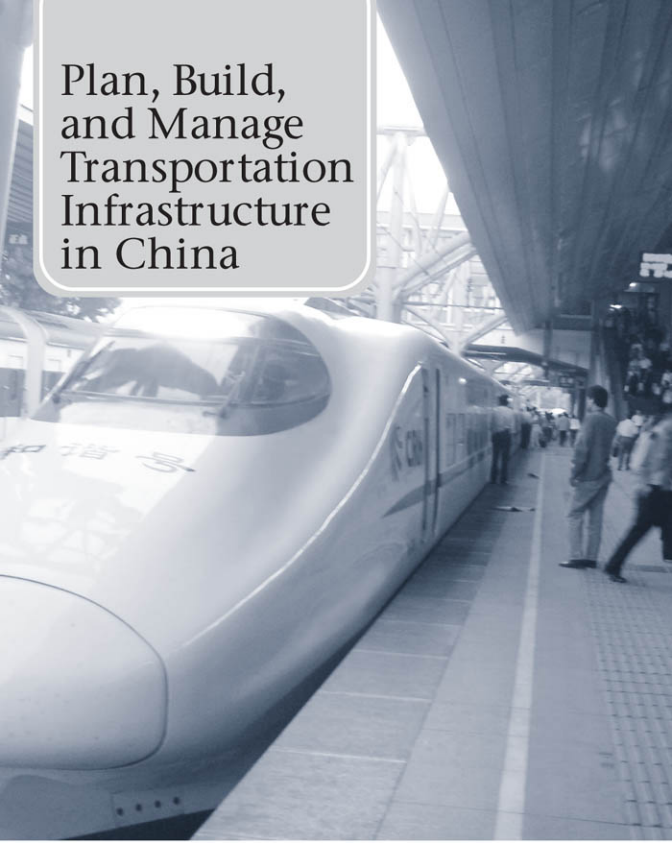
Plan, Build, and Manage Transportation Infrastructure in China



ASCE

Proceedings of the Seventh International Conference of Chinese Transportation Professionals (ICCTP)

Edited by Rongfang (Rachel) LIU, DongYuan YANG, and Jian John LU

T&DI
TRANSPORTATION AND DEVELOPMENT INSTITUTE
A S C E

PLAN, BUILD, AND MANAGE TRANSPORTATION INFRASTRUCTURE IN CHINA

PROCEEDINGS OF THE SEVENTH INTERNATIONAL CONFERENCE OF CHINESE TRANSPORTATION PROFESSIONALS (ICCTP)

May 21–22, 2007
Tongji University
Shanghai, China

SPONSORED BY
North America Chinese Overseas Transportation Associates (NACOTA)

The Transportation & Development Institute (T&DI)
of the American Society of Civil Engineers

The Shanghai Highway Association

Tongji University

EDITED BY
Rongfang (Rachel) LIU, Ph.D., P.E.
DongYuan YANG
Jian John LU, Ph.D., P.E.

Published by the American Society of Civil Engineers

Library of Congress Cataloging-in-Publication Data.

International Conference of Chinese Transportation Professionals (7th : 2007 : Shanghai, China)
Plan, build, and manage transportation infrastructure in China : proceedings of the seventh International Conference of Chinese Transportation Professionals (ICCTP) : May 21 and 22, 2007, Tongji University, Shanghai, China / sponsored by North America Chinese Overseas Transportation Associates (NACOTA), the Transportation & Development Institute (T & DI) of the American Society of Civil Engineers, the Shanghai Highway Association, Tongji University; edited by Rongfang (Rachel) Liu, DongYuan Yang, Jian (John) Lu.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-0-7844-0952-7
ISBN-10: 0-7844-0952-8
1. Transportation engineering--China--Congresses. I. Liu, Rongfang (Rachel) II. Yang, DongYuan. III. Lu, Jian, 1957- IV. North America Chinese Overseas Transportation Associates. V. Transportation & Development Institute (American Society of Civil Engineers) VI. Tong ji da xue (China). Shanghai Highway Association. VII. Title.

TA1005. I5225 2007
388.0951--dc22 2007049676

American Society of Civil Engineers
1801 Alexander Bell Drive
Reston, Virginia, 20191-4400

www.pubs.asce.org

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.

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 (http://pubs.asce.org/permissions/requests/). 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 http://pubs.asce.org/support/reprints/.

Copyright © 2008 by the American Society of Civil Engineers.
All Rights Reserved.
ISBN 0-7844-0952-8
978-0-7844-0952-7
Manufactured in the United States of America.

Preface

The exponential growth in travel demand and related transportation investment in China during the past two decades may stun most ordinary minds especially when compared with the developed countries of the Western World and compared to its own stalled development since the early days. The combination of economic growth, improved standard of living, and urbanization has led to motorization in China, and the process is accelerating. Massive infrastructure building is the current response to the ever-growing travel demand. Governments at all levels and the private sector are all trying to improve the transportation infrastructure, particularly in road and highway construction.

The theme of the Seventh International Conference of Chinese Transportation Professionals (ICCTP) focuses on critical issues associated with planning, building, and managing transportation infrastructure in China. The 7th ICCTP, held in Shanghai, China in May 2007, continues the summer annual conference tradition started seven years ago by North America Chinese Overseas Transportation Associates (NACOTA), a federally registered non-profit organization in the United States. The 7th ICCTP was jointly organized by Tongji University, NACOTA, and the Shanghai Highway Association. The Transportation and Development Institute (T&DI) of the American Society of Civil Engineers (ASCE) was one of the co-sponsors of the conference.

Established in January 1996, NACOTA strives to promote transportation development in China by providing technologies and expertise through its members; strengthen the links between transportation professionals and students in North America and their counterparts in China; promote networking among Chinese overseas transportation professionals and students; and disseminate information on transportation developments and major events in China. NACOTA has more than 500 members and friends in North America and other parts of the world. Starting in 2001, NACOTA has collaborated with a number of universities in hosting its annual summer conference in various Chinese cities including Shanghai (2001 and 2007), Nanjing (2002), Beijing (2003), Wuhan (2004), Xian (2005), and Dalian (2006). The 7th ICCTP was in conjunction with the 100 year anniversary of Tongji University on May 21 and 22, 2007.

More than three hundred people attended the conference and heard more than 100 presentations among the three main tracks:

- Transportation Planning and Policy
- Traffic Operation and Safety
- Infrastructure/Pavement

In order to preserve the sprit of the conference and promote dialogs among various researchers, scholars, and practitioners around the world, we have selected about two thirds of the best papers to be included in these proceedings.

I would like to thank Mr. Dongyuan Yang, Vice President of Tongji University, who co-chaired the conference with me. Professor Yang has led the collaboration effort with

NACOTA and provided his full support for the 7th ICCTP despite his full throttle involvement with the 100 year Anniversary Celebration.

My sincere appreciation goes to a number of administrative and technical leaders in the College of Transportation, Tongji University. Professors Zhongyin Guo, Lijun Sun, Shouen Fang, and Xiaohong Chen, among others, came to Washington DC two years in a row, first to invite the NACOTA delegation to their new Tongji Campus for this conference in conjunction with the 100 year anniversary celebration, and second to pursue me and other key NACOTA members to publish these proceedings. In the end, I would like to say it has been all worth it.

My deepest gratitude goes to my dear friend and mentor, Eva Lerner-Lam. Our breakfast of Chinese crepes with Starbucks coffee in Beijing has "conspired" to produce a number of fruitful projects, including a series of conferences, a transportation glossary, and several potential publications. As the first elected President of T&DI (2005-2006), Eva has trail blazed a series of China initiatives. Her successors, Louis Cohn, Kam Movassaghi, and the T&DI staff Director, Jon Esslinger, are equally enthusiastic and persistent in these efforts. Without their whole-hearted support, these proceedings would not be possible.

I am also indebted to Dr. Jian (John) Lu, the Chair for the Technical Review Committee and its members: Ning Wu, Linbin Wang, Zhongren Peng, Junyi Zhang and Chiapei Chou. Dr. Lu is not only the pillar of the 7th ICCTP but also instrumental in organizing all the past NACOTA China conferences.

Last, I would like to acknowledge the dedication and thorough work of Xiaolei Zou, Tongji University. Mr. Zou has painstakingly gone through all the papers, forms, and other administrative tasks, which made my job as the chief editor so much easier and enjoyable.

Finally, it is a great team effort to put the conference and these proceedings together. It would not be possible without the participation and contribution of all conference attendees, particularly our NACOTA delegates. It would be all worthwhile since I am sure much useful information contributing to the areas of transportation infrastructure planning, building and managing will be found here. I hope to see you in Chengdu at the 8th ICCTP conference in August 2008, one week before the 2008 Beijing Olympic Games.

Rongfang (Rachel) Liu, Ph.D., AICP, P.E.

President,
North America Chinese Overseas
Transportation Association (NACOTA)
http://www.nacota.org

海内外共创育人新天地

百年风雨 同舟共济

十龄学童 齐心协力

The surrounding antithetical couplet was composed by Dr. Rongfang (Rachel) Liu, President of NACOTA, to celebrate the 100 Anniversary of Tongji University and the opening of the 7th International Conference of Chinese Transportation Professionals in Shanghai, May 2007. A brief translation is: the wise and experienced Tongji University, celebrating her 100 year birthday, and NACOTA, a 10 year old youth and energetic organization, will work together in China and internationally to educate the future generations of transportation professionals.

During and After Conference Tour

Contents

Transportation Planning and Policy

Brief Introduction of the Shenzhen Urban Transport Simulation System 1
Dongyuan Yang, Xiaohong Chen, Qun Lin, Hangfei Lin, and Zhichao Guan

Estimating Congestion Index at the Link Level with TransCAD GIS 7
M. Zhang and T. Lomax

Planning Bicycle Corridor for Shanghai Central City 18
Xiaohong Chen, Wen Xiong, and Zhaoyi Huang

Transportation Planning: Travel Demand Modeling Forecast Model Conversion from TransCAD to Cube Voyager 30
Ka-Fai Wong

Design and Development of Planning a Decision-Oriented Transport Information Platform 43
Hangfei Lin, Chao Yang, and Xiaohong Chen

An Origin-Based Algorithm for a Combined Distribution, Hierarchical Mode Choice, and Assignment Network Model 52
Meng Xu, Hillel Bar-Gera, and Ziyou Gao

Computing Job Accessibility with Integrated GIS and DBMS 58
Jiawen Yang

Study on the Relationship between Highway Passenger Transfer Terminals Planning and Urban Spatial Development 67
Xiucheng Guo, Jun Fan, Shuangyan Chen, Haibo Zi, and Hainan Huang

Research on the Equivalent Capacity of Urban Streets and Its Validation 76
Minhua Shao, Lijun Sun, and Xiaoning Zhang

A New Braess's Paradox with Queue Spillovers Considering Dynamic User Equilibrium 90
Xiaoning Zhang and H. M. Zhang

Resident Trip Characteristics Comparison in Small Cities between a Developed Zone and a General Developed Zone 100
Jian-zhi Gao, Yao Zhao, He-ling Ma, and Zhi-yun Zou

The Strategies of Multimodal Transport Transfer System for World Exposition 2010 Shanghai Based on Trip Chain Analysis 111
Wusheng Liu, Ya Wen, Kefei Yan, and Chaowei Yu

Harmonious Urban Development and Strategic Transportation Planning in China 121
Kuo-Ann Chiao, Xi Zou, and Fei Yang

Towards a Hub-and-Spoke Network: A Study on the Chinese Mainland Hub Airport Planning 134
Bo Zou, Huapu Lu, and Qiang Liu

Traffic Distribution Forecast of Urban External Traffic Hub Based on Uncertainty 147
Shunying Zhu, Juxiang Guan, Wuxiong Peng, and Yuan Xu

A Land Use Transport Modeling Framework and Its Design and Development for the Province of Alberta 159
Ming Zhong, John Douglas Hunt, and John Edward Abraham

Evaluation of Yangshan International Hub Port Competitiveness for Container Transshipment Business 170
Rong Zhang and Tao Chen

Selective Evaluation of Traffic-Tools Based on Fuzzy Analytic Hierarchy Process 184
Jianyou Zhao, Mourong Gao, and Meng Du

Predicting the Peaking of Holiday Traffic (Victoria Day Example) 192
Zhaobin Liu and Satish Sharma

Research and Application of Cross-Nested Logit Model in Traffic Mode-Split 203
Gu Wang, Xiucheng Guo, and Jie Yang

Study on Effect of Taxi Quantity to Taxi Split Rate 213
Chaofan Chang, Baohua Mao, and Tuansheng Chen

Intermodal Coordination: A Vital Element for the Success of Shanghai Maglev Services 222
Rongfang (Rachel) Liu, Yi Deng, and Gao Li

Research on Basic Concept and Key Technologies of Experimental Traffic System 235
Zhizhou Wu and Mei Tong

Research on Appropriate Road Grade Proportion of Urban Road Network 243
Haiyuan Li and Wei Deng

Bus Transit Service Optimization—State-of-the-Art, State-of-the-Practice, and Challenges 254
Fang Zhao

Price Negotiation Model in E-Commerce 268
Jinkai Wang

Traffic Operations and Safety

Modeling of Cargo Vehicles Choice in Variable Pricing of Freeways 276
Qiaojun Xiang, Yingwei Ren, Jian Lu, and Shourong Wu

Logical Definition and Application of Expressway Network Guide Sign System 288
Yuren Chen, Yueru Wang, and Shaojun Chen

Research on Pedestrian-Vehicle Collision Model and Simulation Calculation 294
Shouen Fang and Junhua Wang

Safety Evaluation Approach of Highway Intersection 301
Li Yuan, Jian Lu, Guoqiang Zhang, and Qiaojun Xiang

Multi-Objective Transit Route Network Design Using Tabu Search 308
Deming Lei and Xinping Yan

Dynamic Model for Evaluation of Advanced Traveler Information System 317
Chao Yang and James Luk

Feasibilities and Challenges of Probe Technologies for Real-Time Traffic Data Collection **328**
Kai Liu, Toshiyuki Yamamoto, and Taka Morikawa

Driving Fatigue Identification Method Based on Physiological Signals **341**
Zhe Mao, Xin-ping Yan, and Chao-zhong Wu

Evaluation Model for the Urban Public Transit Network Connectivity Based on Graph Theory **352**
Xiaoguang Yang, Chao Zhang, and Zhuang Bin

UCF Driving Simulator's Research and Application in Traffic Safety and Operation **360**
X. Yan, E. Radwan, M. Abdel-Aty, X. Wang, R. Harb, and C. Santos

Further Research and Application of Train Interconnecting Network System **371**
Xi Li, Xiaoning Zhu, and Guoqiang Cai

Safety Research during Entrance and Exit of Highway Tunnel Based on Drivers' Eye Movement Experiment **381**
Xiaodong Pan, Zhigang Du, and Zhen Yang

Safety Analyses at Signalized Intersections: Research Strategies, Modeling Techniques, and Significant Factors **389**
Xuesong Wang, Mohamed Abdel-Aty, and Xiaohong Chen

Research on Train Operating Scheme of Intercity Rail Transit **401**
Xingfang Xu and Jingsong Xiang

Survey and Analysis of Driver's Response to Dynamic Traffic Information **408**
Guiyan Jiang, Zuduo Zheng, Zhu Bai, Leilei Dai, and Jiaqi Zhao

Development of an Application Programming Interface (API) in Simulation to Evaluate Real-Time Signal Control **419**
Fang Clara Fang

Non-Motor Vehicle Trip Characteristics and Its Application for Disabusing Related Policies in Shanghai **433**
Wen Xiong and Xiaohong Chen

The Application of Combinational Evaluation in Traffic Transfer System Based on Multi-Objective Programming **445**
Bin Shuai, XiCheng Zhang, and JiangHong Wang

Performance Measurement on Arterial Roads in Real Time **451**
James Y. K. Luk

The Application of Visual Information into Freeway Operating Speed Prediction Modeling **465**
Zhiqing Yang, Xiaoli Du, and Zhongyin Guo

Optimizing Passenger Train Services: A Model and a Genetic Algorithm Solution **476**
Jing Teng, Baoming Han, and Winggun Wong

Two Proposals for Solution to Transport Problems in Wuhan **492**
Qi Cheng, Akio Kondo, Akiko Kondo, and Hideaki Shimada

Optimal Coordination of Artery System Based on Modified Cell Transmission Model and Particle Swarm Algorithm **500**
Zhengwu Wang, Dayong Luo, and Zhongxiang Huang

A Modified Heuristic Search Algorithm for Pedestrian Simulation **512**
Peng Gao, Ruihua Xu, and Xiaolei Zou

Signalized Intersection Design and Operations: Experiences in Germany and Japan **520**
Keshuang Tang and Hideki Nakamura

Analysis of Railway Orientation in the Freight Express Service Network **530**
Shuai Chen, Youwang Sun, and Zhihong Yin

An Improved Ant Colony Algorithm for UE Traffic Assignment Problem **542**
Ruijun Qiao, Xiaoning Zhu, and Yanhu Liu

Transit Priority Oriented Signalised Junctions—Design and Realisation in Geneva, Switzerland **549**
Jaubert Jianan Tan

Integrative Architecture of ATMS for Urban Elevated Expressway Network in Shanghai **566**
Yuchuan Du, Wenping Yuan, and Lijun Sun

Capacity at a Signalized Intersection Approach with Short Lanes Related to the Cycle Length **575**
Jinyong Jiang and Peikun Yang

Optimization of Signal Timing by Equilibrium Technique **583**
Ning Wu

Simulation Research about the Number of Lines in a Passenger Rolling Stock Servicing Yard on ARENA **596**
Lixin Zhou and Xiongbin Yang

Serve Sequence Optimization Approach for Multiple Bus Priority Requests Based on Decision Tree **605**
Wanjing Ma and Yu Bai

Improving China Traffic Signal Operations through Signal System Retiming **616**
Edmond Chin-Ping Chang

Route Diversion Probability Model Based on Guidance Utility and Its Application **627**
Meiping Yun and Shoupeng Tang

ITS Architecture for EXPO 2010 ..... **637**
Yu Bai and Kun Xue

Can We Predict Road Safety by Simulation? **649**
Henk J. van Zuylen

Mechanism of Dynamic Multiple Ant Colony Optimization for Mass Service Routing and Assignment in Urban Rail Transit Stations **660**
Xiao-lei Zou and Rui-hua Xu

A Study of Parking Behavior and Parking Information Requirements in Shanghai CBD **673**
Yikui Mo, Baoyu Zhang, and Kefei Yan

Exploring Driving Simulator as a Traffic Safety Test Tool at Signalized Intersections **683**
M. Abdel-Aty, X. Yan, E. Radwan, X. Wang, and X. Chen

Impact of ETC on Traffic Safety at Toll Plaza 695
Jiping Ding, Fan Ye, and Jian Lu

Infrastructure/Pavement

Piecewise Optimum Delay Estimation for Intersection Control 702
Chun Shao and Ping Yi

Performance Evaluation of Stone and RAP Interlayers in Soil-Cement Pavements under Accelerated Loading 712
Zhong Wu, Zhongjie Zhang, and Louay N. Mohammad

Asphalt Pavement Statistical Temperature Prediction Models Developed from Measured Data in China 723
Lu Jia, Lijun Sun, and Yonghua Yu

Relationship between Fractal and Road Performance of Asphalt Mixture 733
Rui-hua Yang, Zhi-hong Xu, and Chao Zhang

DCP for Low Volume Road Design/Construction 743
Shie-Shin Wu

Laboratory Study on Ultraviolet Aging Behavior of Asphalt Binder and Mixture 752
Yiqiu Tan, Jiani Wang, Zhongjun Xue, Zhongliang Feng, and Huining Xu

Reasonable Spacing Model of Chevrons 760
Shunying Zhu, Wuxiong Peng, Qinmei Sun, Yuan Xu, Juxiang Guan, and Hongli Dong

Asphalt Rubber and Its Potential Use in China 776
Terrie Bressette, Haiping Zhou, Anne Stonex, and R. Gary Hicks

Laboratory Evaluation of Performances of Asphalt Rubber and Gap Graded Mixtures 786
Rongji Cao and Qifeng Bai

New Technical Method for Production Homogenous and Storage Stable Polymer Modified Asphalt 799
Ai Al-Hadidy and Yi-qiu Tan

Perpetual Pavements: Applying the Approach to Very Heavy Loads 811
Xishun Zhang, Gerald A. Huber, David H. Timm, and Yongshun Yang

Experiment on Efficiency of Nano Titanium Dioxide Photocatalytic Materials Degrading Factors of Automotive Emission 825
Qiang Guan, Zhanyu Wang, and Haiying Bai

Discussion of Coverage for Airport Asphalt Pavement Design 833
Hongduo Zhao and Jianming Ling

Dynamic Moduli Test for M-E Design of Asphalt Pavements 841
Zhanping You, Sanjeev Adhikari, Shu Wei Goh, and Qingli Dai

Laboratory Study of Fatigue Limit for LSM Base Mixture 851
Zhixiang Zhang and Qi-feng Bai

Compaction and Testing Method of Wind-Blown-Sand Subgrade 858
Ming Xue, Dandan Song, and Jun Zhang

Micromechanical Finite Element Models for Micro-Damage and Complex Constitutive Behavior of Asphalt Mixtures **867**
Qingli Dai and Zhanping You

Analysis of Strain in Asphalt Pavement Using FRP-OFBG Sensors **877**
Fengchen Chen, Yiqiu Tan, Hao Liu, and Baoxin Wang

Indexes

Subject Index **887**

Author Index **889**

BRIEF INTRODUCTION OF THE SHENZHEN URBAN TRANSPORT SIMULATION SYSTEM

YANG, Dongyuan[1]. CHEN, Xiaohong[1]. LIN, Qun[2].LIN, Hangfei[1].GUAN, Zhichao[2].
[1] School of Transportation Engineering of Tongji University, Shanghai, 200092, China, Tel.: 86-21-65982127, 86-21-65982274, chenxh@mail.tongji.edu.cn
[2] Shenzhen Urban Transport Planning Center, Shenzhen, Guangdong, 518034, China, 86-755-83949462, linq@sutpc.com

Abstract: Transport systems of most Chinese cities are in a large-scale construction period. To make the decision of planning and construction more scientific and to make urban transport management finer is the chief task of Intelligent Transport System (ITS) construction in the big cities of China. "Shenzhen Urban Transport Simulation System" is based on a multi-source data collection, taking the transport simulation technology and the model as a core, making the intelligent transport shared information platform as an information exchanging and sharing pivot, and taking the support of transport planning, decision-making analysis and transport information service as a direction.

Keywords: *Dynamic Traffic Surveillance and Information, Comprehensive Transport Information Platform, Transport Planning*

1 Background

Transport systems of most Chinese cities are in a large-scale construction period. Facing the rapid development of urbanization and motorization, local authorities attempt to solve and prevent transportation problems through the planning of a transport system, the construction and management of transport infrastructure, and so on. This characteristic has decided that the construction and development of an Intelligent Transport System (ITS) in China should have an advanced pattern and a development path which are different from those in developed countries. ITS is regarded as an important method of enhancing the operating efficiency of existing facilities, improving the transport environment, and promoting traffic safety by using advanced technical methods (1). Considering the current infrastructure condition and transport technology capability in China, ITS provides technique and decision-making support in the transport planning process in order to make the decision of planning and construction more scientific and to make urban transport management finer. This is the chief task of ITS construction in the big cities of China.

ITS is an important part of the urban information basic infrastructure and the handhold of implementation of urban transport strategy. It provides technology support for planning construction decisions. The design target and functional localization of "Shenzhen Urban Transport Simulation System" (SUTSS) are

supporting the intelligence of urban management, including providing support for the whole process of urban transport planning, urban transport construction and transport system operation. The key achievement of SUTSS is using existing traffic information resources (taxi GPS data) and low-cost data collection, integrating the advanced computer and network technology, and then establishing an urban comprehensive transport information platform which is facing public users, decision makers and transport technical personnel. SUTSS is based on the multi-source data collection, taking the transport simulation technology and the model as a core, making the intelligent transport shared information platform as an information exchanging and sharing pivot, and taking the support of transport planning, decision-making analysis, and transport information service as direction.

2 System Compositions

SUTSS makes use of advanced information system integration technology, intelligent transport system technology and transport simulation (macro-scope, meso-scope, micro-scope) technology. It directs the key target of providing decision-making support to the transport planning, construction plan making, and transport system renovation. It adapts the long term development request of ITS construction. This system integrates standardization and mature reliable technology, guarantees data compatibility, the versatility and the scalability, satisfies the resource sharing and data exchange requests, reserves Shenzhen and Hong Kong inter-areas transport shared information platform interface, and provides the communication service for related systems. SUTSS is composed of "one network, four platforms" as illustrated in Figure 1 (2):

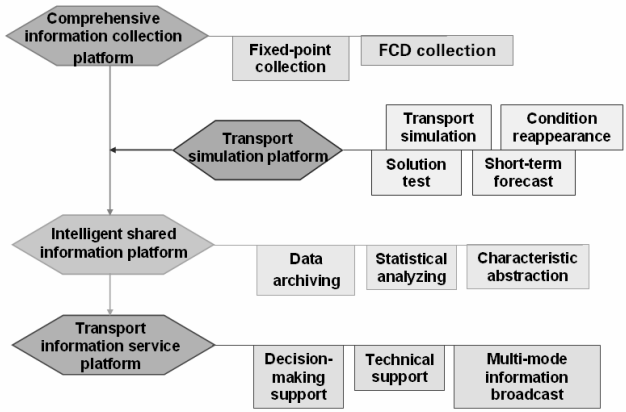


Figure 1. System composition

A comprehensive information collection platform gathers the position and speed information of about 5000 taxis of several taxi companies. These data make about 4.2 millions records every day. The real-time road information data covers more than 70% of the roads in the Shenzhen area. It uses 67 sets of radar monitor equipment (wireless data transmission, solar energy power supply), monitoring the traffic parameters (volume, occupancy, speed per lane per vehicle type every sixty seconds) at the pass, screen-lines and the main traffic corridors.

Based on the Oracle database and spatial database, intelligent shared information platform archives, statistical analyses, and characteristic abstracts, manage and maintain the dynamic, static and historical data, and is the exchange pivot of the whole system.

A transport simulation platform takes the mature simulation software as the core, fuses real-time data and historical data. It realizes the integration of the models of four levels: strategic, macroscopic, meso-scopic and microscopic.

A transport information service platform provides the information service needed by the public, technical personnel, and decision makers through website, plasma display panel (PDP), personal digital assistant (PDA) and web-based intranet applications.

Figure 3. PDP in public place

Figure 2. Web-site for real-time road information -

3 System Functions

SUTSS provides eight function modules, facing three types of users.

3.1 Public users

Public users can go over the traffic map in the Shenzhen area, inquire the name, mileage, real-time speed of links through website, PDP and mobile terminal devices. The following three functions are realized: real-time road traffic condition examination, route searching optimization, and traffic development trend analysis. These functions help the public know the traffic conditions, especially the congestion links, so that they can choose the best route and improve the quality of the trip. These

functions also help them to avoid the current congestion links, prevent the spread of the congestion area, and provide the information basis for trip optimization as well as transport pattern alternation.

3.2 Transport technical personnel

Users can ask for thousands of combinations of inquiry factors about time, space, links, and traffic parameters. They can also get abundant graphs, including the variation of traffic parameters by hour and by week. The realization of the system takes the place of a large-scale artificial traffic survey, saves survey and data processing costs greatly, enhances the transport data use efficiently, and provides powerful data support to transport planning and management. It provides feasible traffic condition inquiry, standardized demonstration of transport simulation and measured data calibration of models, which provides the technical safeguard system for the reliability of transport planning and design.

SUTSS realizes nine kinds of simulation model analysis and design templates and graphs, according to the planning process; realizes the fusion of dynamic data monitoring and simulation models (the simulation model parameters can be calibrated); realizes the share of simulation tools and achievement under the simulation foundation data security conditions, enhancing the reliability of simulation model quality.

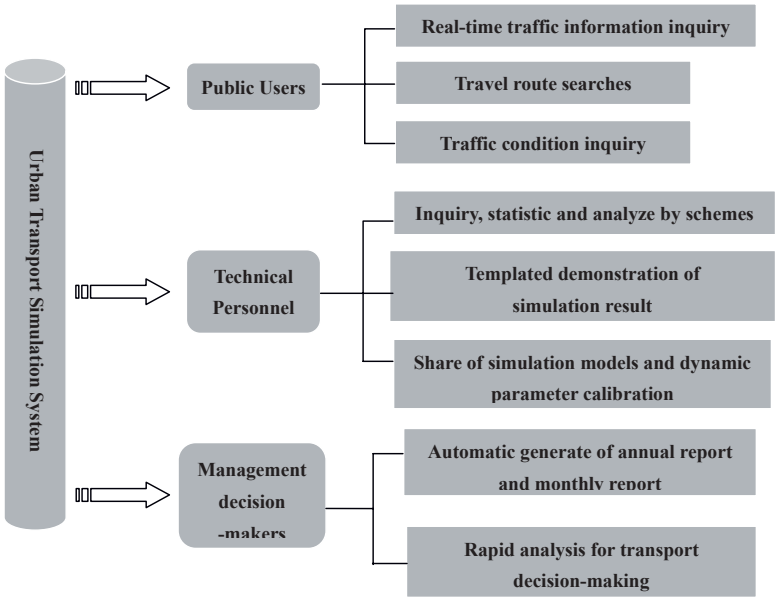


Figure 4. System functions (3)

3.3 Management decision makers

SUTSS realizes the automation and intelligence from data collection and analysis to performance evaluation and decision-making support. It provides an annual report and monthly report generation system and a traffic condition rapid analysis method, including the comparative analysis of traffic condition indices by time, space, topic, condition, etc. It manages and maintains the dynamic transport data, which makes transport work in different departments use the same analysis source, ensuring the coordination of business analysis and decision in order to form the technical method that uses dynamic quota data as the support of planning decision-making. By using real-time transport data in transport decision-making, the system obviously improves the timeliness of the traffic report and the reliability of data, evaluates in full scale the benefit of significant projects, and improves the transport decision-making quality.

4 System Characteristics

SUTSS adjusts the traditional transport system's open interrupting mode to a control process of continuously observing, data accumulating, continuously learning and close feedback. It demonstrates the ability of system recognition, condition and development reappearance, trend analysis and forecast analysis. It creates a wholly new conception of a decision-making support method, promoting the urban transport construction and management to a newly high level.

