McCarran Boulevard	Segment	EB	0.08	5.57	0.85	0.00
		WB	-0.31	6.15	0.55	0.05
	Segment/ 1000 ft	EB	-0.32	7.61	0.81	0.01
		WB	1.06	1.70	0.32	0.12
	Sliding Window	EB	-0.30	5.12	0.00	0.02
		WB	0.11	3.13	0.11	0.00
South McCarran Boulevard	Segment	EB	0.10	4.80	0.83	0.01
		WB	1.31	2.71	0.25	0.31
	Segment/ 1000 ft	EB	-0.03	4.09	0.96	0.00
		WB	1.25	1.93	0.16	0.42
	Sliding Window	EB	0.60	0.88	0.00	0.14
		WB	0.64	2.36	0.00	0.07

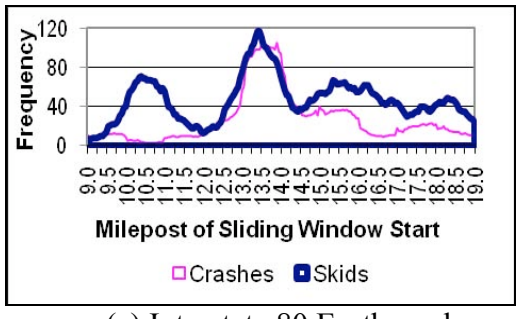
In these graphs, the x-axis represents the cumulative distance of the starting point of each sliding window of a 1000-foot length. The y-axis shows the frequency of

crashes and skid marks within a window starting at a certain point; thus, larger areas under the broken lines have higher frequencies of crashes or skids and as such could be considered as dangerous or problematic locations. Notice in the figures that the segment with the highest number of crashes does not correspond to the highest observed skid marks. Also, the frequency trends of skid marks and crashes appear disparate on each study segment.

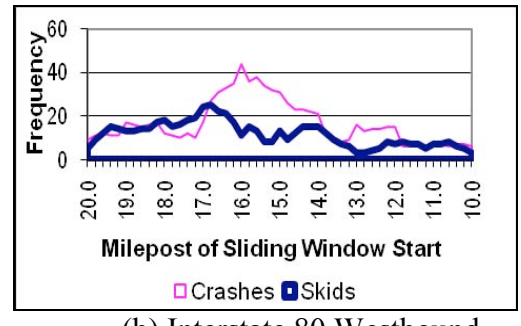
Table 1 presents the results from the linear regression analyses for the study arterial segments. Notice from the table that negative slopes exist, which indicates that crashes do not always increase as skid marks increase. Also, given directional data for each arterial segment, the regression lines may have resulted in different trends (positive or negative) by applying different segmentation scales. Further, with the purpose of investigating if the skids mark frequency is a useful factor for predicting crashes; the P-values of regression slopes were also obtained. When using the most popular confidence level -- 95%, the slopes are statistically significant in the following two scenarios (highlighted in darker shade in Table 1): 1) both directions of South McCarran Boulevard under sliding-window segmentation scales and 2) the eastbound direction of North McCarran Boulevard with sliding-window outputs. This indicates that skid marks are reasonable to be used in predicting crashes in these circumstances It was also noticed that when confidence level is reduced to 90% (which is also widely used), regressions are significant for all scenarios using sliding-window output for the arterials (p-value=0.11 for North McCarran Boulevard eastbound can be approximately considered as 0.10), which outperforming other segmentation methods.. This might be due to the larger data set that sliding windows produced. Additionally, R-squared values were provided to evaluate the degree of linear association between tire skid marks and crashes. In general, very low R-squared values were obtained, suggesting weak linear correlation between crashes and skid marks. The only exception is the westbound direction on South McCarran Boulevard under the two segmentation scales (highlighted in lighter shade in Table 1). However, the low R-squared values do not necessarily mean that skid marks and crashes are not related. Using a three-year period of crash data (for the purpose of increasing sample size) against the few months skid mark data may have skewed the analysis.

Skid Marks on Freeway Segments. The same method was used to analyze two selected freeway stretches, namely Interstate 80 and U.S. Highway 395. Linear regressions were conducted using data generated by sliding windows and the milepost segmentation. Mileposts are visible mileage markers erected alongside the highway every mile (1.61 km), and are commonly used in the United States. Multiple scenarios of window length and forward step increment were initially studied, including 1-mile (1609.344 meter) windows with 0.2-, 0.1-, and 0.05-mile (321.87-, 160.93-, 80.47-meter, respectively) steps, 0.5-mile (804.67 meter) windows with 0.2-, 0.1-, and 0.05-mile steps, and 0.25-mile (402.34 meter) windows with 0.1- and 0.05-mile step. Based on their outputs, the scenarios which produced relatively greater R-squared values were finally adopted for further analysis. Thus, 1.0-mile windows with 0.05-mile step were chosen for Interstate 80 Eastbound and U.S. Highway 395