4.1 Innovation in design concept

By introducing dynamic transport information in the urban transport planning and design, SUTSS establishes the intelligent simulation platform which integrates the strategic, macroscopic, meso-scopic, and microscopic simulation models. By using an intelligent simulation technical method in urban transport planning and design, performance evaluation, transport solution selection and transport construction decision-making, it improves the scientific character of urban transport planning and design decision-making as well as the whole efficiency. It lays the technological base for the urban construction and the transport planning decision-making process optimization.

4.2 Innovation in planning process

By changing the traditional close and static method in transport planning, it establishes the new transport planning technology based on real-time and continuous transport information collection and processing, realizing the modernization and automation of basic data collection. SUTSS has not only used intelligent technical methods in survey, solution, and decision-making, but also changed the artificial evaluation pattern in the decision-making period. By introducing an intelligent simulation platform concept, it realizes the quota and a scientific transport planning solution. By using advanced Web-GIS technology, it realizes multi-modes transport

information broadcast, providing information for transportation enterprises and public, leading citizen's reasonable transport behavior and optimizing transportation structure.

5 Conclusions

SUTSS is a transport comprehensive information service system which is adapting the gradual characteristics of the development of Chinese cities.. Its design concept, system structure, function design and key technologies are applicable to medium-sized and larger cities in China for their transport information service system construction and intelligent public transport system. In cities with taxi or other vehicle location data, the basic functions of the system could be realized in about six months with about 10-20 million RMB investment, through data comprehensive utilization, without adding collecting equipment, so that it can provide road condition of road network and travel information, which has good promotion prospects.

The accomplishment and application of the SUTSS project, whose main content is multi-source real-time data collection, planning decision-making support, transport simulation and information service, have obtained a breakthrough in aspects of timeliness of services, technical advancement, diversity of form, and wideness of service areas. It has great significance in ameliorating urban traffic congestion, reducing urban environment pollution, and promoting urban ITS construction.

References

Wang, XJ, Qi, TY and Cai, H, (2004). "Principle and Application of System Architecture for The Intelligent Transport Systems." China Rail Publication House, Beijing.

Shenzhen Urban Transport Planning Center, Tongji University and Shanghai Baosight Software Corp. (2006). "Detail Design Specification of Shenzhen Urban Transport Simulation System." Shenzhen, 2006.6.

Shenzhen Urban Transport Planning Center, Tongji University and Shanghai Baosight Software Corp. (2006). "Technical Research Report of Shenzhen Urban Transport Simulation System." Shenzhen, 2006.11.

Estimating Congestion Index at the Link Level with TransCAD GIS

M. Zhang[1] and T. Lomax[2]

[1]Community & Regional Planning, The University of Texas at Austin, 1 University Station, Stop 7500, Austin, TX 78712, USA; PH (512) 471-1922; FAX 512-471-0716; email: zhangm@mail.utexas.edu

[2]Texas Transportation Institute, College Station, TX-77843-3137, USA; PH (979) 845-9960; FAX (979) 845-9356; email: **t-lomax@tamu.edu**

Abstract: This paper presents a GIS-based procedure to estimate congestion index (CI) at the highway link level. Using the 1995 traffic assignment data for the Lubbock, TX area, the exercise produces CI values at the network link level. These disaggregate measures of CIs are then presented in two summary forms. One is the statistical summary of aggregating link-level CI values by area type and by highway functional classes. The other is a visual summary with maps on which network links in the study area are grouped and shaded by five congestion levels measured by the CI. Benefits and limitations of using TransCAD GIS for CI studies and the like are discussed.
Key words: *Congestion Index, GIS, Link-level, Texas, Programming*

1 Introduction

Roadway congestion is a chronic problem facing metropolitan areas across the United States. To closely monitor congestion conditions and trends of change, federal and state transportation agencies have sponsored multi-decade research projects to report on an annual basis of regional congestions (Texas Transportation Institute, 2005). While the reports have portrayed big pictures of metropolitan mobility performance in over 80 metropolitan areas, they do not provide detailed information of intra-regional congestion conditions. Congestion in general displays a highly concentrating pattern in space and time. To more effectively tackle with congestion-related problems, it is desirable to know where and when the congestion occurred or likely to occur based on the forecasted future travel demand. With this spatial and temporal congestion information, the policy maker can thus allocate the limited resources for most effective solutions to congestion mitigation.

This paper presents a GIS-based procedure to evaluate link-level roadway congestion. The procedure takes assigned traffic volumes of regional travel demand model as input, calculates congestion index (CI) for each highway link, and visualizes congestion levels in maps and in aggregate tables. The study presented here uses

Lubbock, Texas as an example for demonstration purposes. Transportation planning agencies in other metropolitan areas may customize the GIS tool and apply it to monitor existing roadway congestion conditions or to examine the congestion relief effects of future transportation infrastructure investments.

Structure of this report is as follows: Part 2 discusses a number of conceptual issues pertaining to congestion measurements. Part 3 describes the method developed for congestion measurement for this study. Part 4 presents the results obtained from a case study of the Lubbock area. Finally, Part 5 discusses the advantages and limitations of using TransCAD GIS for CI modeling.

2 Conceptualizing Congestion

Congestion by the dictionary definition refers to the state of an abnormal or excessive accumulation. When traffic accumulates on a highway segment, traffic density increases (density being defined as the number of vehicles per lane per mile). Continuing increase of traffic density forces the vehicle flow to slow down because presence of other vehicles limits the driver from driving freely, i.e., at the free-flow speed. When density increases to a point where vehicles are literally bumper to bumper, the traffic flow breaks down, reaching a state of jam-density and maximum congestion.

One difficulty in measuring congestion is to determine the break point between a normal accumulation and an excessive or congested accumulation of traffic. Traffic engineers set a threshold ratio between traffic volume and the capacity of a facility. Congestion occurs when the traffic volume exceeds the capacity. Based on this conception, congestion may be measured in terms of traffic volume, for example, the volume-to-capacity ratio. A more intuitive measure is based on travel speed or travel time, because the direct consequences of congestion are speed reduction and travel delay.

Several characteristics of congestion make congestion measurement a rather complicated task.

1) Congestion in most cases is specific in time and location. Congestion occurs largely during the morning and afternoon peak-periods, although there are increasing trends of midday congestion and of the spreading of morning and afternoon peaks in many communities. Location-wise, congestion occurs in particular facilities of highway network. The facility may be a lane segment in the highway network, an intersection, an on- or off- ramp, or an entire commuting corridor. To better understand and closely monitor congestion conditions, it is important to have a relatively precise measure of congestion at the roadway link level.

2) Level of congestion relates directly to the actual performing capacity of a facility, not necessarily to the design capacity of the facility. While the design capacity of a facility may be determined by design manuals, many natural and human factors may reduce the facility's performance to a great extent. For example, the design capacity of a freeway link may be 2,200 vehicles per hour per lane. In a raining or snow day, however, the facility's capacity will decrease significantly. The capacity of the same type of facility also varies spatially due to the influence of geographical, geometric, and many contextual factors affect the performance of the facility. For instance, a freeway link in the CBD area typically has a smaller carrying capacity than that in the rural fringe area due to such factors as spacing of exit/entrance points, shoulder and median width, curvature and slopes, and pavement and tunnel lighting conditions.

3) Congestion evaluation contains both the objective and subjective elements. People in cities with different sizes or in different parts of the same city tend to have different view of congestion. Because congestion usually comes hand-in-hand with economic and social activities, congestion is often interpreted as an evidence of economic and social vitality (Taylor, 2002). Therefore congestion may never be totally reduced due to the 'laws of convergence' (Downs, 2002). The concept of 'Optimum Congestion' captures the objective and subjective characteristics of congestion (Small, et al. 1989, Stopher, 2004).

4) The optimum congestion concept is a combination of factors including expectations of the residents, businesses and travelers in the area and realistic assessments of the solution alternatives. Optimum congestion also allows for differences between congestion at a place such as a major job center, and congestion between activity areas or between job centers and homes.

    - Job center congestion is the result of activity density. It typically covers a smaller area, imposes a shorter travel time penalty and in most cases has a variety of solutions including walking, biking, shuttle vehicles, schedule changes, etc.
    - Congestion "between" areas imposes a more significant travel time penalty and brings with it unreliable travel conditions. The congestion causes both travel time penalties and business or personal decisions that have a significant affect on the economy, land use patterns and social structures.

5) Coverage and effectiveness of such practice as Travel Demand Management (TDM) and system management have strong implications to congestion. In general, there are two types of congestion—recurring and non-recurring (Lomax, 2003a). The former occurs at the same place and the same time recursively, especially on weekdays. Recurring congestion results from the peaking effect of travel demand and from the bottleneck effect of facility supply. Non-recurring congestion arises from temporary conditions, such as a

vehicle breakdown, accident, or temporary road works. Many cities offer a variety of TDM programs and system management services, for example, telecommuting, ridesharing, ramp metering, traffic signal coordination, and incident management. These programs and services help reduce peaking or bottleneck constraints and thus reduce level or duration of congestion. In measuring congestion, they should be taken into consideration.

3 Methodology

Based on the above conceptualization and consideration of factors pertaining to congestion, a measurement method is developed. The method includes two major components. One is the formulation of congestion index (CI) for measuring levels of roadway congestion. The other is implementation of CI using GISDK, a scripting language for TransCAD GIS. The following highlight the CI formulation. For detailed explanations, see (Lomax, 2003a, 2003b).

3.1 Congestion Index (CI)

The basic formula for calculating CI is given below in Equation (1):

$$CI = \frac{t_c}{t_f} = \frac{t_f + t_d * (1 - r)}{t_f} \tag{1}$$

t_c: Estimated travel time under the congested condition;
t_d: Travel time delayed resulting from congestion;
t_f: Travel time at the free-flow speed;
r: Travel time reduction benefit (%) resulting from the congestion improvement techniques

A number of subsequent equations detail ways in which these travel times are estimated. Note that times are measured in vehicle hours of travel, unless specified otherwise. Specifically, travel time in the free-flow speed is given by the ratio of vehicle miles of travel (v) over free-flow speed (s_f) on the link.

$$t_f = \frac{v}{s_f} \tag{2}$$

Similarly, travel time under the congested condition is given by the ratio of vehicle miles of travel (v) over congestion speed (s_c) on the link.

$$t_c = \frac{v}{s_c} \tag{3}$$

The vehicle miles of travel for the congested hours are calculated from the assigned traffic volume multiplied by link length and adjusted by time-of-day factor. The free-flow speed is input from the user and varies across different roadway functional classification and area type for different cities.

Congestion speed is estimated using a speed-delay function (Bell, et al. 2003).

$$s_c = \frac{60}{\frac{60}{s_f} + \alpha * e^{\beta(v/c)}} \tag{4}$$

where α and β: constants, typically taking values of 0.15 and 4.0, respectively;
c: link capacity
e: the base of natural logarithm

Travel time delay includes two components: recurring delay t_r and incident delay t_i. Define that

$$t_r = t_c - t_f \tag{5}$$

$$t_i = (t_c - t_f) * \theta \tag{6}$$

where θ is incident factor, indicating the likelihood that incidents occur in the network system. Therefore,

$$t_d = t_r + t_i = (t_c - t_f) * (1 + \theta) \tag{7}$$

3.2 Implementation of CI in TransCAD GIS

Implementation of CI in TransCAD GIS involves the following three steps.

- Data assembly

Measuring link-level congestion with the CI method described above needs input data in quite different spatial and temporal characteristics. Some are network-based, for example, traffic volume and network capacity. Others, for instance, the coverage of incident management are area-based. Still others such as V/C ratio cut-off points for congestion classification are generic throughout the city. Time-of-day specifies the time period for which CI is calculated, mostly focusing on the peak period or peak hours. TransCAD GIS, like most other GIS packages, offers a relational database management tool to relate these data tables on common ID fields. Figure 1 below illustrates the way in which data tables are organized and

assembled for the CI study. Each box represents a data table. The core data table contains the traffic assignment output, which is part of the attribute table of the GIS spatial data (not shown here). Other data tables relate to the core table through the corresponding common ID fields.

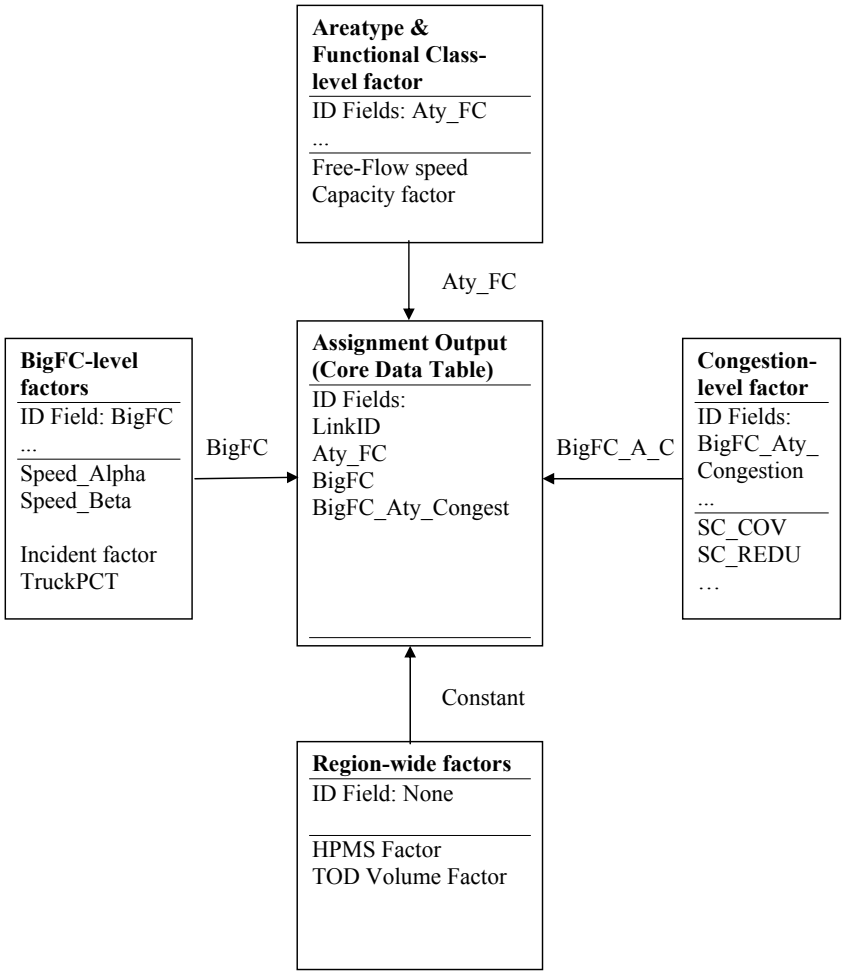


Figure 1. Organization of Input Data for CI Modeling in TransCAD

- Programming of CI calculation

The main task of this working step is to develop a GIS-based interface for the user to input city-specific factors and parameters. When the user completes data input, TransCAD calculates CI values for each roadway link in the study area and stores the results in a separate table that is also related to the network file. For this project CI programming utilizes Caliper Script, a scripting language of GISDK that comes with TransCAD GIS (Caliper Corp, 2000). The final program contains a series of macros and dialogue boxes for specific input-reading and intermediate CI calculations. It has approximately 1,800 lines of coding and therefore is not included in this paper.

- GIS mapping and table presentation

This step is user-specific. It utilizes TransCAD built-in features to generate maps, tables, and other forms of presentations. The user may choose to create thematic maps to visualize spatial distribution of congested roadway links in the study area, to highlight congestion 'hot-spots' in specific locations, or to report average congestion levels as measured by CIs with various methods of data aggregation, for example, by roadway functional class, by area type, and by time of day.

4 An application example: Lubbock, TX

This section presents an application example of the GIS-based congestion measurement procedure in Lubbock, TX. Applying the procedure to other regions follows the same steps.

The Lubbock metropolitan area is located in northwest Texas, with a total population of approximately a quarter of a million in year 2000 (The Census Bureau, 2007). Compared with other metropolitan areas in Texas, Lubbock has a relatively good traffic condition region-wide. The travel modeling output data needed for this study are obtained from the TxDOT. The Lubbock metropolitan planning organization (MPO) provided data on 24-hour free-flow speeds and capacities.

To run the procedure, Lubbock's network file should be opened first in TransCAD GIS. The procedure is programmed as an Add-In, a built-in customization tool of TransCAD. When the procedure name "TCI Macro" is specified, a number of interfaces/windows will pop up, asking the user to provide input factors and base speed and capacity information. The information will be saved in a temporary folder and does not need to be re-entered when the procedure is repeated except for the items modified by the user. After the input is completed, the procedure will generate a set of CI estimates, store them in a separate table, and link the table with the network attribute table. All the calculations are carried out at the link level. The user decides the aggregation level in which the congestion measurement is summarized and reported. For illustration purpose, this exercise reports only the AM and PM peak hour average CI values. Specifically, the average hourly traffic volumes between 6:00-8:00

in the morning and between 4:00-6:00 in the afternoon are used as the initial input data for calculating the AM and PM CI values.

Table 1 reports two sets of CI estimates that are aggregated by area type (row) and by highway functional classes (column). One is for passenger car mode in the morning peak, whereas the other for car and truck combined average of morning and afternoon peaks. The average CI value for a given area or a given functional class is shown as the marginal mean. The figure in the lower right cell indicates the region-wide average TCI. Any cell with "--" means that there is no data for CI estimate for that specific highway class in that specific area type. Note that these aggregate TCIs are the arithmetic means of the link-level values. An alternative way to aggregate the TCIs is to obtain the VMT-weighted means.

Interpretation of the results is straightforward. Take the CI modeling output for the AM peak for car as an example. Region-wide, Lubbock has a CI of 1.07, meaning that in 1995 the average travel time in the congested condition during the AM peak was 7% longer than in the free-flow condition. For the same time period, the CBD-fringe and other urban areas (AreaType 2 and 3) appeared to be relatively the most congested areas in the region (CI=1.09). Of the seven functional classes with CI estimates (excluding Class #0, which is the Centroid Connector), principle arterial (Functional class #4) was the most congested highway class, with average travel time in the congested condition 19% longer than in the free-flow condition.

Similar congestion patterns can be observed for the PM peak period (not shown here). Overall, the results indicate levels of congestion in the Lubbock region that are consistent with the common knowledge about the region.

The set of CI matrices presented above in Table 1 show statistical summaries of congestion conditions by different area type and functional classes. Another way to summarize the results is through visualization of the CI values. Network links can be colored or shaded with different color schemes or gray tone scales based on the magnitude of the link-level CI values. Figure 2 shows an example, mapping out the link level CI estimates for the Lubbock area. The map shows that congestion occurred mostly in the southwest quadrant of the region. Severe congestion is evident in approximately ten intersection areas, with congested travel time more than twice longer than the free-flow travel time. The CI pattern displayed on the map highlights the locations or links that are most needed in terms of congestion reduction and management.

Table 1. Summary Congestion Index by Area Type and by Roadway Functional Class, Lubbock, TX

Congestion Index Estimates for Passenger Car: AM Peaks

	Functional Class	F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	MEAN
Area Type	**1**	1.00	--	1.09	--	1.19	1.09	--	1.03	--	1.05	*1.07*
	2	1.00	1.13	1.05	--	1.27	1.11	1.03	--	--	1.04	*1.09*
	3	1.00	1.14	1.10	--	1.23	1.12	1.03	1.04	--	1.04	*1.09*
	4	1.00	1.07	1.07	--	1.11	1.08	1.05	1.04	--	1.03	*1.06*
	5	1.00	1.05	1.06	--	1.15	1.10	--	1.03	--	1.03	*1.06*
	Mean	*1.00*	*1.10*	*1.07*	--	*1.19*	*1.10*	*1.04*	*1.04*	--	*1.04*	*1.07*

Congestion Index Estimates for Car and Truck Combined: Average of AM and PM Peaks

Area Type	**1**	1.00	--	1.12	--	1.31	1.09	--	1.04	--	1.06	*1.10*
	2	1.00	1.23	1.06	--	1.50	1.13	1.03	--	--	1.05	*1.14*
	3	1.00	1.23	1.14	--	1.37	1.14	1.03	1.05	--	1.06	*1.13*
	4	1.00	1.10	1.09	--	1.15	1.09	1.07	1.05	--	1.04	*1.07*
	5	1.00	1.05	1.07	--	1.20	1.11	--	1.03	--	1.03	*1.07*
	Mean	*1.00*	*1.15*	*1.10*	--	*1.31*	*1.11*	*1.04*	*1.04*	--	*1.05*	*1.10*

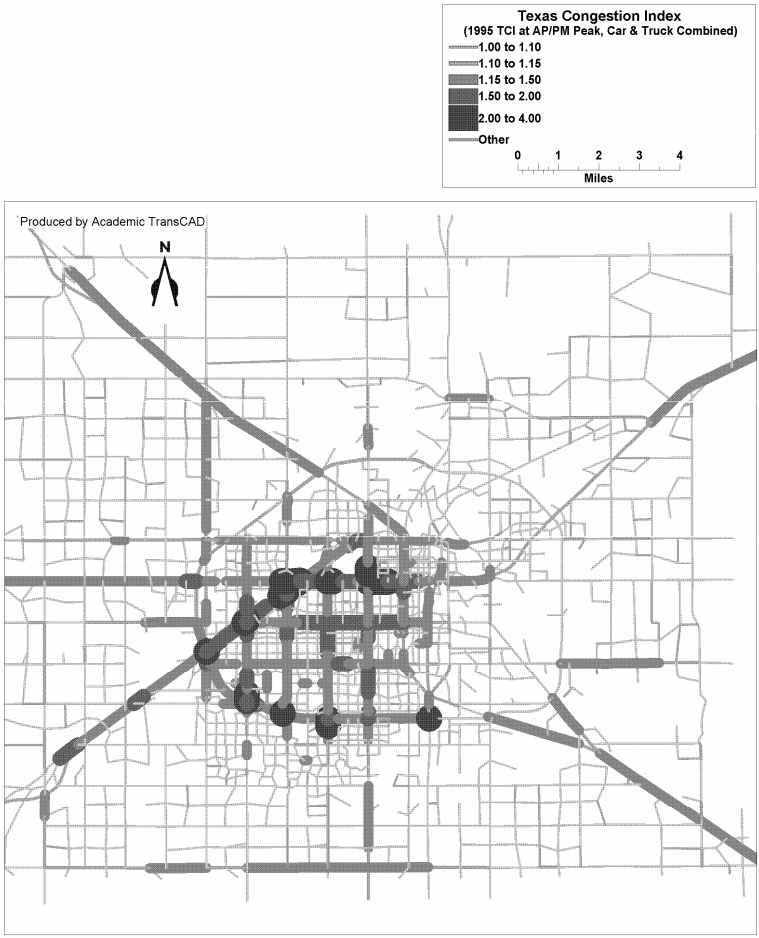


Figure 2 Visualizing CI Values in TransCAD for Lubbock, TX

5 Summary and Conclusions

This paper presents a GIS-based procedure to measure congestion at the highway link level, portraying a detailed image of congestion conditions in the study area. The results can be effectively presented at least in two ways. One is in maps to visualize the link-level CI. Illustrating congestion conditions in different color

schemes or gray tone scales, the CI maps create visual summaries of network performance in the study area. They are intuitive and informative. From the map, 'hot spots' with severe congestion (i.e., very high CI values) can be easily identified in the network context, which helps the analyst and policy makers focus their attention to the most resource-needed portions of the network.

Another way to present the disaggregate CI results is to summarize statistically the link-level values to the chosen level, for example, by area type, functional class, county, or the entire metropolitan area. Compared with the mapping approach, this approach provides more accurate, quantitative evaluation of congestion conditions throughout the study area. For example, from the CI values, one can find out the extent to which arterials in the CBD are more congested than those in the suburb. Of course, these two approaches can be used jointly, with each serving specific purposes.

The procedure presented here is not free of limitations. A critical limitation of TransCAD (in fact, of all GIS of the current generation) is its incapability to handle temporal data. For example, should CI values be produced for each hour of a day, the procedure presented here would have to be repeated 24 times! This is due to the limitation of GIS data structure. Additional programming would be necessary to perform time-of-day CI analysis (TransCAD has a built-in time-of-day subroutine to split trip O-D matrix from daily volume to each of the 24 hours). Furthermore, operating TransCAD requires substantial technical training. Purchasing and maintaining the software demands a significant amount of resources.

References

Bell, C., Boardman, M. and White, L.D. (2003). *Texas Mobile Source Emissions Software User's Guide: Mobile6*. Texas Transportation Institute: College Station.

Caliper Corp. (2000). *TransCAD (4.7 Academic Version) User's Guide*. Caliper Corp.: Newton, MA.

Downs, A. (2002). *Still Stuck in Traffic: Copying with Peak-Hour Traffic Congestion*. The Brookings Institution Press, Washington, D.C.

Small, K., Winston, C., and Evans, C. (1989). *Road Work: A New Highway Pricing and Investment Policy*. The Brookings Institution Press, Washington, D.C.

Lomax, T. (2003a). *Texas Congestion Index Measurement Concept*. Retrieved on August 16, 2003, at URL: http://mobility.tamu.edu/tci/ .

Lomax, T. (2003b). *Texas Congestion Index Calculation Process/Detail*. Retrieved on August 16, 2003, at URL: http://mobility.tamu.edu/tci/ .

Stopher, P.R. (2004). "Reducing road congestion: a reality check." *Transport Policy*, 11: 117-131

The Census Bureau (2005) retrieved on July 30, 2005 at URL: http://factfinder.census.gov/

Taylor, B. (2002). "Rethinking traffic congestion." *Access*, 21: 8-16.

Texas Transportation Institute (2005). *Urban Mobility Study*, http://mobility.tamu.edu

PLANNING BICYCLE CORRIDOR FOR SHANGHAI CENTRAL CITY

Xiaohong CHEN [1], Wen XIONG [2] and Zhaoyi HUANG [3]

[1] Professor, Tongji Univ., 1239 Siping Road, Shanghai, 200092, P. R. of China,
Tel) +86-21-65982127, Fax) +86-21-65982274, Email) tongjicxh@163.com
[2] Ph.D Candidate, Tongji Univ., 1239 Siping Road, Shanghai, 200092, P. R. of China,
M-Tel) +86-13636460813, Fax) +86-21-65982274, Email) xwart@126.com
[3] Ph.D, Shanghai Road and Traffic Research Center, 200092, P. R. of China,
Tel) +86-21-65981860 Fax) +86-21-65988660, Email) Huang_zhaoyi@lycos.com

Abstract: In Shanghai central city, numerous bicycles are currently used as the most important commuting vehicles. Due to the lack of alternative running-through corridors, many cyclists have commuted along main roads for many years where conflicts between motor vehicles and non-motor vehicles are seriously intensified as a consequence of insufficient road availability and these cyclists are also terribly exposed to traffic pollution with the rapid development of motorization. Therefore, it is necessary to provide more convenient, safer, and continuous bicycle corridor systems for cyclists. This paper proposes the notions of Pedestrian & Cyclist (P&C) Island and P&C Nucleus to remove the division by arterial roads and locate corridors for a main flow of cycling trips. Based on the principles of "Connecting P&C Island", "Passing P&C Nucleus", "Cutting and Straightening", and "Prioritizing then Attracting", the bicycle corridor network planned for Shanghai central city and the tidal management policies designed for different corridors are raised this paper.

Keywords: *Bicycle corridor, Pedestrian & Cyclsit (P&C) Island, P&C Nucleus*

1 Introduction

In 2004, the ownership of bicycle on hand reached 9374751 in Shanghai, while the actual usage degree is only 50%. Despite the fact that the percentage of bicycle trips has been declined from 41.7% in 1995 to 25.2% in 2004, the bicycle still has the superiority of time-saving, convenience, and economical characteristics within distance of 1.5 to 6 km. What's more, the transfer model of "Bicycle Park & Transit Ride" will become an important supplement to transit trips in Shanghai.

Compared with motor vehicles drivers, the traffic violations of cyclists are more common, which results in the prejudice of policy-makers toward bicycles. The policy-makers try to carry out measures to separate the flow of motor vehicles from non-motor vehicles, which means that the parallel branches of main roads should be modified and built to serve as non-motor vehicle passages so that the arterial roads can be used by motor vehicles exclusively. However, cyclists don't like driving on some

non-motor vehicle passages with poor network connectivity, many zigzags, and long-distance round links. During peak time, more or less, some cyclists always ride along the motor vehicles exclusive lane, which brings accident threats and influences the operating efficiency of motor vehicles.

The authors expect to plan more continuous, safe, and environment-friendly bicycle trips circumstances for numerous cyclists by the planning of a Bicycle Corridor.

2 Planning Background

2.1 Deficiency of alternative convenient bicycle lane

In 2004, daily trips by bicycle reached 10.7 million person trips in Shanghai. Figure 1 shows the OD distribution of trips by bicycle among different zones of centre city and suburb.

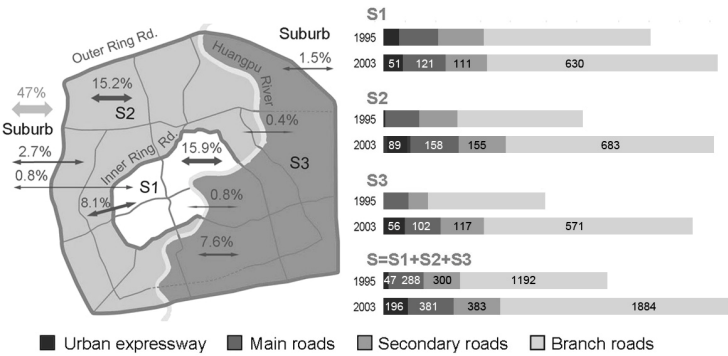


Figure 1. The road length and OD distribution of bike in Shanghai central city

In Zone S1: Since the main roads lack of reconstruction potentials (hardly changed during past 8 years), policy-makers have planned to convert the non-motor vehicles lanes into bus only lanes on the three vertical and three horizontal main roads gradually, and to link and rectify the parallel branches serving as non-motor vehicle passages. However, these imperfect parallel branches bring many kinds of inconvenience for riders. Take Yan'an Road and Lu Jiabang Road where "motor vehicles only" has already been enacted, for example, there are still many cyclists reluctant to detour while riding along the bus only lanes during the peak hours, because those alternative bicycle passages are indirect and even interrupted.

In Zone S2: Though the conflicts between the motor and non motor vehicles are very serious on several main radial roads, those branches which are not well-through are incapable to release transportation pressure on main roads. Large scale patch of

land development like communities, universities usually intensifies the spacing of branches and main roads.

In Zone S3: Arterial network planning is based completely on motor vehicles demand, which results in large spacing between roads. For example, the spacing between South Pudong Road and Yanggao Road is 1.4km to 2.2km, while there is only one branch road between them. Traffic conditions have already been aggravated during peak periods.

Obviously, the reason why a huge numbers of bicycles converged on arterial roads is that alternative convenient bicycle paths are not enough. Therefore, in order to reduce the conflicts between motor and non-motor vehicles on main roads, it is necessary to reconstruct enough branch roads for laying out more continuous and convenient bicycle corridors

2.2 Safety analysis

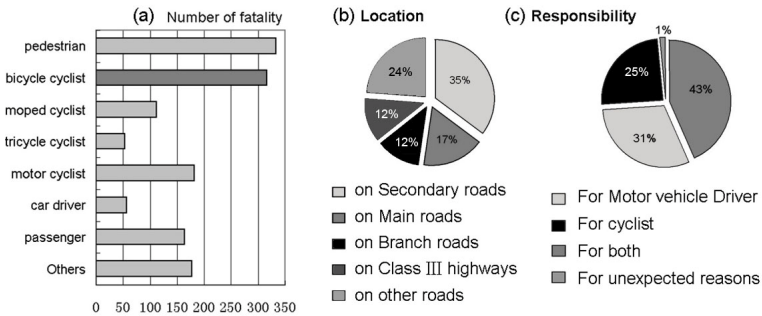


Figure 2. The number of fatalities, road distribution and responsibility constitution of bicycle accidents in Shanghai, 2005

Figure 2 reveals the number of fatalities among all traffic accidents in Shanghai, in which the ratio of cyclists is 23%. More than half of cyclist fatalities occurred on secondary roads and main roads, and the number of deaths on branches is listed as the third one due to the lack of separation facilities between bicycle and motor vehicles. Cyclists bear the responsibility for 25% fatalities, for instance, inappropriately crossing roads caused 39 fatalities, rushing on lanes meant for other vehicles resulted in 22 people's death, and making turns unexpectedly led to 14 people dead. Motor vehicles undertake the responsibility for about 31% cyclist fatalities because of the drivers' improper behavior, such as improperly controlling the engine, depriving lanes meant for bicycles, and drunk driving. Because of the statistic problems, about 43% victims are dead due to either motor vehicles' responsibilities or that of cyclists, including not obeying rules, traffic signal offences, and so forth.

In order to reduce cyclist fatalities, some active measures must be carried out so as to provide relatively safe spaces for cyclists.

2.3 Cycling trips exposed to heavy-polluted road environment on main roads

The bicycle is recognized as a kind of green transportation mode by the virtue of its calming, non-emission, non-energy-consumption, and so on. However, in Shanghai, the riding environment for cyclists on many main roads is startling. The traffic noise is approximately distributed between 70 and 75db during daytime. The noise on the main roads is generally higher than that of branches. The NO_x concentration on main roads is higher than national standards. Under some elevated roads where the air circulation is very poor, motor vehicle emissions (especially diesel smoke) during the peak hours accumulates to become vision pollution.

Therefore, it is necessary to build some low-noise, low-emission bicycle corridors utilizing brancheskeeping the cyclist far away from all kinds of traffic and pollution.

3 Planning Notions and Methodology

3.1 Notion 1: Non-alternative for high-bicycle-volume arterial roads

With the increasing deterioration of the conflicts between motor vehicles and non-motor vehicles, policy-makers expect to separate bicycles on main road completely by modifying and establishing alternative parallel branches, thus achieving the “Motor vehicle only” policy by converting the non-motor vehicle lanes into motor vehicle lanes.

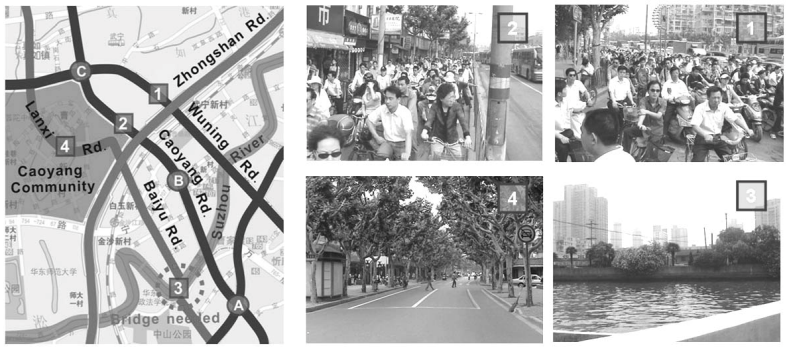


Figure 3. Inappropriateness for Motor Vehicle Only — Case for Caoyang Rd

As shown in Figure 3, Zhongshan Road isan elevated expressway, Caoyang Road and Wuning Road are main roads, Lanxi Road and Baiyu Road are branch roads. If Caoyang Road is selected as the “Motor vehicle only” policy, bicycles have to choose

Wuning Road as a bypass (If Wuning Road also complies with the policy, there would be no way for cyclists to cross the river.), A to B has to bypass 0.9km, and A to C has to bypass 0.6km. If Baiyu Road Bridge has been built, then cyclists may choose Baiyu Road and Lanxi Road as a bypass route, then A to B has to bypass 0.8km, and A to C has to bypass 1.6km. According to the Stated preference surveys in 2005, the acceptable detour distance of most cyclists in Shanghai is 0.25 to 0.37 km. Obviously, above-mentioned detouring distances largely surpass the cyclist's tolerance.

In Shanghai central city, it is not only seen on Caoyang Road that bicycles on main roads can't be shifted away completely because of the high spacing and imperfect shape of road networks. However, the overflowing number of bicycles on the main road can be attracted and transferred partially by bicycling corridors in future. As Figure 3 shows, when Baiyu Road Bridge is built, the bicycle corridor on Lanxi and Baiyu Road will attract lots of cyclists lived in Caoyang Community. As for previous cyclists who use Caoyang Road as commute routes, young people who prefer tranquility and freedom may probably transfer to the bicycle corridor, while those care about physical strength and time consumption can still ride on Caoyang Road.

3.2 Notion 2: Pedestrian & Cyclist Island and Bicycle Corridors

In Shanghai central city, with the development of Mobilization Process, rapidly constructed expressways and main roads, like Huangpu River, completely separated the space for pedestrians and cyclists into pieces. The surging traffic flows cause not only inconvenience, but also noise, emissions and unsafe feelings for both pedestrian and cyclists. Figure 4 shows 24 isolated zones by the segregation of Expressways and Huangpu River. These zones are defined as Pedestrian & Cyclist Islands (P&C Islands). We encoded these zones as A to X. P&C Island which are segmented into P&C Isles by creeks and arterial roads with a smaller scale.

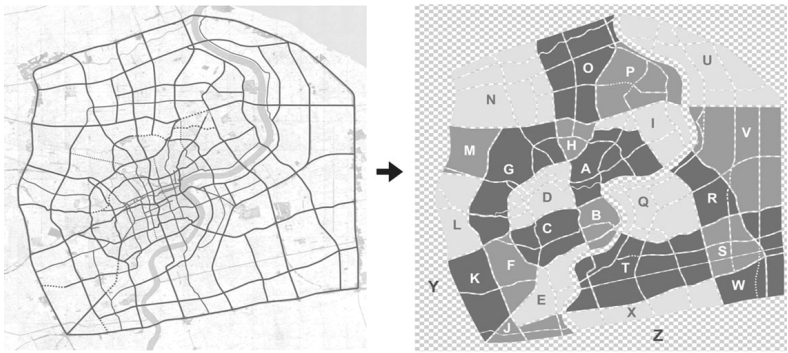


Figure 4. The Distribution of P&C Islands and Isles in Shanghai central city

As for central city, P&C island covers 28.5 km^2 onaverage, where pedestrians and cyclists are insulated for the lack of crossing facilities between P&C islands. P&C isle covers 4.8 km^2 on average, where residents are surrounded by motor vehicles, noise, emission, and dangers. What's more, most cyclists have to ride along the main roads due to the lack of continuous branch roads.

In the light of these statements, we have come up a new planning proposal named "Connecting P&C Island". The main concept is selecting proper branches to build bicycle corridors, connecting all these P&C Islands, and linking most of P&C Isles. Bicycle corridors will make up the framework of bicycle lane networks.

3.3 Notion 3: Pedestrian & Cyclist Nucleus and Bicycle Corridors

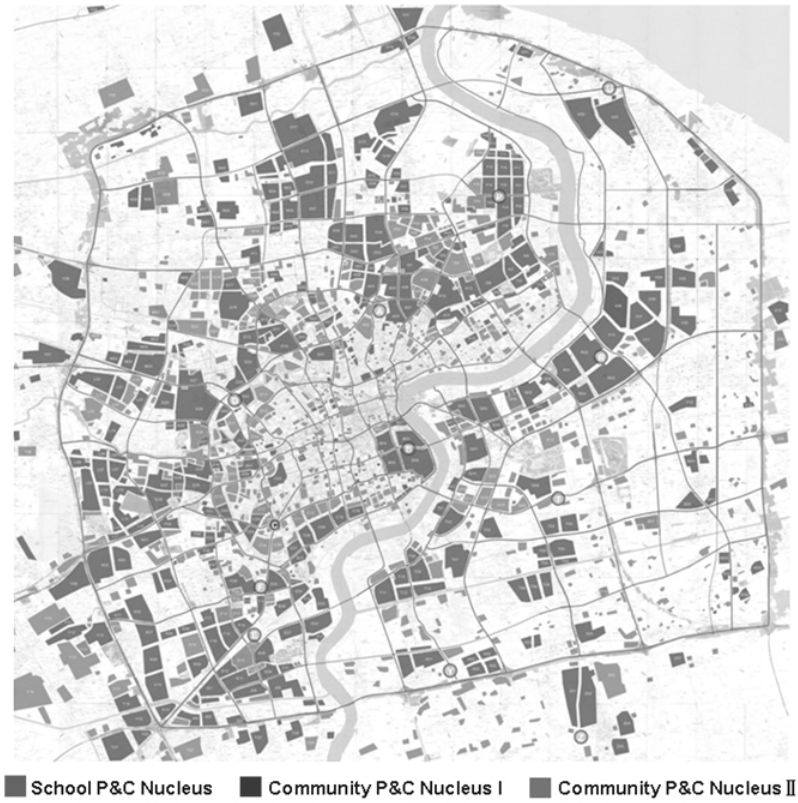


Figure 5. The Distribution of P&C Nucleus in Shanghai central city

Considering the lack of OD data among zones and uncertainty of flow distributions, this paper comes up with the concept of Pedestrian & Cyclist Nucleus (P&C Nucleus), which means high frequency, and high intensive areas of slow mode transport generation and attraction. According to trip purposes, P&C Nucleus can be divided into four classes: School P&C Nucleus, Community P&C Nucleus□, Community P&C Nucleus□, and Commercial P&C Nucleus. Referring to satellite maps, senior citizens inquiries, and field surveys, we identify the scale and distributions of former three classes, which are shown in Figure 5.

In 2004, the ratio of schooling trips was 15% of the total cycling trips (and 16% of that of walking trips). The ratio of commuting trips and pleasure trips in 2004 to total cycling trips are 51% and 11% respectively (18% and 13% of that to total walking trips). The residential clusters constituted of a large number of households which are exactly the origin of these trips.

Figure 5 indicates two types of Community P&C Nucleus referring to the communities data: Community P&C Nucleus □ represented in green, with high-density and low-class residential clusters, where the bicycle utilization ratio is very high. Community P&C Nucleus□ presented in orange , has a high density of high-class residential clusters, where the bicycle utilization ratio is much lower, while the walking trips ratio is relative high.

The ratio of shopping trips and reaction trips in 2004 among total cycling trips was 16% and 4% respectively (31% and 21% of that of walking trips). Commercial P&C Nucleus can be identified through surveying visitors' numbers and bicycles parking around shopping malls and business blocks. Since they are scattered sparsely, no more research was carried out. (it is very important for Pedestrian Planning)

Our research group has sorted out the first three typical P&C Nucleus, totally 371 units. Utilizing the electronic maps we carried out the corridor analysis to ensure the selected branches can go through the nucleus, and therefore, arranged the names and required engineering measures of south-north and east-west corridors.

4 Criteria and Planning Program

4.1 Criteria of Plan

(1) Connecting P&C Islands: planning corridors for those necessary to cross main roads and link P&C Islands together. (similar as expressways).

(2) Passing P&C Nucleus: corridors layout should cross through or approach more P&C Nucleus so that more cyclists can take advantage of the corridor system as soon as starting trips.

(3) Cutting and Straightening: bicycle corridors should be laid out along straight roads by reconstructing (linking) curved (broken end) sections to make it straighter.

(4) Prioritizing then Attracting: Providing a preferential environment to attract parts of cycling trips on main roads to release the traffic pressure on main road.

4.2 Layout of Bicycle Corridor

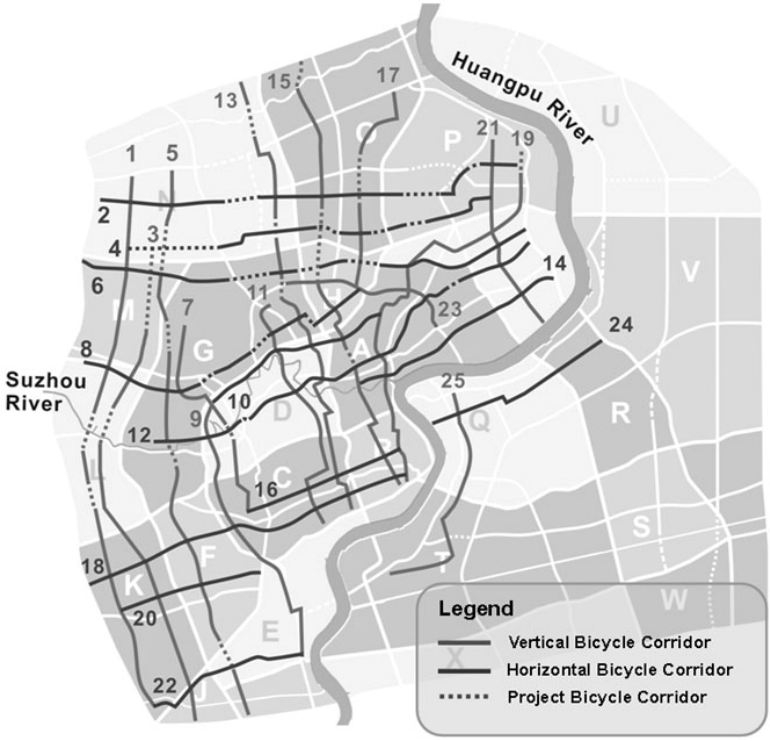


Figure 6. The Distribution of Bicycle Corridor in Shanghai central city

Figure 6 reveals the basic planning of 25 bicycle corridors (totally 314km) in Shanghai central city. Purple lines represent 13 vertical bicycle corridors, totaling 183.6km; blue lines represent 12 horizontal bicycle corridors, totaling 130.3km. Vertical corridors are coded in odd numbers from west to east, and horizontal ones are coded in even numbers, so that citizens can remember the directions and locations. Dotted lines represent three typical important engineering projects, including new bridges, underpass rail corridors, and stub-end roads linkage.

Taking Corridor 7# as an example, this corridor (see Figure 7) is composed of 7 sections of branches, in total 7 kilometers. It crosses through Inner Ring Elevated Road, Yanan Elevated Road, and links Islands D, G and C; and it also crosses through some P&C Nucleus, including G22, G23, G26, G27, G28, D03 (East China University

of Politics and Law), D04, D06, D07, D08, D09, D10,D11, and so on. This corridor would attract some of the bicycle flows on Caoyang Road, Jiangsu Road, and Huashan Road. In the future, it would become a corridor network with the linkage of corridor 05#, 08#, 09#, and 12#. Suzhou River New Bridge is an important project in the planning, which will bring more convenience to cycling trips in Caoyang Community.

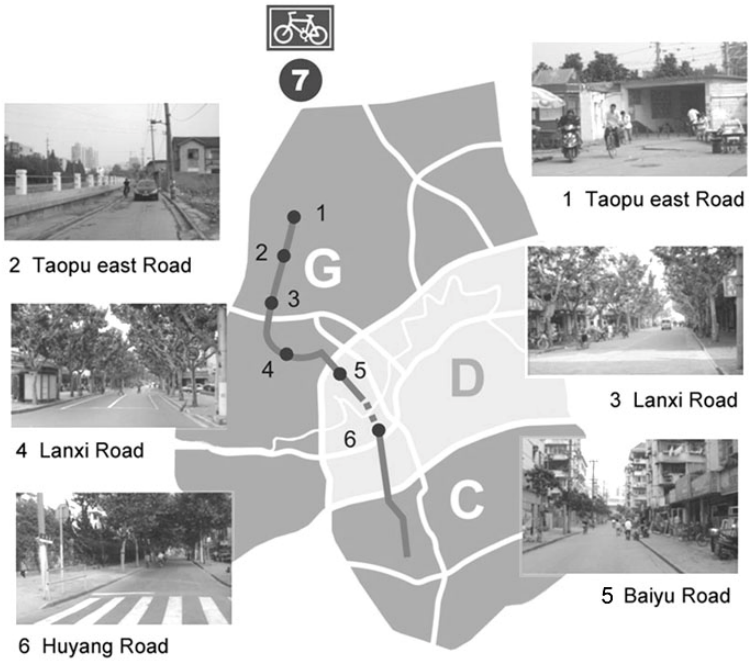


Figure 7. Alignment and typical sections for Bicycle Corridor N0. 7

4.3 Corridor Section Design

One bicycle lane needs 1m width at least. The capacity of a bicycle lane is suggested to adopt 2100 vehicles per h.m in China. Considering the influence of mopeds, the capacity of a bicycle corridor is advised to use 1000 vehicles per h.m.

During the peak time, bicycle flow is about 2000 to 4000 vehicles per hour in a single direction on most main roads in Shanghai, thus the minimum width of bicycle corridors should be no less than 2.5m. To meet the demand of three low-speed bicycle fleets or two high-speed fleets forward through, the recommendatory width of Bicycle corridor should be 4 to 6m.

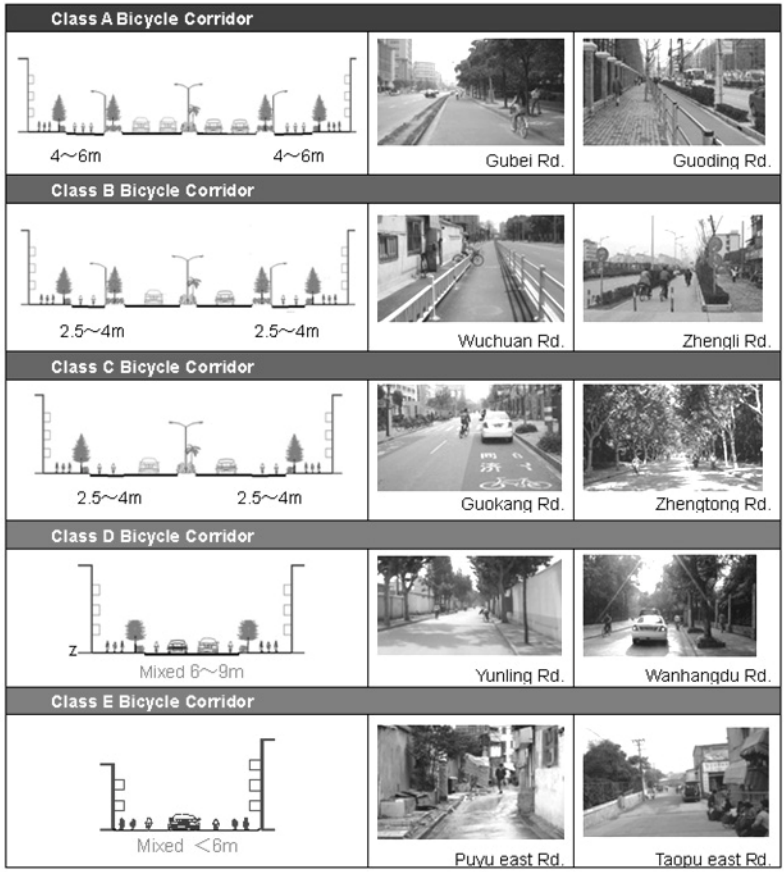


Figure 8. The Distribution and Typical cross-sections of Bicycle Corridor

As indicated in Figure 8, based on the field observations of 306 typical cross-sections, we designed 5 class standards of bicycling corridors in central city. Type A and B both have a physical median with motor vehicles, and the bicycle lane width in each direction is 4 to 6m, 2.5 to 4m respectively. Type C is separated by line markings, with 2.5 to 4m bicycle lane width in each direction. Type D is a mixture of bicycles and motor vehicles but contains a specific sidewalk, with a 6 to 9m road width. Type E is mixture used by motor vehicles, bicycles, and pedestrians, and the road width is generally less than 6m.

When designing and building these bicycle corridors in Shanghai central city, bicycle traffic volume on them should be considered as should selecting a high-level cross-section standard and fence separations. On those low-level corridors, "Bicycle priority" policy should be carried out during peak time.

4.4 Dynamic management of Corridors

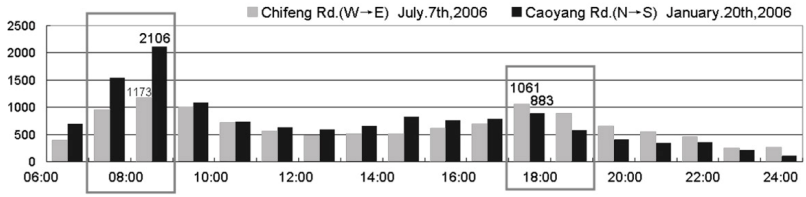


Figure 9. Daily bicycle volume variation in Chifeng Rd. and Caoyang Rd.

Figure 9 reveals that lots of cycling trips happen around 7:00 to 9:00 am or 17:00 to 19:00 pm. In Shanghai central city, the tidal feature of cycling trips is so obvious that dynamic management is needed on different periods of daytime in order to utilize corridors efficiently.

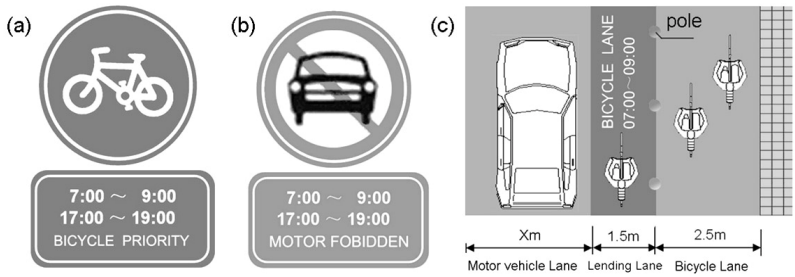


Figure 10. Dynamical managements on Bicycle Corridor (Level C to E)

In Figure 10, (a) indicates that it is necessary for C to E types cross-sections to carry out "Bicycle priority" policy during the peak hours, (b) reveals some sections with heave traffic conflicts should adopt "Motor forbidden" management policy during the peak hours, while (c) represents that installing poles with equal intervals can avoid the intrusion of motor vehicles into bicycle lanes. During the peak hours, bicycles (especially mopeds) on those over-saturated corridors can "borrow" one lane from motor vehicles.

5 Conclusion

In Shanghai central city, more and more bicycle lanes are reconstructed into motor vehicle lanes to release motor traffic pressure on main roads. Based on both efficiency and equity, it is essential to create an accessible, safe and continuous trip environment for cyclists utilizing newly constructed bicycle corridors. Considering that more and more urban roads are being used by motor vehicle, it may be the last opportunity for planners to do something for cyclists.

In 1990s, with the Nanpu Bridge and Yangpu Bridge opened to traffic, the history of motor vehicles isolated on two sides of Huangpu River declared the end. Similarly, when bicycle corridors are built, cycling trips will have more convenience between P&C islands and become more comfortable in P&C islands. What's more, it is more likely for motor vehicles and non-motor vehicles to coexist harmoniously.

Acknowledgment

The authors are extremely grateful to Wenping Yuan, Sumei Sun and Qing Li who assisted in the administration of the survey and to Jian Li who completed much valuable work on finding P&C Nucleus, and without whom this research would not have been initiated. Other additional support was provided, in part, by Shanghai Municipal Public Security Bureau.

References

The people's government of Shanghai Municipality.(2002). "Shanghai Metropolitan Transport White Paper in 2002"

Bureau of the leading group for the second comprehensive transportation survey of Shanghai.(2004). "The third comprehensive transportation survey of Shanghai "

Bureau of the leading group for the second comprehensive transportation survey of Shanghai.(1998). "The second comprehensive transportation survey of Shanghai"

Transportation Planning: Travel Demand Modeling Forecast Model Conversion from TransCAD to Cube Voyager

Ka-Fai Wong, Dowling Associates, Inc.

Dowling Associates, Inc., 180 Grand Ave, Suite 250, Oakland, CA 94612; PH (510) 839-0871 ext 129; FAX (510) 283-0871; email: kwong@dowlinginc.com

Abstract：
The key to develop an accurate, reusable, and representative transportation demand model requires extensive planning prior to beginning any coding. After all, the selection of which transportation software application to deploy the transportation demand model will greatly affect the outcome of the run. For example, when applying/using the same modeling approach on transportation demand forecasting, it can be expected that two different model applications provide comparable results. However, one model application could take one tenth of the run time of the other. Run time duration is one of the many factors transportation professionals should consider. The difference is not only a matter of computer time – as the run requires professional monitoring that, in turn, transfers directly to cost. When evaluating applications, one should consider factors such as ease of maintenance and output information along with runtime duration. It may be even beneficial to convert the model to a new platform. Despite the initial investments needed to convert or redevelop transportation model, the benefits do add up expeditiously. This paper will study the reasons behind the decision of converting the Transportation Demand Forecast Model from TransCAD to the Cube software package. Lastly, the pros and cons of two different applications will be compared and analyzed.

Key words: *Demand Forecast Model, TransCAD, Cube*

1 Introduction

General Assumptions, Facts, and Theories

Traffic Demand Modeling helps government agencies setting the base case for existing traffic condition. In addition, the model predicts future travel demand given forecasts of future land uses, highway improvements, and transit improvements. The result information forecasted will be predominantly used to monitor traffic circulation, congestion, and safety.

Regardless of what software applications is used to build the transportation demand model, most traffic planners/modelers will agree on the conventional standard 4-step traffic demand modeling process: Trip Generation, Trip Distribution, Mode Choice, and Assignment. The trip generation step converts land uses into estimates of daily trip

generation; the step identifies all the origins. The trip distribution step determines where the trips will go; the step determines the destinations. The mode choice step determines the origin-destination trip whether will be made by private vehicle or by public transit. Then the assignment step assigns peak hour trips to specific streets and freeways.

2 Background of the Project

The Contra Costa Transportation Authority (CCTA), in the San Francisco Bay Area, maintains a traffic demand model which was developed in the TRANSCAD application for Contra Costa County, which is one of the 9 counties that make up the San Francisco Bay Area. The traffic model includes all of the Bay Area land development and roadway network data. It also includes county-to-county travel data for forecasting San Joaquin County trips through the Tri-Valley and to other Bay Area destinations.

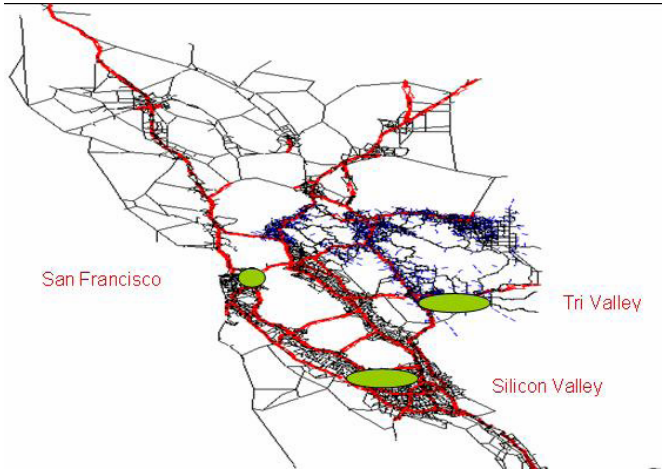


Figure 1. San Francisco Bay Area Network

This project covers the vicinity of three cities: Dublin, Pleasanton, and Livermore. These are so-called the Tri-Valley. The network infrastructure was built based on three corridors of freeway or highway: I-580, I-680, and Route 84. As the three cities have been undergoing massive growth in land use and employment, the primary transportation corridors located in the study area are over capacity during the peak hours. The high volume of traffic diverts through the study area, which is situated between the Bay Area's affordable housing sprawling east and the job epicenter in San Francisco and Silicon Valley. It is because of this sprawling housing phenomenon, Alameda County Congestion Management Agency (ACCMA) initiated the study. The goal was to create a long-range plan for sequencing improvements to achieve practical traffic relief in the I-580, I-680 and Route 84 corridors in a cost effective manner consistent with the transportation needs of the area.

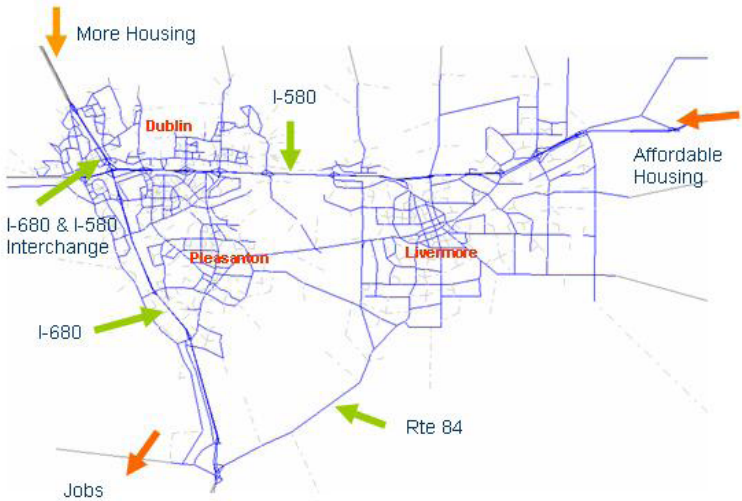


Figure 2. Tri-Valley – Transportation Infrastructure & Travel

The Tri-Valley adopted the CCTA regional model and have been utilized it to run traffic modeling process. The complete process combines a traditional demand forecasting tool and an advanced traffic micro-simulation tool. The step of traffic demand forecasting was implemented in the TransCAD software application to generate and distribute traffic throughout the system network, based on study scenarios. It provides a macroscopic, system-wide analysis within the traffic network system. For detailed traffic operations modeling analysis, Tri-Valley deploys detailed traffic micro-simulation in Synchro. The combination of the two provides macro and micro analysis within the vicinity. The outputs present assessment of freeway, local street congestion, and the inter-relationships between the two.