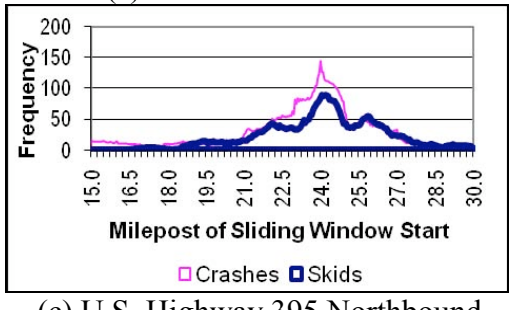
Northbound, while the 1.0-mile windows with 0.20-mile step were chosen for Interstate 80 Westbound and U.S. Highway 395 Southbound.



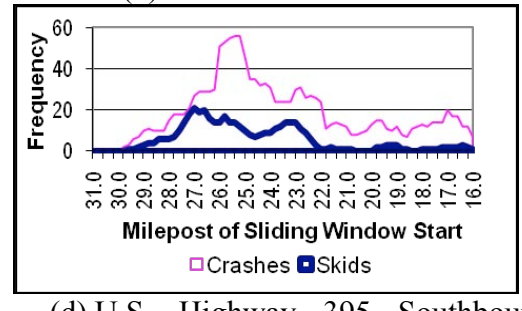
(a).Interstate 80 Eastbound



(b).Interstate 80 Westbound



(c).U.S. Highway 395 Northbound



(d).U.S. Highway 395 Southbound

Figure 4. Sliding window crash and skid frequency on freeways.

Based on the sliding window outputs, frequencies of crash records and skid marks were plotted in Figure 4. In these graphs, the x-axis represents the milepost location of the starting point of each mile-long sliding widow. The frequency of crashes and skid marks of any 1-mile window starting at any milepost can be determined by the y value at that point. It is noticed that the curves of skid marks and crashes follow the same general pattern in Figure 4(a), 4(c) and 4(d), respectively. However, this is not observed from Interstate 80 westbound in Figure 4(b), which is possibly due to relatively smaller data set that sliding window output provided.

Table 2. Regression outputs for freeways.

Freeway Stretch	Division Scale	Direction	Regression Equation		P-value of Slope	R-square
			Slope	Intercept		
Interstate 80	Milepost	EB	0.82	-7.49	0.02	0.47
		WB	0.46	12.72	0.42	0.07
	Sliding Window	EB	0.81	-10.99	0.00	0.51
		WB	0.74	8.20	0.00	0.17
U.S. Highway 395	Milepost	NB	1.48	0.81	0.00	0.85
		SB	1.66	10.50	0.00	0.52
	Sliding Window	NB	1.30	0.00	0.00	0.88

	SB	1.69	9.01	0.00	0.55

Regression outputs were provided in Table 2. Given that all slopes are consistently positive, it is reasonable to conclude that crashes generally increase as the skid marks increase in our cases. Additionally, except for the westbound direction of Interstate 80 under milepost scale, the P-values for regression slopes are small (shown in darker shaded cells), which provides strong evidence that skid mark frequency contains useful information in explaining and predicting crashes, given a confidence level of 95%. Also, the R-squared values of all scenarios were examined to evaluate the correlation between frequencies of skid marks and crashes, and the closer the value is to 1, the greater the degree of linear association between them. Thus, by comparing the outputs of the same freeway stretch of any direction under two division scales, it is found that regressions supported by sliding windows yield higher R-squared values than those by mileposts. This suggests the linear association between skid marks and crashes is relatively stronger by using sliding windows. In addition, given the fairly high R-squared values on most of the stretches (except for Interstate 80 westbound), linear correlation between skid marks and crashes might exist in these cases (highlighted in lighter shade in Table 2).

From the statistical analyses above, linear correlation was found to be weak and regression line trends were reported to be inconsistent in some of the scenarios, especially for arterial segments. All these findings are likely to be influenced by certain inevitable errors that occurred during data collection, including inconsistent time periods of data collection, indistinct skid mark clusters on pavement, observer inconsistency, and incomplete crashes records.

Summary and Conclusions

This paper is used to determine if tire skid marks can be used as a crash surrogate in traffic safety studies. Based on data collected from selected freeway and arterial segments in the Reno-Sparks area, characteristic appearances of skid marks as well as their quantitative relationship with crash records were examined. Nine parameters were extracted to describe and classify all types of skid marks observed in the study area. Statistical analyses were utilized to investigate if tire skid mark frequency is valid in predicting crash frequency and further if there is any linear correlation between them.

Based on the analyses of the data acquired from the case study, the following findings and conclusions were reached:

- regression outputs showed a consistent trend for none of the scenarios on arterials, which indicates the increase of crash frequency does not always correspond with the increase of skid mark frequency. However, crashes generally increased with increasing skid marks at freeway sites.