The Tri-Valley model project could have been first extrapolated from the existing CCTA without any customization. Alternatively, the model could either be revised and customized based on the existing CCTA. Either case, the traffic generation details are essentially broken down into the number of residents and jobs within each traffic analysis zone. The roadway network possesses roads and connectors that represent the reality of how traffic enters from one zone to another. With the supported GIS (Geographic Information System) features, the roadway within the system contains many decision factors like travel distance, number of travel lanes, lane capacity, maximum travel speed, average travel speed, and facility types. Due to the size of the CCTA complete regional network and multiple of numerous attributes, the time to finish all the matrix computations and assignments alone will take many hours to run. This is one of the primary reasons why the proposition of a customized traffic demand model is necessary to ensure the model being cost effective. The remaining question is

whether to convert the model from TransCAD to any comprehensive traffic demand model application.

The forecast model utilizes a combination of the City's existing and approved future land uses and the existing and future roadway network to project traffic volumes based on calculated travel times. The model routes traffic as necessary to produce traffic volume based on the shortest travel time available under each scenario. The fundamental of the traffic model remains unchanged regardless which application the new Tri-Valley model will be developed. In addition, the model will have to be calibrated and validated to match the model traffic volume and distribution projections closely with the actual traffic counts collected at multiple locations throughout the city. The calibration process is assumed the model represents as close as possible of City's regional traffic origins and destinations. Hence, the fundamental of the model will remain valid despite of when traffic conditions and travel patterns changed. The model will reflect great realism in the traffic projections with future land development and additional transportation infrastructure. The time and effort to put into later procedure will be substantial.

The preliminary research strongly indicated Cube Voyager, the proprietary comprehensive traffic modeling application developed by Citilabs, is an appealing candidate that up for the Tri-Valley project. Recently, the research center of Florida Department of Transportation evaluated several transportation model applications. The top-rated software applications were TransCAD and Cube Voyager. The research committee recommended that both TransCAD and Cube Voyager are equally effective in meeting the planning and modeling needs of the State of Florida. Its recommendation caused further intrigue in the selection process of selecting the application for Tri-Valley traffic demand model and necessitated further research between the TransCAD and Cube Voyager.

3 TransCAD

TransCAD is a flexible and capable travel demand modeling software. It supports various travel demand modeling approaches from sketch planning methods, UTPS-style four-step demand models, advanced disaggregate modeling techniques and simultaneous models for multiple choices. It also comes with a set of traffic assignment models.

TransCAD features its integration of GIS and transportation modeling capabilities on one platform. Its integrated GIS engine is equipped with special extensions designed for transportation. As the result, the application can handle mapping, visualization, and spatial analysis. It also better conveys the ideas in transportation flows and network characteristics.

3.1 Strengths and Weaknesses

TransCAD is also known for its matrix calculation. It provides many statistics for any given loaded matrix without any scripting effort. This handy feature will allow the planner to validate the matrix with a sense of whether the numbers are in the expected

range. The simple statistics of a matrix reveals so much information that it serves as an invaluable time-saving tool during the debugging process.

As data involved in planning increase, the matrixes grow in multi-dimensional while the data become more valuable. The trend of storing the data migrates towards to any conventional ODBC platforms, such as Oracle, SQL Server, or MS Access. TransCAD built-in database interface protocol can connect directly with any ODBC compliant data source. It can also be integrated with web server application and most application development languages.

TransCAD also provides a GIS engine with special extension for transportation. It is a function that most transportation modeling applications do not feature. This is particularly useful for conducting spatial analysis while maintaining geographic data sets, such as geocoding, polygon overlay, and buffering.

Because of the embedded GIS engine, TransCAD is designed to provide more spatial and visual display to linear-referenced data sets without conversion, such as accident locations, pavement condition, etc. Furthermore, TransCAD provides additional tools that further display better visualization on data that is not easily illustrated using conventional GIS applications without a few customizations or twists. Such data vary from intersection diagrams that illustrate turning movements and flow volumes, to turn restrictions and delay, to strip charts that represent facility characteristics. Furthermore, it can be applied to analyses and illustrations of the correlation between demographic and geographic characteristics. That leads to its strength in territory management and site location modeling.

However, the strengths of TransCAD in certain areas become its weaknesses as well. The learning curve to expertize in GISDK and GIS extension can be steep. The idea of integrating the transportation modeling with scripting and GIS extension together requires careful planning. In order to fully implement the GIS and/or GISDK into the foundation of transportation demand modeling, it certainly takes more in-depth understanding of GIS and object-oriented scripting knowledge than expected. It can be a significantly complicated process to utilize the sophisticated yet intricate features to work well together. The add-on features may be advanced for unaccustomed users. In practice, a GIS analyst and Transportation Planner are two separate identities and professions.

3.2 Deliverables/Outputs

TransCAD can export data in ASCII, shape files, dxf, and dbf formats. There are four features related to information generation and deliverable results which have been repeatedly utilized: census database, intersection turning movement graph, bandwidth traffic volume, and equilibrium traffic assignment matrix.

Census data are valuable in urban transportation analysis. TransCAD can import data from the Census Transportation Planning Package (CTPP) based on any customized

segregation by parcel or state. The database can be incorporated with GIS supported network. It is a valuable time-saving tool.

Intersection turning movement graph presents all twelve turning movement volumes with scaled bandwidth. The feature allows user to generate graphic information with minimum effort. It is definitely an indispensable function.

Bandwidth traffic volume allows user to look at the traffic assigned network in a macroscopic perspective. It instantaneously reveals a lot of traffic conditions within the model. It is a great feature to provide a clear visual for user with a few clicks away.

Equilibrium traffic assignment matrix stores all the origin-destination traffic volumes. Often it holds the key to the model validation. It is another feature that any diligent and prudent user should keep referring to.

4 Cube

Like TransCAD, Cube provides a comprehensive package of software products for transportation system planning and analysis. With its comprehensive module extensions, Cube package encompasses a wide range specialty in:

- Urban regional forecasting
- Long distance travel forecasting
- Multimodal microsimulation
- Matrix estimation
- Air quality analysis
- Freight forecasting

The core of the package, Cube Base, is the user interface for the entire Cube system. There are 5 optional extensions built around Cube. Each of the extensions specializes in a unique perspective related to transportation planning analysis.

Cube Voyager - incorporates methods ranging from four-step to discrete choice to activity-based models for personal travel forecasting.

Cube Me - developed for estimating and updating statistically optimized base year travel matrices.

Cube Cargo - based on commodity-based approach, offers freight forecasting. It is capable to test from pricing strategies to freight-specific facilities.

Cube Polar - estimates air quality emissions and energy consumption depending on network assignments, speed profiles, vehicle classifications, fuel characteristics, or any attribute that would affect emission.

Cube Dynasim - provides sophisticated, flexible methodologies to tie data and results from other extensions to conduct detailed traffic micosimulations including automobiles, trucks, buses, pedestrians and rail vehicles in 3D graphical displays.

4.1 Strengths and Weaknesses

The sophisticated integration of ArcGIS from ESRI is an important strength of Cube. The tight incorporation with ESRI, the world leading vendor in GIS technology has greatly enhanced Cube's broad range of capabilities to provide analyses in a wide

range of planning aspects from new public transit development, testing alternatives, road pricing strategies, to freight movement analysis. Unlike TransCAD, Cube possesses a feature called Catalog, a nifty application development interface to provide the flow chart visual. It is an incredible feature any software application should adapt. It can present any model from the beginning to the end, and it is flexible to edit and control the flow of input, output, and feedback procedures with a few clicks and drags with a mouse control.

Figure 3. Cube Catalog shows 4-step Model

Cube Voyager comes with a set of model templates available in its library. The comprehensive library allows users to build any conventional travel forecasting models easily. Models built with Cube Voyager can be traditional four-step models, modified four-step with feedback, activity-based demand, and combined equilibrium models. Another key in Cube Voyager is the ability to represent signaled and non-signaled controlled intersections, roundabouts, and ramps as part of the pathfinding and assignment process. In addition, the library possesses functions for modeling and analysis of passenger transport systems: roadways, public transit, pedestrians and bicycles. It is particularly useful for any country, city, or community where bicycling is a dominant modal.

Aside from its comprehensive library, Cube Voyager supports its proprietary scripting language, which allows users to easily design and represent any model methodology, with no limitation of customizations. Models can be of any form, plus any feedback can be looped back to the model if decided. Feedback feature with user-controlled criteria is one of few crucial features in modeling. Like TransCAD, with the scripting function, Cube Voyager can accommodate any complex demand process.

4.2 Deliverables/Outputs

There are four features in Cube substantially enhance the user-ability and the quality of the model work: select link volume trace, multi-bandwidth posting, equilibrium traffic assignment matrix, and statistics and comparisons.

Select link tracing can be achieved by a less than 10-line script, with a built-in function of Cube Voyager. How would a new project affect the traffic in its vicinity? It

can be answered with a dedicated traffic analysis zone for the specific project before running the model. With the assigned network, a select link script can trace the traffic volume produced and attracted by the project. This excellent feature works great for planning and zoning tasks.

Multi-bandwidth allows posting up to 8 link attributes simultaneously. In addition to allowing the user to view the traffic assigned network in a macroscopic perspective, it demonstrates the correlation among the attributes. It is a useful tool to tie the relationship between assigned traffic volumes with link capacity, free flow speed, or physical constrains of a roadway.

Equilibrium traffic assignment works similarly like TransCAD does. It keeps all the origin-destination traffic volumes. The model work will not be completed without censoring the matrix.

Every model run comes with a simple statistics and comparisons at the end of its log file. It will list all the variables involved in the model with number of observations, total, average, minimum, maximum, and RMS (root mean square). The log file is a good source to start checking the model run and should never be overlooked.

5 Conversion

5.1 Reasoning

The Tri-Valley traffic model had to be completed and delivered in 4 months. The schedule will not be attainable if the regional model with sparse network and zone details within the study area, land use data from City General Plans, and comprehensive sets of 2005 AM & PM peak hour traffic counts were not available immediately. Although the network and zone details were in TransCAD format, the information were exported and converted with reasonable effort, without causing any impediment.

The final decision to convert the CCTA regional model to Cube Voyager for the ACCMA Tri-Valley project was based on the following reasons:

- Matrix Estimation Module
- Select Link Feature
- Faster Run Times
- Deliverables

Cube ME has been designed specifically for estimating and updating base year automobile, truck and public transit trip tables. It allows specific projects to exploit extended variety of data that affects and contributes to matrix development and update. The matrix estimation step greatly determines the run time as the number of origin-destination pairs increases.

The select link feature provides great flexibility in computing select link numbers including the ability to trace volumes based on any origin or a group of origins. It can isolate and identify the individual city in Tri-Valley.

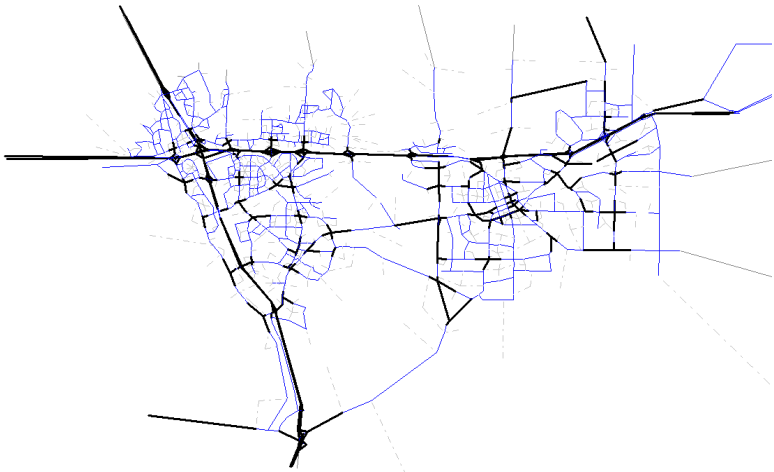


Figure 4. Traffic counts available, highlighted in black

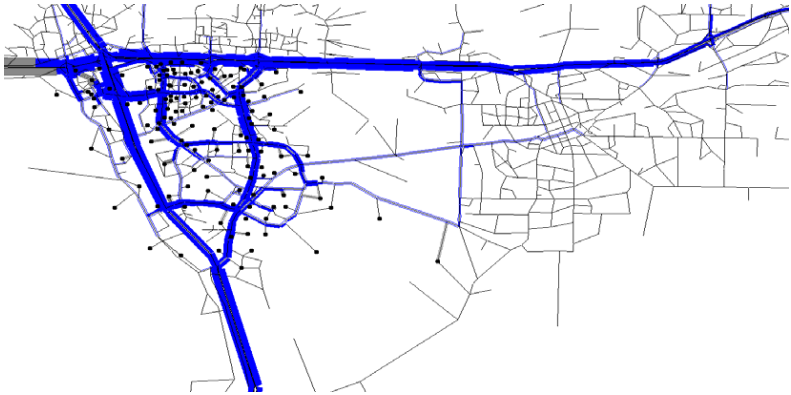


Figure 5. Select link by city; Trips with origins and destinations in Pleasanton

Based on the processing speed benchmark test conducted and published independently by the Florida Department of Transportation, Cube Voyager performed freight model assignment to 4,059 zones network under 5 minute and 32 seconds for 2 iterations, compared to 6 minutes and 36 seconds for single iteration in TransCAD. Evidently, Cube Voyager greatly shortens the model run time.

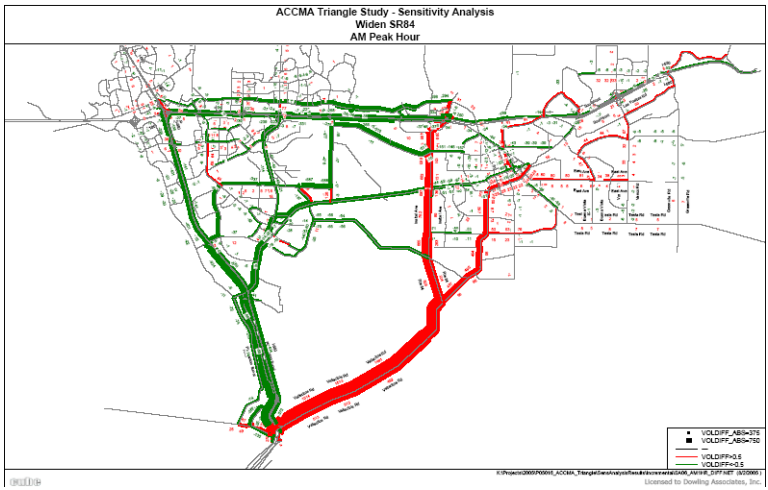


Figure 6. AM Peak Hour – Widen SR84

The necessities of producing wide range of deliverables from maps to plots hinge on the ease of use. Apparently, Cube Voyager appeared to be well balanced in terms of its ease of use and options in deliverables.

5.2 Steps

The following are the tentative steps for the conversion:

1. Updated land use and network in the CCTA regional model in TransCAD
   a. Added network and zonal detail attributes in the Tri-Valley study area
   b. Updated the land use to match the general plan of the three cities
2. Ran the regional model to obtain the 4-hour AM & PM peak assignments
3. Converted the network and trip table to Cube Voyager format
4. Used the Sub-Area extraction process (4-hour trip table) for the Tri-Valley area in TransCAD
5. Used matrix estimation to obtain a 1-hour trip table (peak hour spreading) from the 4-hour trip table
   - Matrix Estimation module was used to compute AM & PM Peak HR trip tables from 4 HR AM & PM trip tables
   - Capacity assigned as counts to the gateway links.
   - High confidence level set for gateway locations
   - Inputs matrix estimation step
     - Intercept file from the assignment step
     - All count locations
6. Assigned the 1-hr trip table
7. Steps 5 & 6 were repeated (Reconciling counts, correcting the network and land use) to obtain a calibrated 1-hour assignment
   - Calibration & validation
   - 1-hour assignments were used to generate validation statistics by city

- Matrix estimation step provides valuable inputs to correct network and land use errors
- It helped a lot in the count reconciliation process
- Matrix estimation process was repeated until a set of acceptable assignments was developed

8. Ran the regional model for Year 2025 (4-hour AM & PM assignment)
9. Repeated Step 3 for Year 2025
10. (Raw 2025 + (Raw 2003 – Calibrated 2003) and anointed FUTURE BASE
11. Ran various alternative packages and compared them to the Future Base
12. Performance measures were calculated for each and compared with respect to future base

5.3 Quality Assurance

For calibration and validation, 1-hour assignments were used to generate validation statistics by the cities. The matrix estimation step provides valuable inputs to correct network and land use errors. Any link not being utilized in matrix estimation is a strong indication that the network is coded incorrectly. Also, any centroid connector that carries no traffic may be a sign of a traffic analysis zone with incorrect land use. It helped a lot in the count reconciliation process as the matrix estimation process was repeated until a set of acceptable assignments was developed.

Among the alternatives, performance measures were calculated and tabulated. The measurements included:

- Average Speed
- VMT (Vehicle Miles Traveled)
- VHT (Vehicle Hours Traveled)
- VHD (Vehicle Hours of Delay)
- PHT (Person Hours Traveled)
- THT (Truck Hours Traveled)

Furthermore, the measurements were aggregated by city and trip classes in order to segregate the nature of the trips.

- Intra city trips
- City-to-Rest of Tri-Valley trips
- Rest of Tri-Valley-to-City trips
- City-to-External trips
- External-to-City trips
- External-to-External trips (a.k.a. Cut-through trips)

6 Conclusion

In many public opinions, TransCAD and Cube package have been deemed equally effective form a technical viewpoint, despite a handful of their niche expertise. It often comes to the other aspects related to costs, technical support, product updates, and services provided by the software vendor.

Often time, it is ultimately the client's decision what application should be implemented, with or without professional advice. Unless it is a new model, it is rare the model will be converted, given time and budget constraints.

6.1 Qualitative & Quantitative Analysis

Beside qualitative analysis focus on features, functionalities, strengths or weaknesses, one can look at it differently from a quantitative analysis. Assess comprehensive future tasks and compare the time saved from using one of the applications with the other one. Multiply the time savings with the estimated hourly rate of a professional planner will be the rough amount of money saved by deploying in one application over the other. Or better, one could multiply the number of hair pulled out due to debugging by one's personal value per strand of hair.

When breaking down the costs, do not solely focus on the initial cost of the software application and initial setup cost. An efficient and user friendly application will cost less to maintain and operate in the long run. Simply determine the net present value by estimating all future costs associated with maintaining and running the model in each application. Each cost is discounted or adjusted by incorporating time value of money. A net present value analysis can be compared with the two alternatives objectively, provided with objective future cost estimations. In financial theory, if there is a choice between two mutually exclusive alternatives, the one yielding the higher net present value should be selected.

6.2 Lesson Learned

Conversion is a big project. It requires a good detail planning before the work should begin. Identify the goal. State the needs and wants. Understand the strengths and weaknesses of the applications available.

Make sure to schedule ample amount of time for the quality assurance step. It is a step that requires great detail of attention, which should be well budgeted. The repeated matrix estimation step can be extremely time consuming. The time it takes will increase exponentially as number of zones increases.

Great planning and well preparation, as complete set of data will save you multiple folds of time later, consider the amount of time invested at the beginning be the best investment in modeling. Modeling requires a great set of technical skills and understanding the fundamental theories. However, with all the sophisticated modeling applications available in the market, traffic demand modeling skill evolves to be a profession which demands both artistic and technical skills. Planning is an art, and so is the model fine tuning.

Despite how well the preparation be, it is inevitable some unforeseeable situations would arise. All in all, the combination of setting up detail planning, visual flow chart, and qualitative and quantities analysis creates a synergy for making crucial decision in any modeling project. Build in some safety margins in terms of time and budget to avoid being caught off guard.

Disclaimer

I do not affiliate with either software application vendors in any way. I do not engage in any activates associated with either vendors.

Reference

Florida Department of Transportation, Research Center,
http://www.dot.state.fl.us/Planning/mtf/subcom.htm

Acknowledgment

Special thanks to Madhav Pai and Rowena Ona, my associates, and Kym Sterner, my manager, for all information and advises.

DESIGN AND DEVELOPMENT OF PLANNING A DECISION-ORIENTED TRANSPORT INFORMATION PLATFORM

LIN, Hangfei. YANG, Chao. CHEN, Xiaohong.

School of Transportation Engineering of Tongji University, Shanghai, 200092, China, Tel.: 86-21-65982127, 86-21-65982127, linhangfei@126.com

Abstract: Based on the infrastructure condition and transport technology capacity of China, the primary objective of Intelligent Transport System (ITS) development of Chinese cities is to provide technical support and decision support of the transport planning process. The construction of planning a decision-oriented transport information platform is to realize the objective above mentioned through information collection and abstraction. This paper analyzes the requirements of the system, proposes the software architecture, and discusses critical technique problems in the system development. The proposed design method was successfully applied in the Shenzhen Urban Transport Simulation System**.**

Keywords: *Comprehensive Transport Information Platform, Decision Support –Oriented, Transport Plan, System Design*

1 Background

Most Chinese cities are at the stage of rapid construction of transport systems. Faced with the urbanization and motorization of the cities, local governments are trying to solve and prevent transport problems through effective planning of the transport system and the construction and management of transport infrastructure. These characteristics determine that the construction and development of Intelligent Transport Systems (ITS) in Chinese cities will have different promotion modes and deployment roads from those in cities of developed countries. It is broadly accepted that an ITS improves the operational efficiency of the existing infrastructure, improves the traffic environment, and increases traffic safety using advanced technologies (1). However, based on the infrastructure condition and transport technology capacity of China, the primary objective of ITS development of Chinese cities is to provide technical support and decision support of the transport planning process. To achieve this, it is important to develop a planning decision-oriented transport information platform (PDOTIF) by solving the technical problems of information collection and abstraction. A transport information platform is considered the core of ITS construction. However, most transport information platforms discussed are on traffic

management and traveler information service purposes. They are rarely considered for the planning decision and construction scheme purposes.

This paper analyzes the requirement needs of the PDOTIF and proposes the software architecture of the system. After that, the critical technique problems on system function and software implementation are discussed. Finally, a show case system, developed using the proposed method, is demonstrated.

2 User Requirement Analysis of PDOTIF

Based on analysis, the users of PDOTIF include decision makers, transport technicians, travelers and system operation and maintenance persons. The duty of system operation and maintenance persons is on the operation management and maintenance of the system, so the requirement of those will not be discussed in detail here.

The decision makers refer to authorities in charge of transport planning, construction and management. Their requirement is the information that supports decision-making. The transport technicians refer to those carrying out the transport plan and design using transport simulation models. In the transport plan, the technicians need reliable data, calibrated model parameters, and transport simulation tools that are easy to use. The travelers refer to the direct users of the transport system, including public transit and private transport. The specific requirements of these users are analyzed as follows:(2)

2.1 Decision makers

The PDOTIF provides decision support tools for decision makers. The tools support the fusion and analysis of transport indices and parameters from different sources and different time periods. It establishes a comparable, intuititive, and consistent benchmark for the analysis of the increasing and developing trends of the transport condition. It will provide figures, charts, reports and suggestions. The decision-support function includes:

2.1.1 Automatic report generation of traffic operation conditions

Automation and intelligence of the data collection, analysis and performance evaluation can be achieved. Supporting data for monthly and annual reports of traffic operation conditions can be provided. The reliability of the data can be increased, and the timeliness of the report can be improved. A uniform data view chart can be established to ensure the consistency of the business analysis and decision making among different organizations.

2.1.2 Traffic condition diagnosis and hot-zone recognition

Based on surveillance data, real-time speed and flow data are analyzed. The congestion condition of each district, main traffic corridors and important areas are summarized. Recognize the hot-zone of congestion based on the frequency of

congestion occurrence. Provide decision support for traffic management and control, traffic improvement schemes, etc. Increase the service capability of the local government.

2.1.3 Demand increasing trend analysis

According to the historical data, the changing trends of performance indices, such as travel demand and supply and level of service, can be analyzed. The travel demand indices include vehicle ownership, vehicle kilometer readings, and traffic volume of screen lines. The travel supply indices include equivalent road mileage and supply capacity of transit. The level of service refers to average vehicle speed, transit penetration, and the accessibility of the CBD area. With these performance indices, the local government can evaluate the effect of critical infrastructure project and traffic management policy.

2.1.4 Benefit demonstration of transport planning projects

The above mentioned performance indices can be used to demonstrate the benefit of the transport planning projects. The relationships among the population, vehicle ownership, road network size, rail network size and performance can be revealed.

2.2 Transport Technicians

For transport technicians, the system provides process support and technical support in the transport planning. More specifically, the system provides a user-friendly interface of the macro-traffic simulation software. Through the interface, the transport technician can carry out a sensitivity analysis of different design scenarios with a simple modification of a base scenario and model parameters. The critical indices of the model output are custom-tailored. A user can also acquire plenty of data, such as real-time data, simulation results and historical data for comparison and analysis.

2.2.1 Consistency of the basic data

Uniform management of the basic data for the transport planning models can be provided, such as network data, basic model information, model parameters, demand analysis results, traffic evaluation indices, and a traffic model supporting database. The transport technician can apply for the access right of those data according to the design project. Besides the base year data, forecasted future year data, such as data of 2010, 2015, etc., can also be obtained. The access right is managed by the system operation and maintenance person.

2.2.2 Usability of the simulation tools

Road network and origin-destination (OD) matrices are used in macro traffic simulation software. The system improves the usability of the simulation tools by establishing an application environment for commercial macro-simulation software. The system develops pre-process and post-process modules so that the input and output of simulation tools are more user-friendly. With the base scenario available, the

user only needs to input the modification of the base scenario through the pre-process model and then get the output. For instance, the user can analyze the change of road network conditions with the increase of vehicle ownership.

The application environment lets transport technicians use the traffic simulation tools more easily, which will promote quantitative analysis in the transport planning and design projects. This environment can also control the reliability of the results of the quantitative analysis with the standardized outputs. More importantly, for the security of critical data such as road networks and OD, the user cannot download these data.

The post-process module of the application environment provides standardized simulation results, such as figures and tables, which can be downloaded and added into reports. This ensures that the presentation of different transport plan projects have the same formulation and content.

2.2.3 Improvement of traffic models

In the most widely used four-step demand analysis model, the parameters of models need to be calibrated with OD survey data. Certain parameters, such as parameters in link cost function and speed-flow curves, are sensitive to the results. The system can calibrate these parameters with real-time data on different road types. Also the real-time data can be used to ensure the quality of the traffic model.

2.3 Travelers

Travelers can view the real-time traffic congestion map on the Internet, the Plasma Display Panel, and mobile terminals. Travelers can also acquire real-time traffic conditions and travel routes. Providing travelers with real-time traffic information can help the travelers use the traffic system more effectively and can lead to a change of travel mode.

3 Software System Design

To demonstrate the software design concept, a real system, Shenzhen Urban Transport Simulation System (SUTSS) that was developed in 2006, is used. SUTSS makes use of advanced information system integration technology, intelligent transport system technology and transport simulation (macro-scope, meso-scope, micro-scope) technology. It directs the key target of providing decision-making support to the transport planning, construction plan making, and transport system renovation. It adapts the long term development requests of ITS construction. This system integrates standardization and mature reliable technology, guarantees data compatibility, versatility and scalability, satisfies the resource sharing and data exchange requests, reserves Shenzhen and Hong Kong inter-areas transport shared information platform interface, and provides the communication service to related systems. SUTSS is composed of “one network, four platforms” as illustrated in Figure 1:

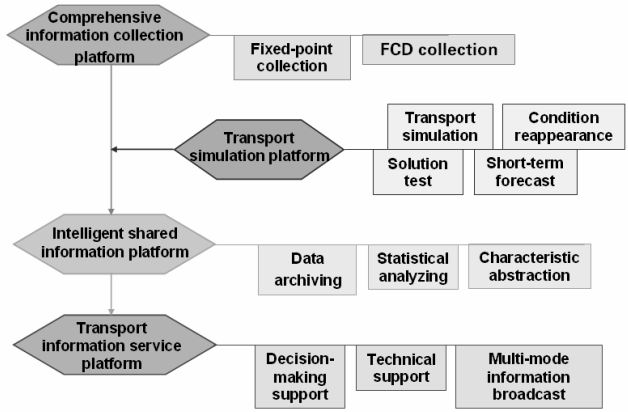


Figure 1. Logical structure of the PDOTIF (3)

Because the software architecture will not be changed in the development of the software, it is critical to grasp the objective of architecture design. Here the objective is set as

(1)Maximal re-usability: Includes the re-usage of components and design modes.

(2)Simplification: Using Middle-ware and N-tie techniques to simplify the design of the software structure.

(3)Flexibility and expansibility: Architecture should support the extension of future functions and second development.

With the increasing number of types and amounts of information systems, the demand for information exchange is also increasing dramatically. The client/server architecture is not adapted to this circumstance. The solution is to use a three or more layer platform structure. This kind of structure is developed based on the object-oriented technique and distributed computation environment.

The PDOTIF is designed with three layer architecture: data access layer, data process layer and data distribution and application layer. The construction of the PDOTIF is the most important task for the integration of traffic information systems. The data access layer achieves the information collection. The data process layer achieves information abstraction. The data distribution and application layer achieves the information and knowledge expression. It is the medium for users and the system interaction.

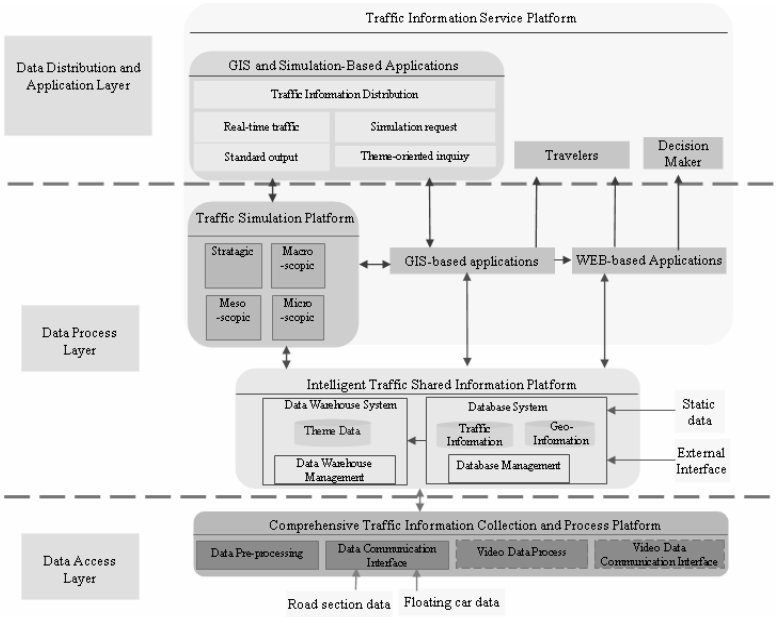


Figure 2. Software architecture of the PDOTIF (3)

4 Hardware System Design

System hardware consists of three parts: data collection devices, computer centre, and information broadcast devices.

4.1 Data collection devices

Data collection devices consist of floating car data (FCD) and road section data. FCD gathers the position and speed information of about 5000 taxis of several taxi companies. These data make about 4.2 million records every day. The real-time road information data covers more than 70% of the roads in the Shenzhen area. The system uses 67 sets of radar monitoring equipment (wireless data transmission, solar energy power supply), monitoring the traffic parameters (volume, occupancy, speed per lane per vehicle type every sixty seconds) at the pass, screen-lines and the main traffic corridors in road section data.

4.2 Computer center

A computer center consists of computer devices and network facilities. Computer devices include a data communication server, Web server, GIS server, traffic simulation server and database server that runs the Unix-version Oracle database management system. Network facilities include Core Switch, switches, firewall,

routers, and isolated network gate as well as the system operation, network application and supporting software.

4.3 Information broadcast devices

The system provides the information service needed by the public, technical personnel and decision makers through a website, plasma display panel (PDP), personal digital assistant (PDA) and web-based intranet applications.

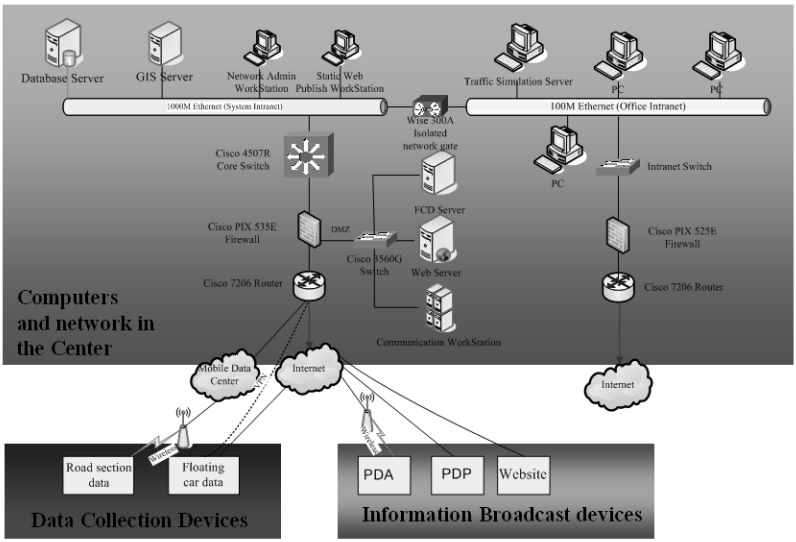


Figure 3. Physical architecture of the PDOTIF (4)

5 Critical Techniques

In the development of the PDOTIF, there are a few technical difficulties in the transportation area, in the computer science area and in the GIS and survey technique. Some difficulties are new to the experts in each area. To solve these difficulties, experts from different areas have to work together. In this paper, only parts of these areas are listed here(5).

5.1 Traffic condition estimation of the whole network

First cleaning of abnormal FCD data is conducted. The data is then processed though map matching, route searching, and link speed analysis. Finally the data is converted to link speed data for network traffic condition expression. For links with road section data, the data fusion is conducted to improve the quality of estimation. For links without real time FCD data, the historical data and data of adjacent links are used in the estimation.

5.2 Digital road network correction

Matching the FCD data with the GIS map is one of the critical techniques in FCD processing. The positioning data of FCD is geodetic coordinates in the WGS84 coordinate system while the road network digital map is based on a local Gauss plane coordinate system. As a result, the geodetic coordinate data need to be converted to WGS84 coordinate data.

5.3 Integration of dynamic data and the transport planning model

To realize the connection between transport planning and real-time factors, the parameters in transport planning models, such as link cost function, link capacity function and peak hour flow model, are calibrated with dynamic detection data.

5.4 Template demonstration of simulation results

With the requirements of data security, result sharing, architecture standardization and fast application, the simulation environment is developed based on a commercial traffic simulation package. The commercial traffic simulation package is embedded in the information platform. The simulation results can be extracted into the database of the information platform so that these results can be expressed as figures and tables with standard templates. The change of the transport system operation is expressed intuitively with critical parameters, such as speed, flow, and degree of saturation. A three-layer architecture (database server/simulation /server/client) of the software is established with ADO and DCOM techniques.

5.5 Traffic simulation business process re-engineering

Instead of searching data in paper files, traffic technicians analyze basic data by inquiring of the database of the information platform. The basic model can be tested by traffic technicians through the simulation environment instead of through the hands of the modeling personnel. At each transport planning project, a different simulation process can be constructed with different source data supporting. The efficiency of the model and basic data usage is improved.

5.6 Multi-modal traffic information broadcasting services

A multiple spatial indexing mechanism of ArcSDE is utilized effectively so that images can be visited quickly by client applications. A seamless space database is realized so that the efficiency of special data storing and management is improved. The network load is decreased with ArcSDE as a middleware to connect client applications and the database system of the servers.

6 Conclusions

The aim of constructing the PDOTIF is to support the transport planning process with technical and decision support by solving the technical difficulties in information collection and abstraction. To develop the PDOTIF, three critical problems must be solved: the definition of user requirement, the design of software architecture, and the breakthrough in technical difficulties. This paper studied these critical problems and

proposed the design method of the system. The proposed method was successfully applied in the Shenzhen Urban Traffic Simulation System project and was proved to be feasible and effective.

SUTSS is a transport comprehensive information service system which is adapting the gradual characteristics of the development of Chinese cities. Its design concept, system structure, function design and key technologies are applicable to medium-sized and above cities in China for their transport information service system construction and intelligent public transport system. In cities that have taxi or other vehicle location data, the basic functions of the system could be realized in about six months through comprehensive data utilization. The system can provide road conditions of a road network and travel information. It has great significance to ameliorate urban traffic congestion, reduce urban environment pollution and promote urban ITS construction.

REFERENCES

Wang, XJ, Qi, TY and Cai, H, (2004). "Principles and Applications of System Architecture for Intelligent Transport Systems." China Rail Publication House, Beijing.

Shenzhen Urban Transport Planning Center, Tongji University and Shanghai Baosight Software Corp. (2006). "Requirement Specification of Shenzhen Urban Transport Simulation System, Shenzhen.

Shenzhen Urban Transport Planning Center, Tongji University and Shanghai Baosight Software Corp. (2006). "Detailed Design Specification of Shenzhen Urban Transport Simulation System." Shenzhen, 2006.6.

Shenzhen Urban Transport Planning Center, Tongji University and Shanghai Baosight Software Corp. (2006). "Hardware Design of Shenzhen Urban Transport Simulation System." Shenzhen, 2006.7.

Shenzhen Urban Transport Planning Center, Tongji University and Shanghai Baosight Software Corp. (2006). "Technical Research Report of Shenzhen Urban Transport Simulation System." Shenzhen, 2006.11.

AN ORIGIN-BASED ALGORITHM FOR A COMBINED DISTRIBUTION, HIERARCHICAL MODE CHOICE, AND ASSIGNMENT NETWORK MODEL

Meng XU[1], Hillel Bar-Gera[2], Ziyou GAO[3]

[1,3]Institute of Systems Science, School of Traffic and Transportation, Beijing Jiaotong University, No. 3 of Shangyuan Residence, Haidian District, Beijing, 100044; PH (8610): 5168-7070; FAX(8610): 5168-7127; email: xumeng@jtys.bjtu.edu.cn, gaoziyou@jtys.bjtu.edu.cn
[2]Department of Industrial Engineering and Management, Ben-Gurion University, POB 653, Be'er Sheva, 84105; PH (9728):646-1398; FAX(9728): 647-2958; email: bargera@bgumail.bgu.ac.il

Abstract: In this study, an origin-based (OB) algorithm is further studied to solve a combined distribution, hierarchical mode choice, and assignment network model (CDHMCAN), where the trip distribution follows a gravity model, the mode choice follows aggregate hierarchical logit structures, and the traffic assignment is a user equilibrium model. The combined model will be first formulated as an optimization model with single user and multiple mode classes, and an equivalent mathematical program is presented. Then, the proposed OB algorithm for the program is given.
Keywords: *travel forecasting, combined models, origin-based algorithm*

1. Introduction

Transportation planning models for urban networks have been receiving attention from researchers and practitioners. Traditionally, this problem was simplified by considering a four step procedure. The inconsistency of predicting the travel demand and determining the level of services becomes one of the principal issues for this procedure (Oppenheim, 1995; Garret and Wachs, 1996; McNally, 2000; Boyce, 2002). To avoid the flaw, one seeks an approach which states the travel behavioral assumptions, translates them into mathematical conditions, and finds solutions that satisfy these conditions. These models in the literature are referred to as 'combined' or 'integrated' models.

For solving combined models formulated as constrained convex optimization problems, the most widely used is the Evans algorithm (Evans, 1976), which is a descent algorithm for continuous optimization problems (Patriksson, 1994). Recently, Bar-Gera and Boyce (2003, 2006) extended the origin-based (OB) algorithm, which was presented by Bar-Gera in 2002, for solving the CDA model. Numerical investigations for the Chicago region road network demonstrated that the OB algorithm can solve the CDA model rapidly and more effectively in achieving highly accurate solutions than the Evans' algorithm.

In this paper, we study an origin-based algorithm to solve a combined distribution,

hierarchical mode choice, and assignment network model (CDHMCAN). The CDHMCAN model was first formulated as an optimization problem. A general algorithm for it is presented, which is based on the origin-based algorithm of Bar-Gera and Boyce (2003).

2. Model formulation

2.1 List of variables

It is assumed flows are averaged over the entire modeling period and over all work days during a specified year. The O-D matrix for travelers is determined by a doubly constrained gravity-type distribution model. The travelers from zone to zone can choose their trip modes according to utility values such as the duration of the trip, the mileage cost of the trip, and their preference of transport mode.

Consider a study area which is divided into zones, during a certain time period of the day in a given year. Let Z denote the set of all zones. For each pair of zone $p \in Z$ and zone $q \in Z$, let d_{pq} denote the O-D flow (persons/h) from p to q. $\boldsymbol{d}$ is the array of O-D flows. Let M denote the sets of mode classes in the network. The set of mode classes M is divided into a set of mutually exclusive groups, $G = (g_1, g_2, \Lambda, g_{N_G})$, such that $M = \mathop{\mathrm{Y}}_{\forall i \in N_G} g_i$, and for any two groups $g_i, g_j \in G$, $g_i \mathrm{I}\ g_j = \phi$, ϕ is an empty set, where N_G is the number of groups. These groups can be further classified into road-based mode groups, $G_1 \subset G$, and transit-based mode groups, $G_2 \subset G$, so that $G_1 \mathrm{Y}\, G_2 = G$ and $G_1 \mathrm{I}\ G_2 = \phi$.

d_{mpq} is travel demand of mode class m from zone p to zone q; d_{gpq} is travel demand of mode group g from zone p to zone q, $d_{gpq} = \sum_{m \in g} d_{mpq}$; d_{pq} is total travel demand from zone p to zone q, $d_{pq} = \sum_{g \in G} d_{gpq}$, $\boldsymbol{d} = \{d_{pq}\}_{p,q \in Z}$; O_p is total travel demand from zone p, $O_p = \sum_{q \in Z} d_{pq}$; D_q is total travel demand to zone q, $D_q = \sum_{p \in Z} d_{pq}$; $\boldsymbol{R}_{mpq}$ is the set of routes from zone p to zone q, $\boldsymbol{R} = \{R_{pq}\}_{p,q \in Z}$; h_{mpqr} is traffic flow of mode class m from zone p to zone q on path r, $\boldsymbol{h}_{mpq} = \{h_{mpqr}\}_{r \in \boldsymbol{R}_{mpq}}$, $\boldsymbol{h} = \{\boldsymbol{h}_{mpq}\}_{m \in M, p \in Z, q \in Z}$; f_{ma} is traffic flow of mode class m on link a; f_a is total traffic flow on link a, $f_a = \sum_{m \in g} f_{ma}$; δ_{mpqar} is equal to 1, if path

$r \in \boldsymbol{R}_{mpq}$ traverses link a; 0, otherwise; $\boldsymbol{\gamma}_{mpq}$ is a vector of non-negative route proportions (conditional probabilities), which describes the distribution of travelers from zone p to zone q among the routes in R_{pq}, $\boldsymbol{\gamma}_{mpq} = \{\gamma_{mpqr}\}_{r \in \boldsymbol{R}_{mpq}}$, $\boldsymbol{\gamma} = \{\boldsymbol{\gamma}_{mpq}\}_{m \in M, p, q \in Z}$; $\Gamma(\boldsymbol{R})$ is a set of all feasible route proportion arrays, $\Gamma(\boldsymbol{R}) = \left\{ \gamma \in [0,1]^{|\boldsymbol{R}|} : \sum_{r \in \boldsymbol{R}_{mpq}} \gamma_{mpqr} = 1, m \in Z, p, q \in Z \right\}$; t_{ma} is generalized travel cost on link a by travelers choosing mode class $m \in g \in G$; t_a is total cost of travel on link a, $t_a = \sum_{m \in g} t_{ma}$, $\boldsymbol{t} = \{t_a\}_{a \in \boldsymbol{A}}$; c_{mpqr} is generalized travel cost from p to q on path r by travelers choosing mode class $m \in g \in G_1$, $c_{mpqr} = \sum_{a \in A} \delta_{mpqar} t_{ma}$, $\boldsymbol{c}_{mpq} = \{c_{mpqr}\}_{r \in \boldsymbol{R}_{mpq}}$; $\hat{c}_{mpq}$ is generalized travel cost from p to q by travelers choosing class $m \in g \in G_2$; u_{mpq} is average generalized travel cost from p to q by travelers who choose mode class m. If mode $m \in g \in G_1$, $u_{mpq} = \boldsymbol{\gamma}_{mpq} \boldsymbol{c}_{mpq}$; and if $m \in g \in G_2$, $u_{mpq} = \hat{c}_{mpq}$. $\boldsymbol{u} = \{u_{mpq}\}_{m \in M, p, q \in Z}$; $\overline{U}_{mpq}$ is total travel disutility from p to q by travelers who choose mode class m. $\overline{U}_{mpq} = u_{mpq} + \varepsilon_{mpq}$, where $E[\varepsilon_{mpq}] = 0$; L_{gpq} is log sum of travel disutility as perceived by travelers who choose mode group g and travel from p to q. $L_{gpq} = -\frac{1}{\beta_2^g} \ln \sum_{m \in g} \exp(-\beta_2^g \overline{U}_{mpq})$, where β_2^g is calibration coefficient; L_{pq} is log sum of travel disutility as perceived by travelers traveling from p to q. $L_{pq} = -\frac{1}{\beta_1} \ln \sum_{g \in G} \exp(-\beta_1 L_{gpq})$, where β_1 is calibration coefficient.

2.2 Mathematical optimal formulation

Let free flow travel times with mode m traveling on link a be tt_{ma}^0, capacity of link a is k_a, length of link a is l_a; for each zone with mode m, parking costs is

pc_{mz} and walking times to or from the parking place is wt_{mz}, (For mode m in group G_1). For traveling by transit (in group G_2) from origin p to destination q (these are fixed regardless of flows), in vehicles, travel time is c_{mpq}^{ivtt}, fares is c_{mpq}^{fare}, and out of vehicle times is c_{mpq}^{ovt}.

Link travel time functions are of BPR form for mode m in group G_1

$$tt_{ma}(f_{ma}) = tt_{ma}^{0}\left[1 + 0.15\left(\frac{f_{ma}}{k_a}\right)^4\right] \tag{1}$$

Auto operating costs, including gasoline consumption, are a linear function of link length and travel time, $oc_{ma} = \eta_{m1} tt_{ma}(f_{ma}) + \eta_{m2} l_a$ (2)

Link generalized costs are $t_a(f_a) = \beta_{time}^{G_1} \sum_m tt_{ma}(f_{ma}) + \beta_{cost}^{G_1} \sum_m oc_{ma}(f_{ma})$ (3)

where the β's are calibration parameters. Parking costs and walking times are components of the additional costs (ac_{pq}) between p and q, defined as

$$ac_{mpq} = \beta_{park}^{mG_1} \frac{pc_{mp} + pc_{mq}}{2} + \beta_{walk}^{mG_1}\left(wt_{mp} + wt_{mq}\right) \tag{4}$$

Route generalized costs by mode m are $c_{mpqr} = ac_{mpq} + \sum_{a \in r} \sum_m t_{ma}$ (5)

OD generalized cost in group G_1 is $u_{mpq} = \gamma_{mpq} c_{mpq}$. The OD generalized cost in group G_2 is $\hat{c}_{mpq} = \beta_{bias}^m + \beta_{ivtt}^m c_{mpq}^{ivtt} + \beta_{fare}^m c_{mpq}^{fare} + \beta_{ovt}^m c_{mpq}^{ovt}$ (6)

The network equilibrium problem can be formulated as the following mathematical program.

$$\begin{aligned} \text{Minimize} \quad & F(\boldsymbol{d}, \boldsymbol{\gamma}) = \sum_{a \in A} \int_0^{f_a} t_a(\omega) d\omega + \sum_m \sum_p \sum_q d_{mpq} ac_{mpq} + \sum_{m \in g \in G_2} \sum_q \sum_p d_{mpq} u_{mpq} \\ & + \frac{1}{\theta_1} \sum_q \sum_p d_{pq}\left(\ln d_{pq} - 1\right) + \frac{1}{\theta_2} \sum_{g \in G} \sum_q \sum_p d_{gpq}\left(\ln d_{gpq} - 1\right) \\ & + \frac{1}{\theta_3} \sum_{m \in M} \sum_q \sum_p d_{mpq}\left(\ln d_{mpq} - 1\right) + \sum_{m \in M} \sum_q \sum_p d_{mpq} \varepsilon_{mpq} \end{aligned} \tag{7}$$

Subject to:

$$\sum_{q} d_{pq} = O_p, \quad p, q \in Z \tag{8}$$

$$\sum_{p} d_{pq} = D_q, \quad p, q \in Z \tag{9}$$

$$\sum_{m \in g} d_{mpq} = d_{gpq}, \quad p, q \in Z, \quad g \in G \tag{10}$$

$$\sum_{g \in G} d_{gpq} = d_{pq}, \quad p, q \in Z \tag{11}$$

$$\sum_{r \in \boldsymbol{R}_{mpq}} \gamma_{mpqr} = 1, \quad p, q \in Z, \quad m \in g \in G_1 \tag{12}$$

$$f_{ma} = \sum_{p} \sum_{q} \sum_{r \in R_{mpq}} \delta_{mpqar} \gamma_{mpqr} d_{mpq}, \quad a \in A, \quad m \in g \in G_1 \tag{13}$$

$$f_a = \sum_{m \in g \in G_1} f_{ma}, \quad a \in A \tag{14}$$

$$\gamma_{mpqr} \geq 0, \quad p, q \in Z, \quad r \in \boldsymbol{R}_{mpq}, \quad m \in g \in G_1 \tag{15}$$

$$d_{pq} = 0, \quad d_{gpq} = 0, \quad d_{mpq} = 0, \quad p, q \in Z, \quad g \in G, \quad m \in M \tag{16}$$

Parameters $\theta_1, \theta_2, \theta_3$ in Equation (7) are calibration coefficients.

3. An OB algorithm for CDHMCAN model

The general scheme of the algorithm is presented in Figure 1. As shown in the figure, the main addition to the origin-based algorithm for the CDA model is a procedure to update mode OD flow with HMCM, while retain route proportions.

Initialization:

Find the initial OD flows

for each p in Z do

A_p = tree of minimum cost routes from p

f_{ap} = All-or-Nothing assignment using A_p with the initial OD flows

end for

Main Loop:

for $n = 1$ to number of main iterations (I_{Main})

Update OD flows with GTDM, retain route proportions

Update mode OD flows with HMCM, retain route proportions

Update approach proportions $\{\alpha_{ap}\}$ for each link a

for p in $\boldsymbol{Z}$ do

Update restricting subnetwork A_p

Update approach proportions $\{\alpha_{ap}\}$ for each link a

end for

for $m = 1$ to number of inner iterations (I_{Inner})
  for each p in $\boldsymbol{Z}$ do
    Update approach proportions $\{\alpha_{ap}\}$ for each link a
  end for
 end for
end for

Figure 1. Origin-based algorithm for the CDHMCAN model

4. Conclusions

Much of the attention has recently been focused on the OB algorithm due to its modest memory requirements, computational efficiency, and ability to achieve highly accurate detailed solutions. As a further study, this paper presented an OB algorithm for solving a combined distribution, hierarchical mode choice, and assignment network model with single user and multiple mode classes.

Acknowledgement

The first author greatly appreciates the unselfish guidance on the urban travel forecasting research of Professor David Boyce. The study is supported by the National Basic Research Program of China (Project No. 2006CB705500) and the National Natural Science Foundation of China (Grant No. 70631001).

References

Bar-Gera, H. (2002). "Origin-based algorithm for the traffic assignment problem." *Transp. Sci.,* 36(4), 398-417.

Bar-Gera, H., and Boyce, D. (2003). "Origin-based algorithms for combined travel forecasting models." *Transp. Res.,* 37(5)B, 405-422.

Bar-Gera, H., and Boyce, D. (2006). "Solving a non-convex combined travel forecasting model by the method of successive averages with constant step sizes." *Transp. Res.,* 40(5)B, 351-367.

Evans, S.P. (1976). "Derivation and analysis of some models for combining trip distribution and assignment." *Transp. Res.,* 10(1), 37-57.

Garret, M., and Wachs, M. (1996). "Transportation Planning on Trial," Thousand Oaks, California.

Oppenheim, N. (1995). *Urban Travel Demand Modeling*, John Wiley and Sons, Inc, New York.

Ortuzar J. de D., and Willumsen L.G. (2001). *Modeling Transport*, 3rd Edition, Wiley, New York.

Patriksson, M. (1994). *The Traffic Assignment Problem: Models and Methods*. VSP, Utrecht, The Netherlands.

Sheffi, Y. (1984). *Urban transportation networks: equilibrium analysis with mathematical programming methods*, Prentice Hall, New Jersey.

COMPUTING JOB ACCESSIBILITY WITH INTEGRATED GIS AND DBMS

Jiawen YANG

Georgia Institute of Technology, Room 204, Architecture Building East, Atlanta, GA, 30332-0155. phone: 404-385-7215; email: jy78@mail.gatech.edu

Abstract: Accessibility is the key concept for transportation and land use planning. However, the difficulty in computing accessibility measures has prevented it from playing a more central role in planning decision making. This paper explores the possibility to compute several kinds of job-accessibility measures with a coupling of user-friendly geographical information systems (GIS) and high performance database management systems (DBMS). The selected case is the Boston Metropolitan Region. The gigantic dataset, Census Transportation Planning Package (CTPP) for the metropolitan area, is stored in the Oracle sever. Both the database server and the GIS client are programmed to streamline the computation and the interaction with the user. This customized system can provide useful experiences for the public who want to know the accessibility impacts of various transportation and land use decisions.

Key words: Accessibility, Database management systems, Geographical information systems.

1. Introduction

In today's planning agenda, topics like urban sprawl, air quality, transit oriented development, highway congestion, open space loss and social segregation, are so heatedly debated that a lot of terms which previously belong to the planning professionals are now frequently appearing in the every day language of the public. During this process, researchers and urban planners have exploited the potentials of new information technology to help facilitate public participation and group discussion by distributing information and knowledge in time and in an easily understandable format. The visualization of traffic condition through web multimedia (Shiffer, 1995) and the scenario-based, policy-oriented planning support system *what if*, which is used to facilitate collective decision making on land use planning (Klosterman, 2001), are examples of this broad stream of efforts

Among the various topics, literature highlights the interaction of urban transportation and land use planning, in the core of which accessibility lies. In transportation research, mobility and accessibility has long been identified as critical issues by the TRB Executive Committee (1997). The need to integrate transportation and land use planning has been repeatedly emphasized. Researchers have called for a paradigm shift from automobility oriented planning toward accessibility oriented planning

(Cervero, 1996). However, accessibility, the key concept of the new paradigm, is still hardly understood or widely adopted in planning decision making for several reasons. Firstly, the notion of accessibility is ambiguous and has been associated with various measures, which are easy to talk but not easy to quantify. The gravity –type accessibility measures, for example, can range from the simplest relative accessibility to integrated accessibility, from supply-side accessibility measure to demand adjusted accessibility measure, from zonal based measures to individual-based measures, from measures based on social physics to those based on individual economic behavior (Morris, Dumble and Wigin, 1979). Usually, the more scientific the measure claims to be, the more difficult to understand it conceptually. Secondly, the computation of accessibility measures usually involves large data set and extensive computing capacity, which is usually inaccessible by the general public and even many planning professionals.

The shift toward a planning model based on accessibility has to be conditioned on the ability to measure and present accessibility before the public. Therefore, computation support on accessibility measures is necessary to facilitate public participation. Through this channel, accessibility may have a better chance to enter the core of land use and urban transportation planning in practice. Toward this end, this paper presents an experiment that supplies the computation capacity for various accessibility measures, which could meet the demand for accessibility information in group discussion.

2. Data and Accessibility Measures

This experiment uses CTPP (Census Transportation Planning Package) commuting data for the Boston Metropolitan Area to compute job accessibility. The data contains information on commuting flow at the level of transportation analysis zones (TAZ). Commuting time between each TAZ pair for both transit riders and auto drivers is estimated based on census long form questionnaire. The advantage of this data set is that it provides necessary information to compute almost all sorts of job accessibility measures if values of required parameters could be correctly supplied. The inclusion of travel time by travel mode makes it especially interesting for computing job accessibility measures because automobile drivers have different levels of accessibility and mobility compared to transit riders. This data set, when combined with other socioeconomic data, can provide accessibility measures for more detailed social group classification. Furthermore, the data are spatially fine-grained. With this data set, the Boston region is divided into around 800 TAZs. Measures at the TAZ scale help reveal the spatial variation of accessibility across different neighborhoods. This experiment computes three categories of job accessibility measures.

a. **Cost based measures**

Planners generally want to know the average commuting time from the interest area to some or all other parts of a region (Ingram, 1971).The interest area could be at different geographical scale. It could be downtown Boston or suburban towns like Framingham. It could be composed of only one TAZ or a group of TAZs. This is to

ask for a relative accessibility measurement only for one TAZ or a group of TAZs.

b. **Accumulative opportunity measures**
Another way to evaluate job accessibility is to find out how many job opportunities are reachable within a certain commuting time from the interest area. This is to ask for a measure of accumulative opportunity within a certain transportation effort. When using this measure to evaluate accessibility, different individuals may want to specify different commuting time.

c. **Demand adjusted accessibility measures**
This experiment also computes the complicated demand adjusted accessibility measures, a measure proposed by Shen (1998). The demand adjustment makes it convenient to compare job access advantage among different locations and different social groups, as for now job supply is normalized by job demand.

Even though this experiment covers only a small set of accessibility measures, the selected three categories are representative for computation purpose. They range from the simplest one, cost based accessibility measure, to the most computing intensive one, demand-adjusted job accessibility. In order to compute a demand adjusted accessibility measure for one TAZ, the computer has to call an exponential function for 1600 times, i.e. two times the number of TAZs.

Even though these accessibility measures can be computed in advance and then stored in the database. There is significant advantage to implement a capacity to compute the measures on the fly, as different individuals may have slightly different rules for defining job accessibility. For example, people may have different spatial decay parameters for gravity-type accessibility measures, or different threshold travel time for accumulative opportunity measures, or they may have different definition of the interest areas.

3. System Consideration

Easiness to use and performance are the two major factors considered in designing the system. They are required by the context of computation. The proposed system should be simple and straightforward, assuming no pre-knowledge of the algorithm and computing environment. The user should be able to easily input parameters to define the interest area and the measures without worrying about how the computation actually works and how the result is returned. The system should also have a high-performance computing capacity that will process requests fast enough and allow group discussion to go on without significant interruption.

These two concerns can hardly be satisfied with standalone GIS or DBMS. DBMS lacks friendly interface to accept input from the user and the capacity to visualize the final result. GIS, on the other hand, lacks the capacity to efficiently deal with the computation with large data set. Take the Boston data as the example. Commuting matrix is the Cartesian product of basic spatial units. Since the data are spatially fine

grained, the data volume for the commuting matrix increases quadratically as the number of spatial units increases. The CTPP commuting data in Boston MSA has a volume over 120 MB when in tab-delimited flat file. The performance could be very bad in a file based standalone GIS system. This drawback is quite serious. As argued by Harris and Batty (2001), computing performance has been an important issue in using planning support systems. Fortunately, today's DBMS is already well developed in handling large data set. For example, a typical table in Oracle could have a volume greater than 2GB (Koch and Loney, 1997). Therefore, the integration of GIS and DBMS is the ideal model to deal with this situation. The use of DBMS is also justified by the nature of the problem: access to the data and computation capacity has to be managed while making them accessible to the public. Enterprise DBMS has the advantage to manage data access as well as mange access to customized computation capacity, an example of which will be discussed in the next section.