- skid mark frequency was proven to be statistically significant in performing crash predictions in 25% of the scenarios for arterials and 87.5% of the scenarios for freeway sites(at a confidence level at 95%).
- generally low R-squared values were obtained on arterials, indicating that linear association might not exist between tire skid marks and crashes. On contrary, in most scenarios, relatively high R-squared values were observed on freeway sites, suggesting skid marks and crashes might be linear correlated

Thus, more case studies are recommended to be conducted on freeways in other urban areas. Also, additional work should be conducted to study how the tire skid marks could be used to help determine crash mitigation.

References

Allen, B., Shin, B.T. (1978). "Analysis of Traffic Conflicts and Collisions." *Transportation Research Record: Journal of the Transportation Research Board*. 667.

Transportation Research Board of the National Academies, Washington, D.C., 67-74.

Amundsen, F.H., and Larsen, H.O.E. (1977). "Traffic Conflicts Technique-Status in Norway. " *In Proceedings, 1st Workshop on Traffic Conflict. Institute of Transport Economics,* Oslo, Norway.
Belmont, D.M. (1954). "Effect of Shoulder Width on Accidents on Two-lane Tangents." *Highway Research Bulletin,* Highway Research Board, 91, 29-32.

Box, P. and Oppenlander, J.C. (1976). *Manual of Traffic Engineering Studies*, *Fourth Edition*. Institute of Transportation Engineers, Washington, D.C.

Carlsson, G. (1996). "Vehicle Travel Speeds and the Incidence of Fatal Pedestrian Crashes." International IRCOBI Conference on the Biomechanics of Impact, Brunnen, Switzerland.

FWHA Road Safety Audit Guidelines. (2006). Publication No. FHWA-SA-06-06. FHWA, U.S. Department of Transportation.

FHWA Road Safety Audits website. (2009). http://www.roadwaysafetyaudits.org/benefits.asp. (May 13th ,2009)

Glennon, J.C., Neuman, T.R., and Leisch, J.E. (1985). *Safety and Operational Considerations for Design of Rural Curves.* Publication FHWA-RD-86-035. FHWA, U.S. Department of Transportation.

Hauer, E., Ng, J.C.N. and Lovell, A.J. (1988). "Estimation of Safety at Signalized Intersections." *Transportation Research Record: Journal of the Transportation*

Research Board 1185, Transportation Research Board of the National Academies, Washington, D.C., 48-61.

Hauer, E. (1999). "Safety and the Choice of Degree of Curve." *Transportation Research Record: Journal of the Transportation Research Board* 1665, Transportation Research Board of the National Academies, Washington, D.C., 22-27.

Hyden, C. (1977). "A Traffic Conflicts Technique for Examining Urban Intersection Problems." *In Proceedings, First Workshop on Traffic Conflicts*, Oslo, Norway, 87-98.

Hyden, C. (1987). "The Development of a Method for Traffic Safety Evaluation: The Swedish Traffic Conflicts Technique." In Bulletin 70, Deptartment of Traffic Planning and Engineering, Lund University, Lund, Sweden.

Hayward, J.C. (1972). "Near-Miss Determination Through Use of a Scale of Danger." *Highway Research Record*, 384, 24-34.

Joksch, H.C., (1993). "Velocity Change and Fatality Risk in a Crash-A Rule of Thumb." *Accident Analysis and Prevention*, 25(1), 103-104.

Kopelias, P., Papadimitriou, F., Papandreou, K., and Prevedouros, P. (2007). "Urban Freeway Crash Analysis-- Geometric, Operational and Weather Effects on Crash Number and Severity." CD-ROM, Paper No. 07-0878. Transportation Research Board of the National Academies, Washington, D.C.

Kruysse, H. (1991). "The Subjective Evaluation of Traffic Conflicts Based on an Internal Concept of Dangerousness." *Accident Analysis and Prevention*, 23(1), 53-65.

Lee, C., Hellinga, B. and Saccomanno, F. (2003). "Real-Time Crash Prediction Model for the Application to Crash Prevention in Freeway Traffic." *Transportation Research Record: Journal of the Transportation Research Board* 1840, Transportation Research Board of the National Academies, Washington, D.C., 67-77.

Madanat, S., Liu, P., Pan, H., and Teng, H. (1997). *Freeway Incident Likelihood Prediction and Response Decision-Making.* Report FHWA/IN/JHRP-96-6. Joint Highway Research Project No. HRP-2090, Purdue University, West Lafayette, IN.

Perkins, E.T. (1957). "Relationship of Accident Rate to Highway Shoulder Width." *Highway Research Bulletin*, Highway Research Board, 151, 13-14.

Zimmerman, G., Zimolong, B. and Erke, H. (1977). "The Development of the Traffic Conflicts Technique in the Federal Republic of Germany." In Proceedings, 1st Workshop on Traffic Conflicts. Oslo, Norway.