4. System Description

The system includes a high performance UNIX server with Oracle and Spatial Database Engine (SDE) ready to handle geographical and non-geographical data, and PCs where ArcGIS 8.2 is installed on an MS Windows operating system. SDE is used to integrate the visualization capacity of ArcGIS and the computation power of Oracle. Besides the CTPP commuting table, a GIS layer of TAZ boundaries and a simple table containing the TAZ list are also stored in the DBMS.

4.1 Procedures in DBMS

The key work done at the data server side is the function created to do the computation. Functions to compute average travel time, job clusters and demand adjusted job accessibility are encapsulated in the same table space as that for the tabular tables. These functions are the elements to compute the accessibility measures. They can be called in standard SQL statement in the same way as we use MAX, MIN, COUNT, the functions coming with standard Oracle package. For example, the following statement shows how the job cluster function is called to compute the accumulative job opportunity in Oracle.

```
Select TAZ, CTPP.JobCluster(TAZ, TravelTime) JobNum
From CTPP.TAZlist
Where zone in TAZSet;
```

This SQL statement computes job cluster for every selected TAZ within a certain travel time. TAZlist is the Oracle table that lists all TAZs in the Boston Metropolitan Area. CTPP.JobCluster is the function created at the server side. TAZSet is the collection of the identifiers of a selected group of TAZs, of which the interest area is composed. TravelTime is the threshold commuting time. JobClulster (1, 20) will calculate the number of jobs reachable within 20 minute commuting time from zone 1. TAZSet is the set of the TAZ identifiers passed from the client, which the user can define using graphic feature selection tool on the ArcGIS interface. In the Oracle

sever, the function JobCluster() could be made accessible to users in an easier manner, just by granting the right of execution. Therefore, the access right to data and functions can be easily managed.

4.2 Client customization

Customization at the client side is designed to enable users to concentrate on what they are interested in without any presumed knowledge of the underlying mathematics and the networked computing environment. Users may even not know the basics of ArcGIS when they define the above parameters for the above SQL statement.

Firstly, four buttons are added to the tool bar within a group named “Accessibility”, as illustrated in the rectangle in Figure 1. One of them adds the GIS layer of TAZ boundary when it is clicked. The other three provoke the functionalities to compute the three categories of accessibility measures mentioned previously. Corresponding VBA script is attached to each button.

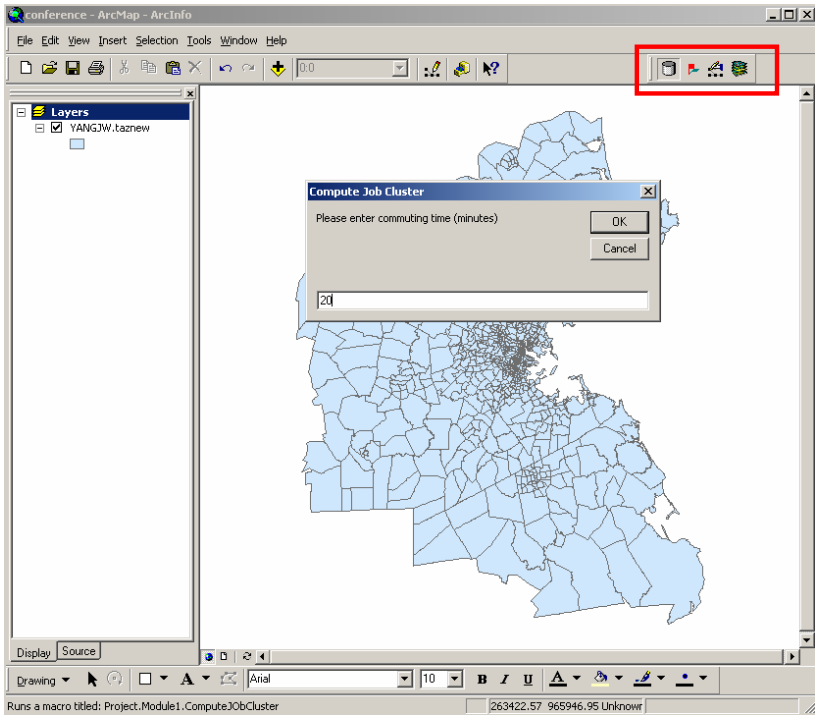


Figure 1. An Overview of the Client Interface

Secondly, interest areas and parameters are specified with the standard selection tool in ArcGIS or simple input box. For example, if a user click the first button, which will provokes the functionality to compute the accumulated opportunity measure, the ArcGIS interface will pop up the above simple input box, asking for a threshold commuting time (Figure 1). The specified travel time will be used as one of the two parameter values for the JobCluster () function. If the user also specifies a collection of TAZ with the graphical feature selection tool, the VBA script will find the identification associated with each selected TAZ, construct the set of the identifiers, i.e. TAZSet, and then further constructs the SQL statement. The completed SQL statement will then be sent to the Oracle server for computation.

Thirdly, when the result is returned from the Oracle server as a table, a relationship class is defined in the VBA code and the table will be joined to the GIS layer on the fly. The user will get a complete layer with accessibility measure for ever TAZ if he selects no TAZ. He can also retrieve the information with the standard identify tool on ArcGIS' toolbar. Figure 2 illustrates the accumulated job accessibility measure with 20 minutes threshold time.

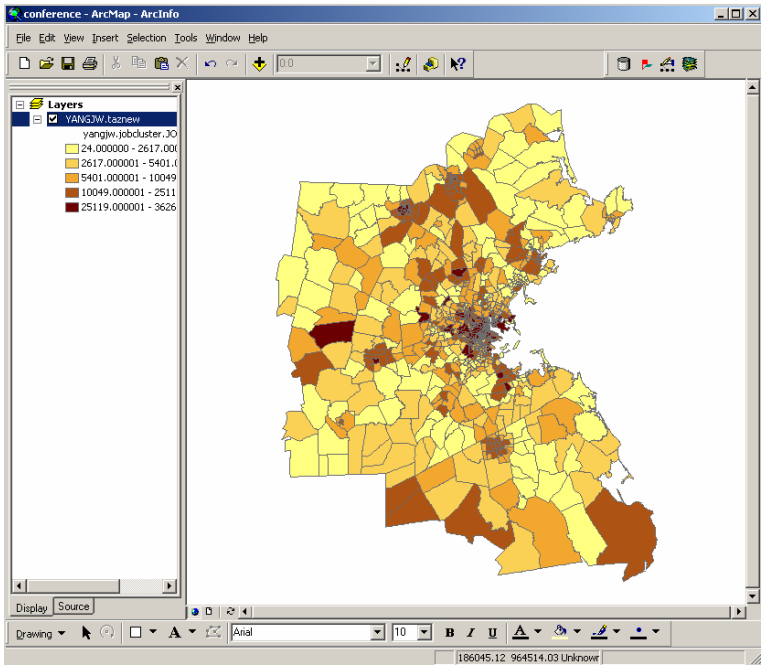


Figure 2. A map of accumulated job accessibility measure with 20 minutes threshold time

4.3 Computing process integration

For every computing request submitted by the user, the whole computing process is distributed between the client and the server, who are further banded together with SDE. The following chart (Figure 3) illustrates the process to compute job accessibility in the integrated environment. The control logic is implemented with VBA script.

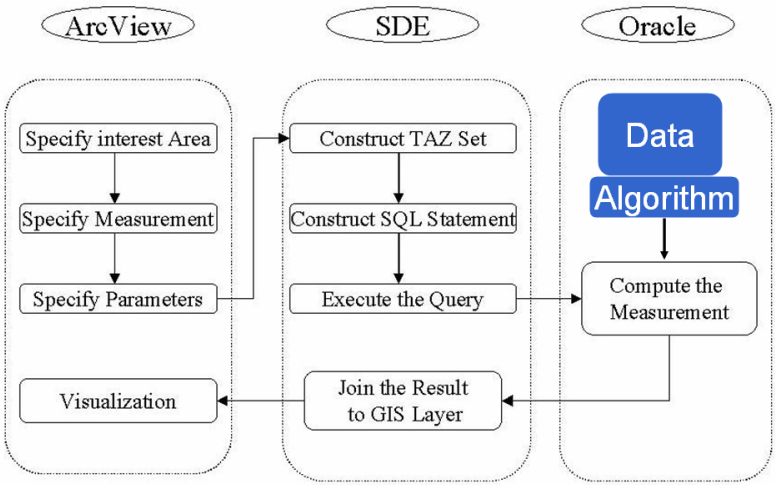


Figure 3. Computing Process

From the programmer's perspective, SDE provides a seamless network computing environment, which is realized by opening an SDE workspace in this case. There is no direct dialogue between the GIS client and the DMBS server. Everything is done in the SDE workspace. The communication between SDE and Oracle could be hidden to the programmer. The GIS data is accessed via SDE, the computation request is sent to DBMS via SDE, the result is maintained in DMBS and accessed via SDE.

4.4 Performance

The computing result is usually returned in several seconds, even for the most computation intensive measure, the demand adjusted job accessibility. Several features of the system help to guarantee high performance. Firstly, the work is distributed between the server and the client, aiming to provide fast computation and easy-to-use interface, as we have talked about in the above sections. Secondly, the system is designed to have the flexibility to reduce waiting time by limiting the interest area to a smaller one. Instead of asking for job accessibility for every TAZ in the MSA, a

typical user can select only a subset of TAZs that are of his special interest.

5. Conclusions

The use of DBMS for planning is not a new topic (Ferreira, 1990), neither is the integration of GIS and DBMS. In recent years, researcher and planners have tried to integrate DBMS and GIS for planning issues where computation and visualization are both important (Cao and Ferreira, 2001). However, instead of relying on a component GIS model that usually involves manual transition between GIS and DBMS, this work experiments an integrated model to automate the process.

The implication of this experiment could be well revealed by the idealized and abstract view of planning support system proposed by Klosterman (1997). He suggests that planning support system should be "designed to facilitate collective design, social interaction, interpersonal communication and community debate which attempt to achieve collective goals and deal with common concerns". It should also "pay particular attention to long-range problems and strategic issues". In today's planning agenda, the interaction of urban transportation and land use planning is a really long-range and strategic issue, but its core concept, accessibility, is still vague and the related computation capacity is still costly. This paper illustrates how a possible model could be designed to deliver computing support on job accessibility in an environment where computation result is demanded on the fly. This model is not far advanced, but it could be a good starting point for further work. For example, computation functionality can be expanded by adding more components to the DBMS server. Without any change on the server side, the system can be migrated to the web. The user can use web browser to type in input, pass request, and visualize returned results.

Transportation and land use planning also involves testing the outcome of various alternatives at the stage of sketch planning. However, realistic modeling requires more and more data set with detailed spatial information to represent the complicated urban landscape. At the same time, computation becomes more extensive to model the interaction of various factors for a more believable forecast result. How can we manage the data? How can we distribute various computing processes? How can we make the data and computation power more accessible to the public? The integration of GIS and DBMS is a possible answer. Binding the computation power of DBMS with GIS reduces the cost and makes computing capacity more accessible to the public and the planning professionals. High system performance can facilitate planning discussion on various alternatives in sketch planning.

References

Cao, L., and Ferriera, J. Jr. (2001). "Integrating GIS and RDBMS to Model Traffic Congestion and Urban Air Pollutants," in *Urban and Regional Systems Association Proceedings*, editors: Mark, S. Long Beach, CA, pp. 678-688.

Cervero, R. (1996). "Paradigm Shift: From Automobility to Accessibility Planning,"

Working Paper 677, University of California at Berkeley, pp.1-22.

Ferreira, J. Jr. (1990). "Database Management Tools for Planning," *Journal of the American Planning Association*, 56:78-84.

Harris, B., and Batty, M., (1993). "Locational Models, Geographic Information and Planning Support Systems," *Journal of Planning Education and Research*, Vol. 12, pp.184-198.

Ingram, D.R. (1971). "The Concept of Accessibility: a Search for an Operational Form," *Regional Studies*, Vol.5, pp. 101-107.

Klosterman, R.E. (2001). "The *What If* Planning Support System," in *Planning Support Systems: Integrating Geographical Information Systems, Models and Visualization Tools*, editors: Brail, R.K. and Klosterman, R.E., ESRI Press: Redland, California.

Klosterman, R.E. (1997). "Planning Support Systems: A New Perspective on Computer-aided Planning," *Journal of Planning Education and Research* Vol.17, pp. 45-54.

Koch, G. and Loney. K. (1997). *Oracle 8: The Complete Reference*, McGraw-Hill: Berkeley, California.

Morris J. M., Dumble, P.L., and Wigan, M.R. (1979). "Accessibility Indicators for Transportation Planning," *Transportation Research A*, Vol.13, pp.91-109.

Shen, Q. (1998). "Location Characteristics of Inner-City Neighborhoods and Employment Accessibility of Low-Wage Workers," *Environment and Planning B*, Vol. 25, pp.345-365.

Shiffer, M.J. (1995). "Interactive Multimedia Planning Support: Moving From Stand-alone Systems to the World Wide Web," *Environment and Planning B: Planning and Design*, Vol. 22, pp. 649-664.

TRB Executive Committee (1997). "Critical Issues in Transportation Planning," *TR News* 174, Sept-Oct. pp1-11.

STUDY ON THE RELATIONSHIP BETWEEN HIGHWAY PASSENGER TRANSFER TERMINALS PLANNING AND URBAN SPATIAL DEVELOPMENT

GUO, Xiucheng[1], FAN, Jun[2], CHEN, Shuangyan[1], ZI, Haibo[1], HUANG, Hainan[1]

[1]Transportation College, Southeast University, Nanjing 210096;
email: seuguo@163.com
[2]Nanjing Academy of Urban Planning &Design Co. LTD, Nanjing 210029; email: Lookerfanjun@sina.com

Abstract: This paper analyzes the relationship between the layout of highway passenger transfer terminals and urban spatial development, discusses the interactive mechanism, and puts forward a new method to research the layout of highway passenger transfer terminals. The method has been applied in the layout of the highway passenger transfer terminals in Jiangyin and obtained a good result. Thus this paper advances a new direction for the layout of highway passenger transfer terminals with the urban spatial distribution.
Keywords: highway transfer passenger terminals, urban spatial distribution, layout planning

1 Introduction

The highway passenger transfer terminals are the very important transportation infrastructure and the link point of urban passenger transfer means and external passenger transfer means. With the expansion of the city, the growth of floating population, the diversification of traffic modes, and the increase of traffic pressure in main cities, the layout planning of urban passenger transfer terminals has been faced with more and more problems. A series of problems such as 1) the overall arrangement are centralized, 2) the function facilities are not reasonably installed, and 3) the operation benefit is unfavorable, all these should be urgently resolved. Therefore, it's necessary to study the layout, construction and operation of highway passenger transfer terminals under the new background.

The layout of the highway passenger transfer terminals is influenced by many factors and urban spatial distribution is one of the important factors that needs to be emphasized. The former theories focus much on the qualitative analysis about the influence of urban spatial distribution, while the project practices place more emphasis on the requirement and the conceiving of the passenger transfer terminals, and it is short of the analysis about interactive action between the urban spatial distribution and the layout of the passenger transfer terminals, especially the quantitative evaluation

about their relationship. This paper puts forward the quantitative analysis method and has applied it in practice, based on the qualitative analysis. This provides a new direction for thinking method about the layout of passenger transfer terminals.

2 The interactive relations between highway passenger transfer terminals and urban spatial development

The highway passenger transfer terminals have to be relied on by cities, as large numbers of passengers come from cities, and the terminal construction and operation demand the cities to provide the ground and funds, as well as the terminals need the urban public transport to aggregate and disperse the passengers, and the traffic organization for passenger transfer terminals requires the urban road network and highway network to support.

Therefore, the layout planning of the urban highway passenger transfer terminals is restricted by the influence of the urban scale, development directions, urban morphology, and the main use zoning distribution. Thus much more about these factors should be considered and understand the inner mechanism between the layout of the passenger transfer terminals and the urban spatial distribution.

2.1 The impact of urban spatial distribution on the layout of the passenger transfer terminals

(1)Urban scale

The urban scale means the difference in amount and hierarchy of the subject and element gathering in urban spatial allocation. It mainly includes three relevance organic parts, such as urban population, economic activity and capability, and urban built-up area. The urban scale is the performance and generalization of the city's characteristic quantity, reflecting the gathering and diffusion of the city element, and it is a quantity concept with the significance of social economy and environment.

With the development of urbanization, the scale of city has expanded, the frame of the urban space has enlarged, and the urban built-up area has widened. The nature of land utilization and the exploitative intensity of land resources has made great changes, as well as the industry, financing and population have gathered in these areas. These changes have led to the new demand to the scale and layout of the urban highway passenger transfer terminals. It can be concluded that the layout of the urban highway passenger transfer terminals develops along with the change of urban scale.

(2)Urban morphology

Urban morphology means the development of the urban spatial form is mutually affected by many factors, such as its natural environment, history, politics and economy, society, science and technology, and culture etc over a period of time.

The distribution form of the city affects the layout of the highway passenger transfer terminals to some extent. The urban morphology appears different according to historical development or policy regulation, mainly containing belt form, lumpy

form, group form and constellation form etc. For different urban spatial morphology, the layout of highway passenger transfer terminals needs to adopt different layout modes according to the spatial characters. Thus the function of terminals could be fully exerted, and meet residents' traffic demands. So it is reasonable that urban spatial morphology determines the layout forms of urban highway passenger transfer terminals.

(3)Urban development directions

Urban development directions point out the focus of urban construction, and it is the key question of urban development, as well as playing a role in guiding and restricting the urban development. This basic principle should be implemented in the city development process.

While holding the urban development directions, we make clear, the way of urban expanding, the direction of new large-scale urban construction, and the shift focus in mass population, Industry, financing. While the layout planning of urban highway passenger transfer terminals keep in accord with urban development directions, it could ensure the terminal system cover more areas, have more passengers, and guard against the exploitation of the land around the terminals.

(4)The main use zoning distribution

The highway passenger transfer terminals should be located near the passenger sources, and the distance controlled under the sustainable limits. In this way, they can supply high quality service, and keep effective operation. Thereby, it is the most important factor to ensure passengers arrive in the terminals conveniently.

The investigation shows that the main inhabitation regions, government agency, central business districts, and college and university are the main passenger orientation areas. The spatial layout of these areas determines the passenger volume and its distribution, thereby going a step further to restrict the layout of the passenger transfer terminals greatly.

2.2 The feedback of the layout of the passenger transfer terminals to urban spatial distribution

Passenger transfer terminals are the important node of the urban road network, as well as an area that involves diversified building blocks with centralized facilities and open space. The area around the passenger transfer terminals has become the new growth point of the city, and brought concentration effects. They are described as cluster zones of terminals including themselves and the surrounding areas. The essential part of the concentration effects is that there is benefit gradient field under definite range. The concentration effects will gradually attenuate from center to the outside, following the distance decay rate. It could be figured by logarithmic attenuation function in theory:

$$d = f(e) = a \ln \frac{a \pm \sqrt{a^2 - e^2}}{e} \mu \sqrt{a^2 - e^2} \quad (1)$$

Where: d ——the distance out from the terminal，unit: meter；

e ——the concentration effects, it could use different measuring method according to different research goals；

a ——is a constant, describes the maximum concentration effects.

The function graphics of Formula (1) is illustrated in Figure 1, as the distance rises from d_1 to d_3, the concentration effects reduce from e_1 to e_3.

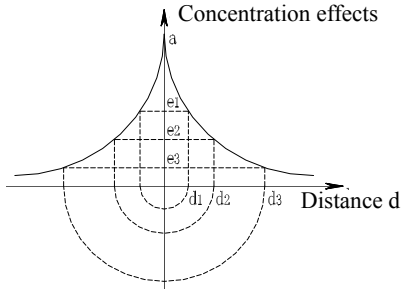


Figure1. Concentration effects distance attenuation function

The concentration effects intensity varies along with the grade and scale of the terminals, and it determines the development models, inhabitant distribution and trip modes in the surrounding area of terminals.

The highway passenger transfer terminals are the transportation infrastructure. Their construction should have prospective thought and adequately reflect the traffic leading function. Meanwhile, the area around the terminals with dense passenger flow has strengthening concentration effects. The effects are able to advance the development of land use and industry especial service industry, spur infrastructure construction, and drive the expansion of urban space.

The layout of urban highway passenger transfer terminals has brought the highway passenger network covering the local region,. It, therefore, enhances the traffic accessibility, expands the urban economic region, and improves the traffic location. Thus people could make abundant development based upon the terminals, and promote the development of new urban zones.

3 Highway passenger transfer terminal layout model based on spatial analysis

Consider the mutual influence between highway passenger transfer terminals distribution and the development of urban space development synthetically, then

search for methods that take these impacts into account to do urban spatial analysis for the rational terminal planning.

3.1 Model principle

(1)Guidance on macro level

Considering the urban development and urban morphology which plays a great role in the layout of highway passenger transfer terminals, the model embodies coordination and is forward-looking on a macro level and gives qualitative guidance.

(2)Land use coordination on micro level

The model incarnates the impact of districts which have high generating rate of highway passenger flow and spatial distribution of facilities on the layout of highway passenger transfer terminals. The model also implements the principle of integrating transport and land use and the principle that the terminals should get closed to the passenger flow barycenter.

3.2 Model calculation method

Considering the urban morphology effect on the layout of highway passenger transfer terminals and the principle mentioned above, the authors put forward the calculation method.

(1)Traffic districts division

According to the functions of districts, the distribution of land, and the characteristics of highway passenger transport, we divide the region into small traffic districts. If the corresponding traffic survey has been done, the results can also be used to the original district, which can reduce workload after merging and appropriate adjustments.

(2)Passenger flow distribution

According to the survey data, the model calculates out the highway passenger generating intensity of every kinds of land. Then the model figures out passenger flow distribution with the capacity of planning land.

$$q_i = \sum_{k=0}^{1} q_{ik} \qquad (2)$$

$$q_{ik} = \frac{\sum_{j=1}^{n} p_{jk} b_j s_{ij}}{\sum_{i=1}^{N} \sum_{j=1}^{n} p_{jk} b_j s_{ij}} \qquad (3)$$

Where: q_i ——the ratio between highway passenger flow in District i and the whole region;

q_{ik} ——the ratio between Sort k highway passenger flow in District i and Sort k highway passenger flow in the whole region;

p_{jk} ——Sort k highway passenger flow generating rate of Type j land use;

b_j ——the cubage intensity of Type j land use;

s_{ij} ——the acreage of Type j land use in District i;

i ——the number of districts， $i=1,2\ldots N$;

j ——the type of land use $j=1,2\ldots n$;

k ——the sort of highway passenger flow, $k=\{0,1\}$, If $k=0$, it means the passenger flow is local population; and if $k=1$,it means floating population;

n ——the sum of land use types;

N ——the sum of districts.

(3)Select the terminal layout patterns and analyze reasonable siting region

Considering urban scales, urban morphology and urban development directions synthetically, and integrating the terminals' status and functions in comprehensive passenger transport system, the paper gives some more reasonable terminal layout patterns which present the practicability and are forward-looking. Then use an analysis model to work out reasonable siting region. The model is described as that:

$$X_M = \sum_{i\in M} q_i x_i \Big/ \sum_{i\in M} q_i \tag{4}$$

$$Y_M = \sum_{i\in M} q_i y_i \Big/ \sum_{i\in M} q_i \tag{5}$$

Where: q_i ——the proportion of highway passenger flow in District i;

(x_i, y_i) ——the coordinate of passenger center of District i

(X_M, Y_M) ——the coordinate of passenger center of District M in theory;

M ——the passenger flow region under different kinds of terminals modes

The spatial analysis model of highway passenger transfer terminals which is based on the interactive relations between highway passenger transfer terminals and urban spatial development, can feedback the impact of urban morphology in quantity. We can choose the reasonable terminals layout modes and the siting regions by the model, adapting and guiding the city's development.

4 Application Example

With the development of urban economy, the urban scale, morphology, characteristic and function of Jiangyin city have changed a lot. The urban configuration has altered from centricity to belt configuration, and the aggregating feature of industry, population and other resources have changed, which lead to the transition of trip distribution. In order to meet the needs of urban development, Jiangyin city decided to adjust the highway passenger transfer terminal planning.

The method mentioned above was applied in the layout planning of highway passenger transfer terminals in Jiangyin city and has achieved good results.

(1)Traffic districts division

Considering the land use types, the dividing line by nature condition (such as rivers, lakes) and urban streets, and the regionalism, we divided the urban area of Jiangyin into 36 districts, showed at the Figure2.

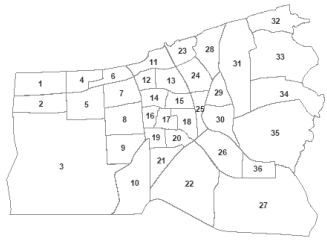


Figure 2 The traffic districts of Jiangyin City

(2) Passenger flow distribution

Using the formula(2) &(3), we obtain the proportion of highway passenger flow .The results are shown below in Table 1.

Table1 the proportion of highway passenger flow unit：100%

District number	1	2	3	4	5	6	7	8	9
Local population	0.89	0.64	0.00	0.42	6.92	1.03	6.17	7.22	16.96
Floating population	2.66	1.90	0.00	1.26	5.06	1.43	5.89	4.70	20.16
District number	10	11	12	13	14	15	16	17	18
Local population	3.37	3.09	1.89	4.48	3.47	3.71	1.53	3.34	2.89
Floating population	3.74	3.80	1.83	3.46	4.39	2.63	1.12	1.66	1.79
District number	19	20	21	22	23	24	25	26	27
Local population	1.20	2.51	2.88	0.00	0.78	5.62	1.61	0.00	0.00
Floating population	0.59	1.25	2.30	0.00	0.39	8.57	0.80	0.00	0.00
District number	28	29	30	31	32	33	34	35	36
Local population	0.59	3.07	1.67	1.80	0.00	2.89	0.43	2.53	4.42
Floating population	1.65	1.97	2.28	3.63	0.00	3.48	1.27	1.58	2.78

(3) Select the terminal layout patterns and determine reasonable siting region

Considering the urban spatial development, the overall plan, the general transportation planning, and the city highway and highway net planning, we put forward four terminal layout patterns to select. Then calculated the perfect layout situs of all layout patterns by formula (4)&(5). We always choose 500-1000 miles as the direct effect region of highway passenger transfer terminals, and found out the rational siting region. The recommend patterns and the reasonable siting regions are described as follows.

（1）A main station in north, an assistant station in south. This mode of terminal layout is near the north and south sides of the central axis in the region. It will perfectly serve the passengers either from east or west side. As Xicheng highway is on the city central axis, this mode is close to the North and South exits of the city, convenient for rapid routes in and out of cities.

（2）Two main stations in east and west, an assistant station in south. The east and west terminals as the distribution center of their respective zones, bear the main passenger flow and attract tourists from the center districts. The south assistant terminal is built to link up the reserved rail inter modal station.

（3）Two main stations in the east and west. This mode fits the urban morphology which is like shoulder pole. These two terminals as the distributing center of their respective zones, bear the main passenger flow and attracting tourists from the center districts. Passengers from the south district can choose the nearest terminals.

（4）A main station in middle, two assistant stations in east and west. In this mode, the main station bearing long-distance routes is good for conformity of long-distance routes and effective utilization of resources. The two assistant stations bear short-distance routes from respective zones which is convenient for passengers from every district.

According to the results calculated by the model introduced above, and acquirability of land use by planning department, we can determine the station sites of every mode, and get the layout projects then use the analytic hierarchy process or other analysis methods to compare all projects, and make out the rational layout project.

5 Conclusion

The paper analyzes the relationship between the layout of highway passenger transfer terminals and the urban spatial distribution, generalizes the elements that urban morphology influence the layout of highway passenger transfer terminals, and puts forward a new method used in the analysis of the terminals' layout which was applied in the layout of highway passenger transfer terminal in Jiangyin city and achieved a good result. The method analyses the impact elements in quantity, and advances a new aspect of terminal planning theory. However, further studied are still necessary on optimizing the model.

References

Eric J. Miller (1984). "Urban Transportation Planning – A Decision-Oriented Approach". McGraw-Hill Book Company.

Levinson, David M (1995). "Evolutionary transportation planning models: structure and application". Transportation Research Record 1493.

Liu lingling, Zhou tianyong (2006). "A new understanding of the city scale theory ". Economic Survey, No.1, 112-115.

Peter White (1995). "Public Transport-its planning management and operation". UCL Press Ltd.1995.

Yuan hong, Lu huapu (2006). "Study on model and method of comprehensive transportation terminal planning". Journal of Highway and Transportation Research and Development, Vo ol.18 No.3, 101-105.

Zheng xin, Lin lin (2002). "A review of studies on the urban morphology since 1990s". City Planning Review,Vol.26 No.7,59-64

Research on the Equivalent Capacity of Urban Streets and Its Validation

Minhua SHAO[1]; Lijun SUN[2]; and Xiaoning ZHANG[3]

[1]Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai, P.R. China, Siping Rd. No. 1239, 200092; PH: (8621) 65980561; FAX: (8621) 65982273. email: minhuashao@163.com
[2]Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai, P.R. China; email: ljsun@mail.tongji.edu.cn.
[3]Institute of Traffic Engineering, Tongji University, Shanghai, P.R. China; email: cexzhang@mail.tongji.edu.cn.

Abstract: Traditionally, the capacity of an urban street is defined for a single direction of travel as the capacity of the through movement at its lowest point (often at a signalized intersection). The results of analysis in this paper show that this is not reasonable. On one hand, the downstream intersection approach may not be the bottleneck of the whole street due to the possible existence of many affecting factors including signal density and it's hard to judge the real bottleneck in practice. On the other hand, the capacity of a bottleneck cannot represent the street capacity effectively, and even might be a higher estimation under some circumstances. So the concept of Equivalent Capacity is proposed here, which can treat segment and its downstream intersection as an integrated system by introducing the travel speed as a variable. The primary evidence of the advantage of this new concept is given by a simple survey. Finally, the generalized definitions of flow, density and speed are used to explain the theoretic implication of equivalent capacity.

Key words: *streets, capacity, traffic signals, intersections, traffic surveys, traffic flow, traffic speed*

1 Introduction

Road capacity is one of the fundamental parameters in the study of traffic flow theory and is also important in traffic engineering for planning, design, operation analysis and management. In this paper, we focus on the vehicle capacity, which is defined as the maximum number of vehicles that can pass a given point during a specified period under prevailing roadway, traffic and control conditions at the point level. At the facility level, the capacity of a highway or a freeway is defined as the capacity of its one uniform section and the capacity of an urban street is defined for a single direction of travel as the capacity of the through movement at its lowest point (usually at a signalized intersection)(TRB 2000).

Although the highway capacity and the urban street capacity were investigated separately in almost all the literatures to reflect the difference between uninterrupted flow and interrupted flow, the underlying methods were the same in reality: a typical section was selected to represent the whole facility at first, then some methods were applied to obtain the capacity of the section based on the theoretical models.

Theoretical research on section capacity initiated from the 1920s with emphasis on estimating the capacity of one lane based on car-following behavior of individual drivers. The 1st edition of Highway Capacity Manual (HCM) contains a synopsis of 23 early studies of highway capacity reported between 1924 and 1941(Gerlough and Huber 1975). Almost all the studies were based on equation 1:

$$Cap = 5280V / S_p \qquad (1)$$

Where Cap is the capacity of a single lane (vehicles/hr), V is the velocity (miles/hr), and S_p is the average spacing (ft) from front bumper to front bumper of moving vehicles, which can be calculated by $S_p = aV^2 + bV + c$, where a, b, c are all constants.

Later, studies on the capacity move to the relationship among volume, density and speed for the development of traffic flow theory since 1930s. Capacity can be obtained by calculating the maximum value of traffic volume through its relationship with density or speed. These relationship functions can be summarized as below (Gerlough and Huber 1975; May 1991):

Greenberg（1959）：$u = u_f(1 - k/k_j)$ (2)

Greenberg（1959）：$u = u_m \ln(k/k_j)$ (3)

Underwood（1961）：$u = u_f \exp(-k/k_m)$ (4)

Drake（Northwestern）（1965）：$u = u_f \exp[-0.5(k/k_m)^2]$ (5)

Where q is the flow, u is the speed, u_m is the speed at maximum flow, u_f is the free-flow speed (FFS), k is the density, k_m is the density at maximum flow and k_j is the jam density.

Also some multi-regime models were developed by Edie (1961), Underwood (1961) Dick (1966) (Gerlough and Huber 1975) et al to avoid the deficiency of one-regime models in reflecting the traffic condition of high concentration or low

concentration.

All these models were built on the regression of field data. Another category of models were based on the assumption of driving behavior. Grazis et al published the first paper about the relationship between car following model and traffic flow models in 1959. Since then, more research on this subject was undertaken. The results showed that many different traffic flow models could be derived directly from the car following model by changing some parameters (May and Keller 1967).

Based on the above theoretical research, the applied research on point capacity was performed. In practice, the methods used to calculate the capacity can be simply divided into two categories: discount parameter method and field survey method. The former was stated in HCM in detail. It can be used in both highway section capacity and the intersection approach capacity analysis. Only in the latter, the green to cycle ratio (g/C) should be included as an important discount parameter. The field survey method usually obtains the capacity by surveying the time headway. For intersection approach, this method can be applied at the stop line or the conflict point of the intersection during the green signal period in the same way as for highway study (Guo 1994). But the g/C must be considered as a discount. Then, according to the definition stated in HCM, these point capacity is chosen to represent the facility capacity.

So, it can be concluded that the essential difference between highway capacity and urban street capacity in current research is the choice of typical section. A uniform section of segment was chosen in highway analysis and usually the section of intersection approach was chosen in urban street. But unlike the uniform performance of sections along the highway, many disturbing factors exist in urban streets, such as the signals of intersections, the bicycles on the street, the business activities along the roadside, and so on. So, the performance of sections along the street is not consistent. Can the capacity of the movement at the lowest point (usually one signal intersection on it) represent the capacity of the whole street? This paper attempts to answer this question.

The remainder of this paper is organized as below. Firstly, problems existing in urban street capacity study are analyzed. Then the new concept of equivalent capacity is proposed from the perspective of street system. Next, the result of a simple survey is given to prove the advantage of the new concept in real life. Finally, its theoretical explanation is presented and the conclusions are provided.

2 Problems Analysis

According to the analysis in the previous section, two hypotheses exist in the capacity study of urban streets:

First, the intersection approach is usually viewed as the lowest point of capacity along the urban street.

Second, the capacity of the lowest point along the street can represent the capacity of the whole street.

The two hypotheses will be discussed in the following.

2.1 Analysis on the first hypothesis

The first hypothesis may not be true under some circumstances. This was also mentioned in HCM: sometimes, there are special disturbing factors at some points between signals that could restrict the capacity of the urban street (Ren 1999). However, from the following analysis, we found that besides these factors, signal density could be another important factor that also affects the location of the bottleneck.

In this study, VISSIM is used to simulate the operational condition of a two-lane street with a downstream bottleneck, where the street changes from two lanes to one. The street is separated into several analysis segments with 100 meters in length. We run the simulation under different given demands and calculate the capacity and FFS of every segment by analyzing the relationship among flow, density and speed with Greenshields model. Figure 1 shows the result.

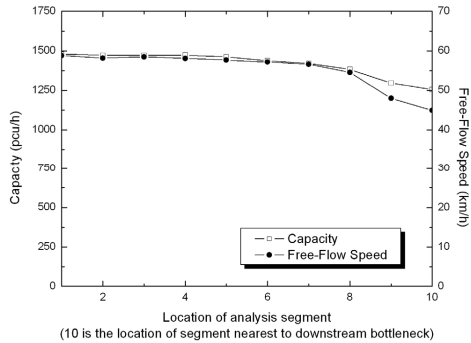


Figure 1. Effect of downstream bottleneck on the capacity and FFS of analysis segment in the simulation.

In figure 1, both the capacity and the FFS of analysis segment decline with its location closer to the bottleneck and the two curves almost overlap. In this process, the jam density of segment doesn't change with its location. So we can conclude that FFS dominates the capacity and other factors affect the capacity through affecting the FFS first.

Considering whether the signal density will affect the FFS or not, a FFS survey was carried out on Zhoujiazui Road and Hutai Road of Shanghai, with intersection spacing varying from 53 meters to 1124 meters. The survey was done in early morning between 0:00 and 4:00 AM, adopting the method described in HCM. The result was shown in figure2. Therefore, we can see that FFS do decrease with the intersection spacing getting smaller.

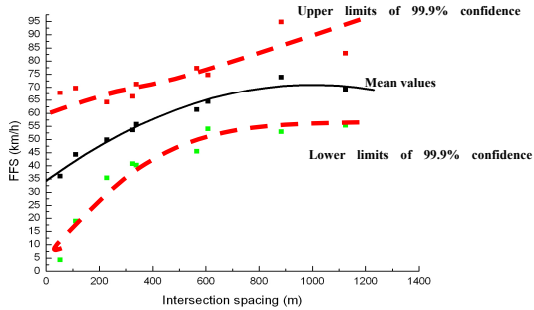


Figure 2. FFS changing with intersection spacing according to the survey results.

Based on the above analysis, the capacity of street unit, which is defined as the segment plus its downstream signal intersection approach, is analyzed in the following. As shown in figure 3, the intersection could reduce the drivers' expected speed on the segment and hence affect the capacity of the segment.

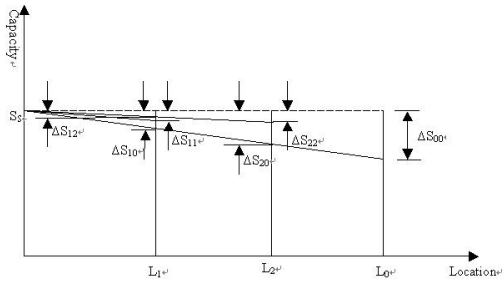


Figure 3. Schematic map of capacity discount of sections along the street.

In figure 3, the approach at intersection is supposed to be not broadened. S_s is the capacity of the street under ideal conditions (without any disturbance factors such as signal intersection, bicycles, roadside activities and so on). L_0 is the location of the downstream intersection, L_1 and L_2 are two locations on the segment where disturbance (like bicycles, roadside activities et al) exists. From the analysis in figure 1, we can conclude that any disturbance factor will result in the capacity drop not only right at the section but also upstream the segment as shown in figure 3. We define the capacity drop of section L_i caused by the disturbing at location L_j as ΔS_{ij} and the

whole capacity drop from ideal capacity at location L_i as ΔS_i. It is easy to know that $\Delta S_i = \sum_j \Delta S_{ij}$. So, in figure 3, we have:

$$\Delta S_1 = \Delta S_{10} + \Delta S_{11} + \Delta S_{12} \quad (6)$$

$$\Delta S_2 = \Delta S_{20} + \Delta S_{22} \quad (7)$$

$$\Delta S_0 = \Delta S_{00} \quad (8)$$

Set g/C as R_g. We can see that, when $\Delta S_i > S_s(1 - R_g)$, the intersection approach is not the bottleneck anymore. This phenomenon happens more frequently when the intersection approach is broadened, since in this case the saturation flow rate of the approach will be greater than the capacity of the segment. However, it's almost impossible to examine the capacity of every section along the street in real situations. So, it is very hard to determine the exact bottleneck of the urban street.

2.2 Analysis on the second hypothesis

Furthermore, even if the bottleneck can be determined correctly, using its capacity to represent the capacity of the whole street is inappropriate. And in some cases, the result is even wrong. In the following, the capacity of a single-lane street unit is analyzed to support this view. The signalized intersection approach is still assumed to be the bottleneck.

Notations

Variables used in this part are defined as below: S is the saturation flow rate of the intersection approach; g is the effective green time for the lane group (also for the approach because there is only one lane); r is the effective red time; C is the signal cycle time, $C = g + r$; c is the capacity of the approach, $c = S \times g / C$; S' is the capacity of section on segment; L is the length of the segment; L_0 is the space headway between queuing vehicles; N_0 is the maximum number of vehicles that can be accommodated in the street unit, $N_0 = L / L_0$; λ is the arrival flow rate; q_0 is the number of vehicles that arrive during red time; Δq is the increment of queuing vehicles at the end of green time with respect to the same time in previous cycle; T_0 is the duration of analysis period, which is recorded from the beginning of the green time of a given cycle without queuing vehicles that are delayed from the previous green time, supposing $T_0 = n_0 \times C$; n_0 is a positive integer, $n_0 \geq 1$; $q_{\max}$ is the

maximum number of queuing vehicles during T_0; Q_{max} is the maximum flow rate of the street unit without considering T_0.

Analysis

First of all, the street units composed by intersection approaches with the same capacity and road segments with different length have different adapting capacity to traffic flow in T_0.

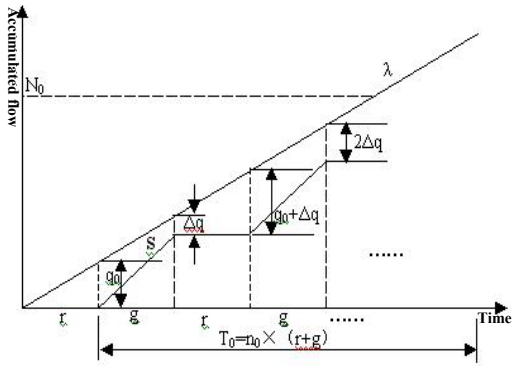


Figure 4. Schematic map of queuing length changing with time.

As shown in figure 4, traffic flow is interrupted by shifts of signals and vehicles arriving during red time wait in the queue. Higher arrival rate will cause some vehicles to queue for the second time and the queue length will increase with time. So, obviously street units with different segment length but same intersection approach can accommodate different queue length during T_0. The capacity of the intersection cannot reflect this point. In order to explain this difference, the concept of maximum adapting flow rate is proposed below.

The maximum adapting flow rate of the urban street unit is defined as the maximum allowable uniform arrival rate, under which queue never occurs on p ($p \in [0,100]$) percent of all sections of the unit during T_0.

Equation 9 can be easily obtained from figure 4:

$$\begin{aligned} \Delta q &= \max(0, \lambda C - Sg) \\ q_{\max} &= \min(N_0, \lambda r + n_0 \Delta q) \end{aligned} \tag{9}$$

According to the concept of maximum adapting flow rate, λ should satisfy:

$$\lambda r + n_0 \Delta q \le N_0 \times (1 - p\%) \tag{10}$$

By some simple deductions, maximum adapting flow rate for different N_0 is expressed with equation 11:

$$\lambda_{\max} = f(N_0, p, n_0, S, r, g) = \begin{cases} \left[N_0 \times (1-p\%) + n_0 Sg\right]/(r + n_0 C) & N_0 \ge cr/(1-p\%)(1) \\ N_0 \times (1-p\%)/r & N_0 < cr/(1-p\%)(2) \end{cases} \tag{11}$$

With the same conditions of the intersection and the same T_0, $\lambda_{\max}$ can be expressed as $\lambda_{\max} = f(N_0, p)$. When condition (1) of equation 11 is satisfied, we have $\lambda_{\max} \ge c$. If p >50, then in more than half sections queue never occurs. Obviously it is inappropriate to choose the capacity of the intersection approach to represent the capacity of the whole unit.

Secondly, even T_0 is not considered as above, it could still be wrong to represent the capacity of the street unit with its intersection approach capacity under some circumstances. The main reason is that the street unit is a system of segment and intersection approach. The two components cannot be considered separately. As discussed in the previous part of this paper, the signal intersection could make the capacity of the section on its upstream segment decline. Similarly, sometimes the capacity of the intersection approach is limited by the length of segment and can't be fully used. In this situation, the capacity of the street unit will be overestimated by the capacity of the intersection approach.

Since the intersection approach is a bottleneck, we have $S' > c$. Also, as discussed in the previous part, considering many disturbing factors exist on the segment, such as lane changing, roadside activity and so on, $S' \le S$ can be true in most situations. Therefore, we assume that $S' \in (c, S]$.

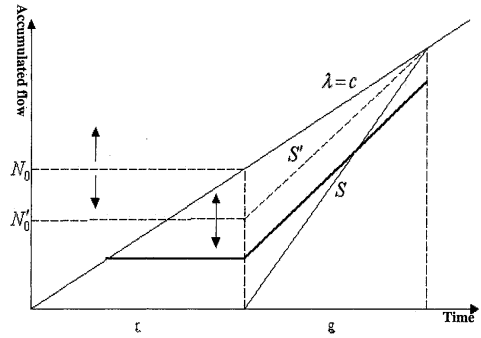


Figure 5. Accumulated flow over time at intersection approach.

As shown in figure 5, N_0' is the number of accommodated vehicles under critical segment length, $N_0' = (S - S')g$. For $\lambda \le S'$, $Q_{\max}$ can be simply calculated with equation 12:

$$Q_{\max} = \begin{cases} c & N_0 \ge (S - S')g(1) \\ (N_0 + S'g)/C & N_0 < (S - S')g(2) \end{cases} \quad S' \in (c, S] \tag{12}$$

When condition (2) of equation 12 is satisfied, the capatity of street unit is limited by N_0 , which is smaller than the capacity of the intersection approach.

3 Concept of Equivalent Capacity

From the above analysis, we reach the following conclusions. On one hand, the downstream intersection approach may not be the bottleneck of the whole street unit due to the existence of many disturbing factors including signal density, and it is hard to judge the real bottleneck in practice. On the other hand, the capacity of the bottleneck cannot represent the street capacity effectively, and it might even be a higher estimation in some cases.

The main reason of these problems is that the methods that people used to analyze urban street capacity are the same as in the analysis of highway capacity. Representing capacity of a whole road with the capacity of some section is not applicable for urban streets although it is applicable for highways. To resolve this deficiency to some extent, the concept of equivalent capacity is proposed in the following.

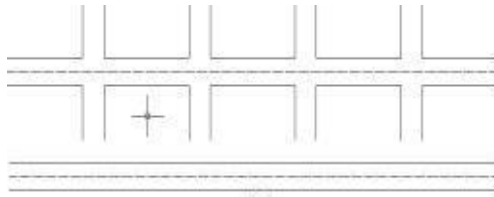


Figure 6. Schematic map of the concept of equivalent capacity

As shown in figure 6, we define equivalent capacity as follow. A real urban street composed by several segments and intersections is simplified to an equivalent segment with no intersections. Then the average travel speed of the real street is taken to be the running speed of the equivalent segment and the average flow of all segments of the real street is taken to be the flow of section of the equivalent segment. Moreover, some traffic flow model concerning the relationship among flow, speed and density is constructed. Thus the capacity of the equivalent segment derived from the model is defined as the equivalent capacity of the real urban street.

In the same way, the FFS from the model is defined as the equivalent FFS and the jam density is defined as the equivalent jam density. Accordingly, the travel speed is defined as the equivalent speed. The average flow is defined as the equivalent flow, and the ratio of equivalent flow to equivalent speed is defined as the equivalent density.

The following notations are given to represent these variables in the paper. q_e is the equivalent flow, C_e is the equivalent capacity, V_e is the equivalent speed, V_{ef} is the equivalent FFS, k_e is the equivalent density, and k_{ej} is the equivalent jam density.

Different from the traditional concept of road capacity, equivalent capacity does not refer to one section of the street. Travel speed is introduced into this definition to reflect all the disturbance factors of the street including bottlenecks on it. So, the urban street can be regarded as an organic system in the definition of equivalent capacity.

Because the prevailing traffic simulation software cannot consider the effect of the intersection spacing to the FFS, field survey method is proposed here. According to the definition, the process to get equivalent capacity can be summarized in figure 7.

After choosing appropriate urban street, we arrange some test cars on it. Exact time of entering and leaving street is recorded in seconds. At the same time, some investigators standing along roadside are organized to count the bi-directional flows of some sections (the mid section is suggested in this paper) on every segment of the street. Then the relationship between average volume and average travel speed during a given analysis period are obtained by regression method. The equivalent capacity can be obtained from the model by taking the maximal available traffic volume.

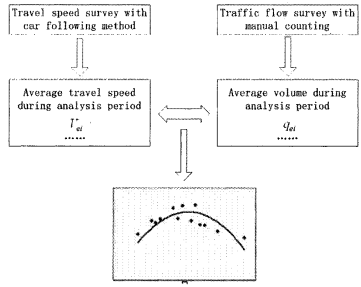
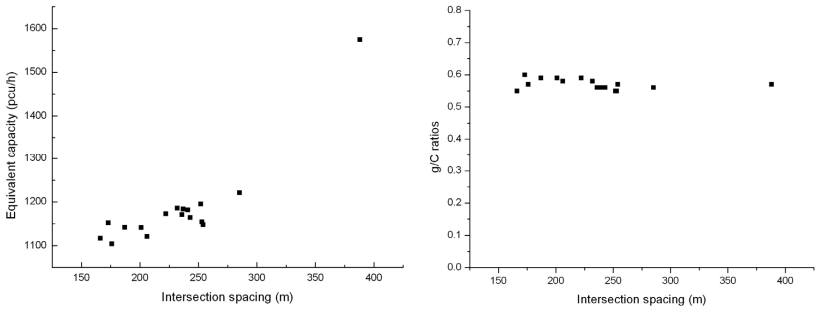


Figure 7. Process depictions of equivalent capacity survey and analysis.

4 Primary Evidence of Advantage

As we know, by introducing travel time into the concept of equivalent capacity, the urban street can be viewed from a systematic perspective. The result of Beijing Road of Shanghai gives a primary proof for this point. 22 street units are assembled into several experimental streets with different average intersection space and equivalent capacity of every street is calculated and shown in figure 8 a. Because other parameters of these streets are similar, including the g/C ratios (seeing in figure 8 b), the difference of intersection space is the main reason for the difference of equivalent capacity.



a) Equivalent capacity changing b) The corresponding g/C ratios

Figure 8. Equivalent capacity changing with intersection spacing.

So, in figure 8 a, the equivalent capacity increases with intersection distance. This tendency is consistent with the result discussed in section 2 of this paper.

5 Theoretic Explanations about Equivalent Capacity

Equivalent capacity can be explained by generalized definitions of traffic flow, density and speed.

Equations 13~15 give these generalized definitions (Daganzo 1997).

A time-space diagram including the trajectories of several vehicles is used to give these definitions. dA is the area of a close extent on this diagram mapped out by a demon observer.

$$q = \frac{\text{total travel distance}(veh\text{-}km)}{\text{total area}(km\text{-}hr)} = \frac{\sum_i dx_i}{dA} \tag{13}$$

$$k = \frac{\text{total travel time}(veh\text{-}hr)}{\text{total area}\ (km\text{-}hr)} = \frac{\sum_i dt_i}{dA} \tag{14}$$

$$V = \frac{\text{total travel distance}(veh\text{-}km)}{\text{total travel time}(veh\text{-}hr)} = \frac{\sum_i dx_i}{\sum_i dt_i} \tag{15}$$

It is easy to know that the relationship of $q = kV$ is still applicable for urban streets under these definitions. Based on this, q_e, V_e and k_e are derived from the previous definitions and explained by the generalized definitions. For analyzing, a rectangle with area $\boldsymbol{T \times L}$ is mapped out on the time-space diagram, where $\boldsymbol{T}$ is length of the rectangle along axis $\boldsymbol{t}$ and $\boldsymbol{L}$ is width of the rectangle along axis $\boldsymbol{X}$. According to the concept and survey process of equivalent capacity, equations 16~20 have been derived as follow.

$$q_e = \frac{\sum_j q_j l_j}{\sum_j l_j} \Big/ T = \sum_j q_j l_j \Big/ LT = \frac{\text{total travel distance}}{\text{total area}} \tag{16}$$

Where q_j is the volume of mid section of the jth street segment during T, l_j is length of the jth segment, L is length of experimental street, and $\boldsymbol{T}$ is duration of analysis period.

$$V_e = \frac{L}{\sum_i t_i / n} = \frac{nL}{\sum_i t_i} \tag{17}$$

Where t_i is the travel time of the ith test car of n cars that traverse the experimental street during T.

Suppose that V_e can represent the travel time of other vehicles traveling on the street partly during $\boldsymbol{T}$. We have equation 18 as follow,

$$V_e = \sum_k l_k \Big/ \sum_k t_k \tag{18}$$

Where l_k is the travel distance of the kth vehicle of those that travel on the street partly and t_k is the travel time on the experimental street of the kth vehicle during $\boldsymbol{T}$.

With equations 17 and 18, hence

$$V_e = \frac{nL + \sum_k l_k}{\sum_i t_i + \sum_k t_k} = \frac{\text{total travel distance}}{\text{total travel time}} \tag{19}$$

Therefore we can see that both q_e and V_e could be explained with the generalized definitions. As k_e is defined as q_e / V_e, it could also be explained with equation 20.

$$k_e = \frac{q_e}{V_e} = \frac{\text{total travel time}}{\text{total area}} \tag{20}$$

Based on above analysis, supposing the urban street is regarded as a time-space container, and then the equivalent capacity can be viewed as the maximum vehicle-kilometers of unit volume of the container.

6 Concluding Remarks

Traditional research on road capacity is reviewed firstly in this paper. Using the capacity of one section as the capacity of the whole urban street is not reasonable, although the same method is used for highway capacity successfully. On one hand, the downstream intersection approach may not be the bottleneck of the whole street unit for the existence of many disturbing factors including signal density, and it is hard to judge the real bottleneck in real cases. On the other hand, the capacity of bottleneck cannot represent the street capacity effectively, and even might be a higher estimation under some circumstances.

Then the concept of equivalent capacity is proposed based on the experimental traffic engineering. By introducing travel speed into its concept, equivalent capacity can regard an urban street as an integrated system, which has been testified by the survey result of Beijing Road of Shanghai. With the generalized definitions of traffic flow, density and speed, the equivalent capacity can be viewed as the maximum vehicle-kilometers of unit time-space volume of urban street.

The recommendation values of equivalent capacity under different conditions can be given after abundant work of survey and theoretic analysis, and they can be used in practice to consider more influence factors in design and planning. Also, the concept is the bridge between point capacity and network capacity. Therefore, it could be the basis of the research on road network capacity.

References

Carlos F. Daganzo (1997). *Fundamentals of transportation and traffic operations*. Amsterdam; New York: Pergamon.

Dick, A. C (1966). "Speed/flow relationships within an urban area." *Traffic Eng. Control*, Vol.8, NO.6, 393-396.

Edie, L. C (1961). "Car-following and steady-state theory for noncongested traffic." *Oper. Research*, Vol.9, No.1, 66-76.

Gazis, D. C., Herman, R., and Potts, R. B (1959). "Car-following theory of steady-state traffic flow." *Oper. Research*, Vol.7, NO.4, 499-505.

Gerlough, D. L., and M. J. Huber (1975). *Traffic Flow Theory: A Monograph*. Special Report 165, TRB, National Research Council, Washington, D.C..

Guo guanying (1994). *Traffic Engineering*. Publishing company of Tongji University, Shanghai, China.

May, A. D., Jr., and Keller, H. E. M (1967). "Non-integer car-following models." *Highw. Res*. Rec. 199, 19-32.

May, A. D (1991). *Traffic Flow Fundamentals*. Prentice-Hall,Inc.. Englewood Cliffs, N.J..

Ren futian (translate)(1999). *Highway Capacity Manual*. Architecture industry publishing company of China, Beijing, China.

Transportation Research Board (2000). *Highway Capacity Manual*. Special Report 209, TRB, National Research Council, Washington, D.C..

Underwood, R. T (1961). "Speed, volume and density relationships." *In Quality and Theory of Traffic Flow*. Bureau of Highway Traffic, Yale University, New Haven, Conn, 141-187.

A NEW BRAESS'S PARADOX WITH QUEUE SPILLOVERS CONSIDERING DYNAMIC USER EQUILIBRIUM

Xiaoning Zhang[1] and H. M. Zhang[2, 1]

[1]Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 200092, P.R. China ; PH (86) 21-65982443; FAX (86) 21-65984583; email: xn.zhang@alumni.ust.hk

[2] Department of Civil and Environment Engineering, University of California, 1 Shields Avenue, Davis, California 95616, USA; PH (530) 754-9203; FAX (530) 752-7872; email: hmzhang@ucdavis.edu

Abstract: This paper explores the network properties when queues spill over network links, by examining a paradox of network expansion. In networks, each commuter simultaneously chooses departure time and travel route to minimize the travel cost including journey time and unpunctuality penalty. Using some necessary conditions of dynamic user equilibrium, the dynamic network flows are exactly obtained on some small networks. Then we use a network with two origins and two destinations to illustrate an interesting paradox that is caused by the FIFO behavior among commuters traveling on a common upstream link of which the exit is blocked by one branch of downstream traffic. The comparison analysis in this example shows that point-queue models often overestimate the network performance and consideration of queuing spatial dimension in dynamic traffic assignment is necessary.

Key words: *Braess's paradox, simultaneous departure time and route (SDR) choices, dynamic traffic assignment (DTA), bottleneck, queue spillover*

1 Introduction

Since Vickrey (1969), the point-queue or bottleneck models have been used in dynamic traffic assignment (DTA). The point-queue models should also be the most solvable DTA models, by simplifying each network link as a free-flow component plus a downstream bottleneck with limited service capacity. In the bottleneck models, it is assumed that vehicles do not take up space, and a queue stacks up vertically at the entrance of the bottleneck. Therefore, only the temporal dimension of traffic is considered, and the spatial dimension in network links is simply ignored. But in reality, vehicles occupy physical space and a network link is limited in spatial storage, thus traffic queues can spill backward and block junctions. Actually, in many metropolitans, junctions are often blocked by downstream traffic, when transportation networks are over congested in rush hours. Therefore, queuing is not an isolated phenomenon

contained on one street, but has strong spatial interactions with neighboring streets. This characteristic has important implications in modeling the traffic movement in transportation networks.

Some recent DTA models, sometimes referred to as physical-queue models, consider both the temporal and spatial dimensions of traffic flow, capturing queue formation and dissipation, queue spillback, and junction blockage. But in these models, computer simulation or numerical methods are generally used, thus analytical and exact solutions are difficult to obtain.

Daganzo (1996) and Daganzo(1998) explored some very interesting and meaningful phenomena with physical queues, using explicit and exact solution methods. Daganzo (1996) presented a continuum model of merge operations, and showed a gridlocked traffic stream without any motion in an improperly designed closed-loop freeway. Daganzo (1998) introduced several interesting paradoxes in networks with physical queues, where adding new ramps or expanding the capacity of existing bottlenecks will result in the abandonment of alternative routes due to the queue spillovers.

Actually, besides Daganzo (1996) and Daganzo(1998), there are some other interesting traffic phenomena with queue spillovers which have not been revealed before. The objective of this paper is to further investigate the distinct network properties when queues spill over in some links, by explicitly considering dynamic user response. We still use the framework of bottleneck models to characterize the performance of transportation networks, but differently we assume that queues take up space and network links can only hold limited vehicles. A dynamic version of Braess's paradox (Braess, 1968) will be illustrated through a small network.

2. Network models and a necessary condition of dynamic user equilibrium

2.1 Network models

A network is composed of a finite set of nodes and a finite set of directed links connecting these nodes. Traffic flows are generated from some origins, and are absorbed to some destinations, where the origins and destinations are subset of the nodes. Each OD pair rs is connected by a finite set of routes (paths), which are acyclic ordered set of links. A link is made up of two parts, namely the main body with infinite capacity and constant travel time h_a^0, and a bottleneck with zero-length and time-dependent finite capacity $c_a(t)$. Moreover, since vehicles occupy physical spaces, each link can hold only l_a vehicles.

It is assumed that the arrival flow leaves the link after the free-flow travel time h_a^0 if there is no queue in the link, otherwise it leaves the link at the bottleneck capacity $c_a(t)$. Therefore, the queues in the links evolve in the following manner:

$$\frac{dx_a(t)}{dt}=\begin{cases}0, & \text{if } x_a(t)=0 \text{ and } i\left(t-h_0^a\right)<c_a(t);\\ i\left(t-h_0^a\right)-c_a(t), & \text{otherwise,}\end{cases} \tag{1}$$

where $x_a(t)$ is the queue length at the entrance of the bottleneck at time t, and $i(t)$ is the instantaneous inflow rate to link a.

The First-In-First-Out (FIFO) principle is respected on each link, which means no overtaking is allowed among drivers. Now we employ the FIFO principle to derive the outflow rates from a link for various paths that pass through a common link, given the path inflow rates to this link. Zhang and Zhang (2007) showed that no matter whether the bottleneck is operating in full capacity or not, the path outflow rate $o_k(t')$ is always proportional to the path inflow rate $i_k(t)$, where t and t' are time instants for the same traffic enters and leaves the link. This condition is used in the example to determine the downstream outflow rates at diverge and merge intersections.

Figure 1 represents a diverge intersection with an upstream link (link 1) and two downstream links (link 2 and link 3). Here we suppose the service capacities and path inflows are time-invariant. The service capacities of the three bottlenecks are c_1, c_2, and c_3 respectively. Let the inflow rate of path ABU be i_u, and that of path ABD be i_d. If $c_1<c_2$ and $c_1<c_3$, there will be no queue on link 2 or link 3 if both links are originally empty, no matter how large is the inflow rate of link 1. Even if queues already exist on the links, they will diminish gradually. In this case, the outflow from link 1 is either in full capacity or equal to the inflow of link 1, depending on whether the inflow rate exceeds the bottleneck service capacity or not.

However, if $c_1>c_2$ or $c_1>c_3$, it is possible that a queue will develop on either link or both. Now we suppose $c_1>c_2$ and $i_2>c_2$, where i_2 is the inflow rate of link 2, so a queue is growing upstream link 2 from the end of link 2 (node U). If this traffic state remains, after a while the queue will reach the junction node B and block it. If originally there is no queue on link 1 (link 1's inflow rate $i_u + i_d$ £ c_1, and $i_2 = i_u, i_3 = i_d$), the queue on link 2 will pass over node B and keep growing along link 1. Whereas if there is already a queue on link 1, or $i_u + i_d > c_1$ and $i_2 + i_3 = c_1$, the tail of the queue on link 2 connects the head of the queue on link 1. The queues on link 1 and link 2 are depicted by the bold lines in Figure 1. What we are emphasizing is that the outflow from link 1 and the inflows to both link 2 and link 3 immediately decrease, after node B is blocked by the queue on link 2. Obviously inflow rate of link 2 decreases from the original value i_2 to the current value c_2. In line with FIFO principle, vehicles cannot overtake each other on a link, therefore at the exit of link 1 the slowdown of traffic outflow of path ABU draws down the outflow of path ABD.

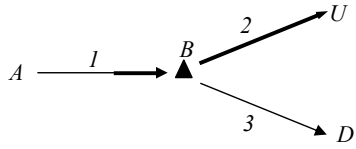


Figure 1. A diverge intersection

Then we have $\frac{i_u}{i_d}=\frac{i_2}{i_3}$ and $\frac{i_u}{i_d}=\frac{c_2}{i_3'}$ corresponding to the traffic conditions before and after the blocking of node B. This implies that the inflow of link 3 reduces to $i_3'=i_3\frac{c_2}{i_2}$. Thus the total outflow of link 1 reduces from i_2+i_3 to $\frac{c_2}{i_2}(i_2+i_3)$, which means bottleneck 1 is operating under service capacity. If $i_u+i_d\le c_1$, or there is no queue on link 1 before node B is blocked, a queue is developing on link 1 in a rate of $\frac{(i_u+i_d)(i_2-c_2)}{i_2}$. On the contrary, if $i_u+i_d>c_1$, the developing rate of the queue that is already on link 1 increases from $i_u+i_d-c_1$ to $i_u+i_d-\frac{c_1c_2}{i_2}$. The bad news of the blocking of node B is that the throughput of link 3 (or path ABD) drops down when $i_3\frac{c_2}{i_2}<c_3$, although the throughput of link 2 (or path ABU) remains unchanged.

Figure 2 represents a merge intersection with two upstream links and a downstream link. The inflow rate of path UBA is i_u, and that of path DBA is i_d. Let the service capacities of the three bottlenecks be c_1, c_2, and c_3 respectively. All links are empty before traffic comes from node U and node D. If $i_u>c_1$, the outflow of link 1 $o_1=c_1$; otherwise, $o_1=i_u$. Similarly, if $i_d>c_2$, the outflow of link 2 $o_2=c_2$; otherwise, $o_2=i_d$. Therefore the traffic out of link 1 and link 2 enters link 3 immediately, and the inflow to link 3 is $i_3=o_1+o_2$. If $o_1+o_2\le c_3$, there is no queue on link 3, and the outflows at node A are o_1 and o_2 for paths UBA and DBA respectively. On the contrary, if $o_1+o_2>c_3$, the bottleneck of link 3 is operating in full capacity and a queue is developing on link 3. In this case, the outflows at node A are $o_1c_3/(o_1+o_2)$ and $o_2c_3/(o_1+o_2)$ for paths UBA and DBA respectively, according to FIFO. After a while, the queue reaches node B and block the intersection. Since there is no more space to hold excessive traffic on link 3, the queue keeps climbing upstream along both link 1 and link 2, or at least either one. Now the inflow of link 3 is no longer the original o_1+o_2, instead it is $i_3'=c_3$. Therefore, the

total outflow of link 1 and link 2 also decreases from o_1+o_2 to c_3. It is clear that the sum of outflows from link 1 and link 2 is c_3, i.e. $o_1'+o_2'=c_3$, but the proportions of o_1' and o_2' are still unknown. Up to now, there is no widely accepted rule to allocate o_1' and o_2', and this is still an open question. As mentioned in Daganzo (1996), a possible method of determining o_1' and o_2' is based on the numbers of lanes of the upstream links.

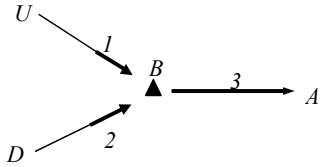


Figure 2. A merge intersection

2.2 Dynamic user equilibrium and necessary conditions

In networks, each commuter simultaneously chooses departure time and travel route to minimize the travel cost including journey time and unpunctuality penalty. A nonnegative schedule delay cost function $\phi^{rs}\left(\tau+T^{rs}(\tau)\right)$ represents the penalty incurred if one leaves the origin r at time τ and arrives at the destination s at time $\tau+T^{rs}(\tau)$. It is further assumed that his/her generalized unit travel cost is

$$C^{rs}(\tau)=\alpha T^{rs}(\tau)+\phi^{rs}\left(\tau+T^{rs}(\tau)\right), \tag{2}$$

where α is the cost per unit travel time, which is a positive value.

We assume that $\phi^{rs}(t)$ is a convex function of clock time t, with a minimum value $\phi^{rs}\left(t^*\right)=0$ at the official work start time t^*. Schedule delay penalty $\phi^{rs}(t)$ increases when t moves away from t^* along the two sides of time axis A typical example of the demand model is the piecewise linear schedule delay function, given as below

$$\phi^{rs}\left(\tau+T^{rs}(\tau)\right)=\begin{cases}\beta\left[\left(t^{*rs}-\Delta^{rs}\right)-\left(\tau+T^{rs}(\tau)\right)\right], & \text{if } \tau+T^{rs}(\tau)<t^{*rs}-\Delta^{rs};\\ 0, & \text{if } t^{*rs}+\Delta^{rs}\geq\tau+T^{rs}(\tau)\geq t^{*rs}-\Delta^{rs};\\ \gamma\left[\left(\tau+T^{rs}(\tau)\right)-\left(t^{*rs}+\Delta^{rs}\right)\right], & \text{if } \tau+T^{rs}(\tau)>t^{*rs}+\Delta^{rs}.\end{cases}$$

(3)

where α is the unit cost of travel time, β the unit cost of early arrival unpunctuality penalty, γ is the unit cost of later arrival unpunctuality penalty, and t^{*rs} is the official work start time for OD pair rs. Interval $\left(t^{*rs}-\Delta^{rs},\ t^{*rs}+\Delta^{rs}\right)$ is the on-time arrival window in which schedule delay penalty is exempted for trips from origin r

to destination s. In accordance with the empirical results (Small, 1982), $\gamma > \alpha > \beta$ holds.

According to the demand equilibrium theory, the departure time choice model is given as below:

$$\begin{cases} i^{rs}(\tau) > 0 & \Rightarrow \quad C^{rs}(\tau) = C^{rs}_{\min} \\ C^{rs}(\tau) > C^{rs}_{\min} & \Rightarrow \quad i^{rs}(\tau) = 0 \end{cases} \forall rs \ \forall \tau, \tag{4}$$

where $i^{rs}(\tau)$ is the departure rate of OD rs at time instant τ, $C^{rs}(\tau)$ is the individual travel cost of OD rs at time instant τ, and $C^{rs}_{\min}$ is the minimal individual travel cost of OD rs among all departure-time instants.

Our route choice model is based on the assumption of dynamic user equilibrium (DUE), which is a dynamic generalization of Wardropian user equilibrium principle. Mathematically the dynamic user equilibrium can be characterized as

$$\begin{cases} i^{rs}_p(\tau) > 0 & \Rightarrow \quad C^{rs}_p(\tau) = C^{rs}(\tau) \\ C^{rs}_p(\tau) > C^{rs}(\tau) & \Rightarrow \quad i^{rs}_p(\tau) = 0 \end{cases} \forall p \in P_{rs} \ \forall rs \ \forall \tau \tag{5}$$

where P_{rs} is the set of all paths from r to s, $C^{rs}_p(\tau)$ is the cost incurred on path p by a commuter who depart the origin at time τ, and $C^{rs}(\tau)$ is the minimum cost traveling from r to s departing at time τ among all alternative paths.

Zhang and Zhang (2006) derived the following necessary condition for the DTA-SDR problem:

$$\frac{i^{rs}_p(\tau)}{o^{rs}_p\left(\tau + T^{rs}_p(\tau)\right)} = \frac{1}{1 + \frac{1}{\alpha}\frac{\mathrm{d}\phi^{rs}\left(\tau + T^{rs}_p(\tau)\right)}{\mathrm{d}\left(\tau + T^{rs}_p(\tau)\right)}}, \forall \tau \in \boldsymbol{\tau}^{rs}_p. \tag{6}$$

where $\boldsymbol{\tau}^{rs}_p = \left\{\tau \middle| i^{rs}_p(\tau) > 0\right\}$, and $T^{rs}_p(\tau)$ is the travel time spent in route p, $p \in P_{rs}$ departing at time τ.

The above equation gives the ratio of path inflow and path outflow, when the same traffic leaves the origin and reaches the destination respectively. Specifically in the piecewise linear schedule delay penalty function (3), the above equation is

$$\frac{i^{rs}_p(\tau)}{o^{rs}_p\left(\tau + T^{rs}_p(\tau)\right)} = \kappa\left(\tau + T^{rs}_p(\tau)\right), \forall \tau \in \boldsymbol{\tau}^{rs}_p \tag{7}$$

where

$$\kappa\left(\tau + T^{rs}_p(\tau)\right) = \begin{cases} \dfrac{\alpha}{\alpha - \beta} & \text{if } \tau + T^{rs}_p(\tau) < \overline{t}^{rs} - \Delta^{rs}; \\ 1 & \text{if } \overline{t}^{rs} + \Delta^{rs} \geq \tau + T^{rs}_p(\tau) \geq \overline{t}^{rs} - \Delta^{rs}; \\ \dfrac{\alpha}{\alpha + \gamma} & \text{if } \tau + T^{rs}_p(\tau) > \overline{t}^{rs} + \Delta^{rs}. \end{cases} \tag{8}$$

Associated with the FIFO principle described in equations, the above expression is used to analytically obtain the DTA-SDR solutions in the following example.

3. A Paradox on a Network with Two Origins and Two destinations

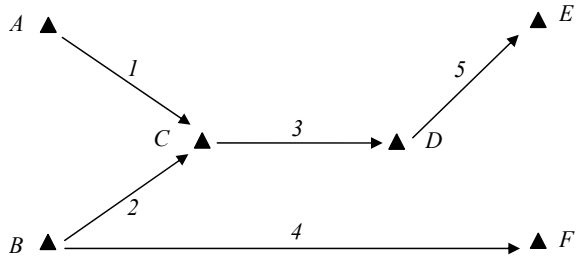


Figure 3. A network with 6 nodes and 5 links

Figure 3 is a network with 6 nodes and 5 links. There are two groups of commuters who use the network for traveling. Group A lives in the residential area represented by node A, and works at node E with start time 8:00am. Group B lives in the residential area represented by node B, and works at node F with start time 8:00am too. The demand of group A is 400 commuters, and the demand of group B is 800 commuters. The unit cost of travel time is α=\$1.0/min, and the unit cost of early arrival time is β=\$0.5/min for both groups. Later arrival is not allowed, and there is no on-time arrival window.

The performance parameters of each link are given below:

constant travel time: h^1=0, h^2=0, h^3=0, h^4=17.5min, h^5=0;

service capacity: c^1=20veh/min, c^2=20veh/min, c^3=40veh/min, c^4=80veh/min, c^5= 20veh/min;

and link storage: $l^1=\infty$, $l^2=\infty$, $l^3=\infty$, $l^4=\infty$, $l^5=\infty$.

Obviously, there is only one path for each group, and cumulative inflows and outflows are shown in Figure 4. Curves EF and EG are the cumulative inflow and outflow for group A, and curves AB and CD are those of group B. Individual travel cost of group A is \$10, and that of group B is \$22.5. Therefore the system travel cost is $\$2.2\times10^4$.

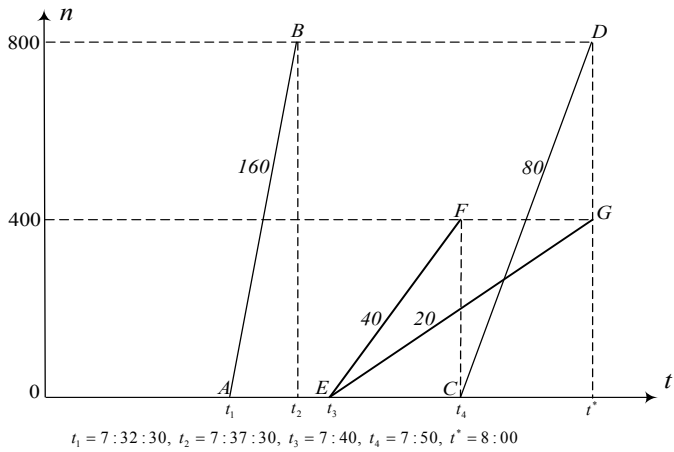


Figure 4. Cumulative inflow and outflow for groups *A* and *B*

Now a link is constructed from node D to node F, with h^6=0, c^6=10 veh/min, and l^6=20veh, as shown in Figure 5. We look at group B first. There are two paths for group B, namely path BF and path $BCDF$. Since the outflow from link 1 is limited by 20veh/min, link 6 can always maintain an outflow of 10veh/min. At the beginning, group B uses path $BCDF$ for traveling, with an inflow of 20veh/min and an outflow of 10veh/min. Hence, a queue is growing on link 6 with a rate of 10veh/min, and 2 minutes later the queue blocks node D. After node D is blocked, the inflow and outflow rates of path $BCDF$ are still 20veh/min and 10veh/min, and the queue keeps growing along link 3 from node D. 17.5 minutes after the first departure on path $BCDF$, path BF starts to be used. The cumulative inflow and outflow of path $BCDF$ are depicted by curves AB and AC in Figure 6, and those of path BF are depicted by curves HB and IC. The number of commuters on both paths is 400, and everyone bears a travel cost of \$20.

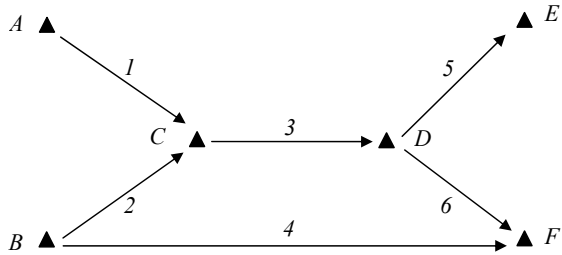


Figure 5. Adding link 6 to the network

Now we look at group A. According to the FIFO principle, if link 6 is not filled up, the outflow of path $ACDE$ is 20veh/min; otherwise if node D is blocked by the vehicles on link 6, the outflow reduces to 10veh/min. In the last two minutes before 8:00am, node D is released from blocking, so the outflow of path $ACDE$ is 20veh/min from 7:58am to 8:00am. And from 7:22am to 7:58am, the outflow of path $ACDE$ is 10veh/min, since node D is blocked during this period. The cumulative inflow and outflow on path $ACDE$ are depicted by curves DEF and DGC in Figure 6 respectively. Clearly the individual travel cost of group A is \$19. After the adding of link 6, the system travel cost becomes $\$2.36\times10^4$, which is increased by $\$1.6\times10^3$.

If we employ the point-queue model, path BF and path $BCDF$ are used for group B in the same way as applying the physical-queue model, so the individual travel cost of group B is still \$20. Differently, group A always has an outflow of 20veh/min, since node D can never be blocked by the downstream queues. Therefore, as in the case before adding link 6, individual travel cost of group A is \$10. So with the point-queue model, the system travel cost reduces to $\$2.00\times10^4$.

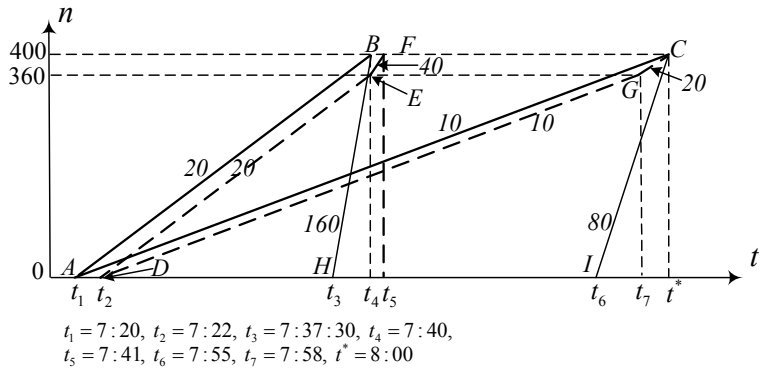


Figure 6. Queuing diagrams for the three paths

5. Conclusion

In this paper, we explored the network properties when queues spill over network links, and provided a paradox example of network expansion considering simultaneous departure and route choices. In networks, each commuter aims to minimize the travel cost including journey time and unpunctuality penalty, by simultaneously choosing departure time and travel route. The time dependent network flows can be explicitly obtained by examining some necessary conditions of dynamic user equilibrium. Then we used a small network to illustrate the dynamic Braess's paradox, which is caused by queue spillovers.

Acknowledgement

This study has been substantially supported by three grants from the National Natural Science Foundation Council of China (Project No. 70401016, 70571058 and 70631002).

References

Braess, D. (1968) Uber ein paradoxen der verkehrsplaning. *Unternehmensforschung*, 12, 258-268.

Daganzo, C.F. (1996) The nature of freeway gridlock and how to prevent it, in *Transportation and Traffic Theory,* (J.B. Lesort, editor), Pergamon-Elsevier, Tarrytown, N.Y., 629-646.

Daganzo, C. (1998) Queue spillovers in transportation networks with a route choice. *Transportation Science*, 32, 3-11.

Small K. A. (1982) The scheduling of consumer activities: work trips. *American Economic Review*, 72, 467–479.

Vickrey, W.S. (1969) Congestion theory and transport investment. *American Economic Review* 59 (Papers and Proceedings) 251-260.

Zhang, X.N. and Zhang, H.M. (2006) An exact expression of simultaneous departure time/route choices and a novel paradox. *Networks and Spatial Economics.* In press.

RESIDENT TRIP CHARACTERISTICS COMPARISON IN SMALL CITIES BETWEEN A DEVELOPED ZONE AND A GENERAL DEVELOPED ZONE

Jian-zhi GAO[1], Yao ZHAO, He-ling MA , Zhi-yun ZOU

[1]College of Traffic Science and Engineering, Huazhong University of Science and Technology, Wuhan, 430074,China;PH027-87556750;email:gaojianzhi2@163.com

Abstract

This paper analyzes the variability of resident trips and the intrinsic cause of the disparity in small cities which are at different economy levels. In the comparative analysis conclusion, economic disparity is the dominant factor which causes a difference in trip times if there was no clear difference on city dimension and layout. Seven typical small cities that are distinct in economy but similar in city dimension have been chosen in a developed zone and a general developed zone. The commonness and disparity of trip mode structure are analyzed, as well as trip purpose in both zones, the inner reason induced difference, average trip time and tolerance time of different travel modes. Verification is done by model building. The qualitative analysis and grey incident degree analysis proved that the economy is the main cause that induced different average tolerance time of small city residents. In conclusion, some suggestions are made on the basis of interrelations between traffic characteristics and economy.

Keywords: Small Cities; Resident Trip; Trip Characteristics; Economy

0. Preface

China's central region and east coast regions are different in urbanization process and economic life, belonging to a developed zone and a general developed zone respectively. Because of the rapid development of urban mechanization, cities in two zones are facing with great traffic pressure. Some small cities find it increasingly difficult to maintain smooth transportation because the traffic jam problem is also spreading to them. This paper tries to compare and analyze the resident trip survey data of some small cities in these two regions, and seeks the universality in resident trip characteristics and individuality in traffic features, which is intimately related to regional economic development. Based on the revelation of the internal relation between urban transportation and economic development, some suggestions are given to solve urban traffic problems of small cities in different economic development level regions.

1. General conditions of comparative cities

Small cities in this paper are defined as the non agricultural population of urban areas and areas with fewer than 200 thousand residents in cities (towns). Combining the participation experience of small cities' transportation planning, some small cities in Hubei Province as a developed zone and the southeast region as a general developed zone are chosen to compare and analyze. The character of the urban society economic development, urban spatial structure (single-center and centralized type) and other factors are also taken into accountant in this paper. Table 1 is the general conditions of comparative cities.

Table 1. General conditions of comparative cities

Category	city	population（10~4）	Area (square kilometres)	Urban residents per capita disposable income(yuan)	Per capita GDP(yuan)	Remarks
A general developed zone(Hu bei Province)	Dangyang	12.1	12.6	8571	13240	2004
	Anlu	12.6	11.9	4936	7639	2001
	Zhongxiang	17.09	16	5287	7266	2001
	Yidu	16.11	15.5	7950	9857	2002
A developed zone (southeast region)	Shengze	13.5	11.8	15325	41659	Jiangsu, 2005
	Nan'an City	18.5	17.18	10895	14925	Fujian, 2006
	Fuqing City	17.56	18.2	13241	20728	Fujian,2003

2. Trip times per person

2.1 Total characteristics of trip

Trip times reflect the capacity and need for mainly resident trips. A total trip time is measuring the index basically of the limit of ability to bear of an urban traffic system(DENG,Maoying,2000).Urban resident trip times per person are basically between 20~3.5 times/day. Since trip costs and trip times in small cities are low, trip times per person in small cities are more than that in big cities.

2.2 Comparative analysis of resident trip times of small cities in these two zones

1) Trend comparison

Table 2 is the comparison of trip times per person of some small cities in a developed zone and a general developed zone.

Table 2. Comparison of trip times per person of small cities

City	Trip times per person	year	City	Trip times per person	year
Dangyang	2.65	2004	Shengze	2.82	2005
Anlu	2.61	2001	Nan'an City	2.68	2006
Zhongxiang	2.54	2001	Fuqing City	2.74	2003
Yidu	2.58	2002			

Factors that strongly influence trip times are urban layout, urban scale, economic situation, life style, traffic condition and communication facilities, etc. Economy development and complete traffic facilities inevitably increase trip times. Because of the convenience of communication and environment quality reduction, trip times will go down. Economy disparity is the dominant factor leading to trip times distinction, if the urban scale and layout have no obvious difference. Small cities in a developed zone invest more funds in fundamental transportation facilities than those in a general developed zone. So the economic life is frequent. Resident trip times per person are distributed between 2.65 and 2.85. In the general developed zone, resident trip times per person of small cities are distributed between 2.54 and 2.65. An average of trip time per person is 2.59 and less than that in the developed zone.

2) Model verification

Establish a multiple linear regression model with trip times per person N (times/person·day), population P (10 thousand), area A (square kilometers), Urban residents per capita disposable income I (yuan), per capita GDP (yuan). The model is as follows:

$$N = 2.574 - 0.005 \times P - 0.0042 \times A + 1.65 \times 10^{-7} \times G + 2.4 \times 10^{-5} \times I$$

$$R^2 = 0.946$$

The calculation results are conformed with the observations. Fitting results are shown in Figure 1.

The correlation coefficient of independent variables and dependent variables is calculated based on the condition of linear correlation, which describes the degree of closeness between dependent variables and trip times. Table 3 is shown the correlation coefficient.

Results show that trip times per person has a positive correlation with the economy index, and a negative correlation with the urban scale. If urban layout and urban scale don't have a great difference, trip times per person will increase when the economy grows, but not increase unlimitedly. On the contrary, trip times per person will decrease when urban scale expands.

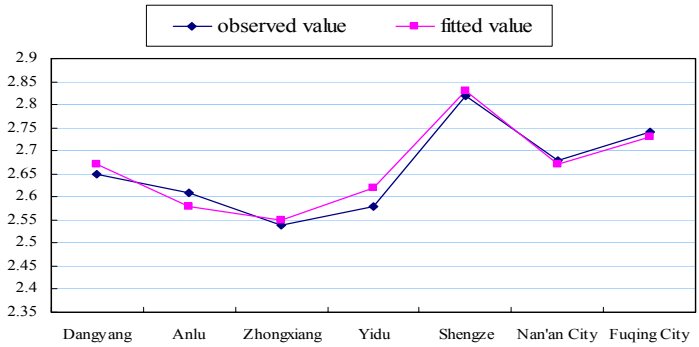


Figure 1. Fitting of observed value and model value

Table 3. Correlation coefficient of trip times per person and dependent variables

Index	Population	Area	Urban residents per capita disposable income (10 thousand yuan)	Per capita GDP (10 thousand yuan)
Correlation coefficient	-0.274	-0.327	0.939	0.856

3. Trip mode characteristic

Trip mode structure refers to the proportion of different trip modes which urban residents chose. It is an important index that reflects the development level of urban transportation. Resident's trip mode can be generally divided into walk, bicycle, motorcycle, public transit, private car, unit car, etc. Different trip mode structures require different urban transportation systems. The common affecting factors of trip mode structure are trip, social economy, urban layout, public transit, private transport, etc. Compared with big and mid-size cities, small cities have enormous differences in urban scale and urban space patterns. Motorcycles, walking and bicycles bring a competitive situation for the selection of trip mode. Table 4 shows the data of trip mode structure of small cities belonging to a developed zone and a general developed zone respectively.

3.1 Total characteristic

The proportion of walking, bicycles and motorcycles is above 81%. Private trip mode proportion in small cities is higher than that in big cities. Compared with big and mid-sized cities, small cities pay little attention to public transportation. Due to the backwardness of public transit, the trip mode proportion of public transit is low.

Table 4. Trip mode structure of small cities belonging to a developed zone and a general developed zone respectively (%)

city	walk	Bicycle	Bus	taxi	motorcycle	car	Unit	other	year
Nan'an	23.9	7.8	6.3	0.3	50.1	6.1	2.4	3.1	2006
Shengze	19.1	24.05	1.53	1.28	41.79	7.48	1.87	2.85	2005
Fuqing	32.9	15.92	8.38	1.53	35.43	3.8	1.19	0.84	2003
Dangyang	44.6	20.51	5.56	3.87	20.23	0.42	1.03	3.71	2004
Anlu	33.3	25.9	2.8	4.7	31.2	-	1.3	0.8	2001
Zhongxia	33.6	35.51	4.75	0.62	24.44	-	0.36	0.7	2001
Yidu	37.5	29.71	3.89	3.45	23.12	0.34	1.17	0.76	2002

3.2 Disparity analysis

There is diversity in the order of walk, bicycle, and motorcycle in different zones. The order of the three trip modes of small cities in a developed zone is motorcycle>walk>bicycle. The motorcycle has become the chief trip mode in Southeast China, because of its maneuverability and flexibility. The order of the three trip modes of small cities in a general developed zone is walk> bicycle> motorcycle.

There is significant disparity in the position of private vehicle trip mode. In small cities of the developed zone, the proportion of private vehicle trip mode is as high as 56.2% in Nan'an, Fujian Province. The lowest level is 39.42% in Fuqing City. In addition, the proportion is usually lower in the general developed zone than that in the developed zone. For example, in Anlu City, the highest proportion of private vehicle trip mode is 31.2%.

The same trip mode has a great difference in proportion. In the four small cities of the general developed zone, the trip proportion of walk in Anlu City is the lowest as 33.3%. In the three small cities of the developed zones, the trip proportion of walk in Fuqing City is the highest as 32.91%, and still lower than that of Anlu City. In the contrast, the lowest trip mode of motorcycle in the small cities of a developed zone, is still higher than the highest proportion of Anlu City in the general zone. The proportion of bicycles of small cities in the developed zone, is totally lower than that in the general developed zone. The private car trip can be overlooked in the small cities of the general developed zone. But in the southeast coastal small cities, the highest proportion of private trips is as high as 7.84% in the urban area of Shengze.

The main reasons for this diversity are the disparity of regional economy, urbanization level, urban layout and transportation management policy. In the small cities of a developed zone, urban residents' income is far higher than that of the general developed zone. Most of the residents tend to buy motorcycles, even cars. Substituting private vehicle trip mode for part of the non-vehicle trip mode will lead to the decline of the proportion of non-vehicle trip mode.

The process of small cities urbanization in the general developed zone has been quite slow. Residential land, commercial land, public land and industrial land mix with each other as they have done traditionally. People can reach destinations on foot or by bicycle. In the developed zone, the level urbanization of small cities is much higher than that in the general developed zone. Urban and rural areas have integrated together. Many residents have broken through the limitation of urban scope in natural trips. Motorcycle trip mode will be the first choice because of the slow development of public transit.

In the general developed zone, small cities advance a new policy towards restriction of motorcycle use in urban after carrying out the Unimpeded Road Project. But in small cities of the developed zone, motorcycle management tends to be more flexible. As a result, the developing trend of motorcycle accelerates.

4. Trip purpose

Trip purpose is trip motivation. This can be generally divided into going to work, going to school, shopping, entertainment, dealing with official affairs, visit relatives and friends, coming back home and others. Among these purposes, going to work, going to school and dealing with official affairs are rigidity trips, shopping, entertainment and other leisure trip can be considered as elasticity trips. Social economic development level, occupation, work (study) time will have an obvious influence on resident trip purpose. Table 4 shows the trip purpose structure statistics of small cities in a developed zone and a general developed zone. The total characteristics is that rigidity trip whose purpose is going home, going to work and going to school; these account for the proportion over than 75%. Besides going home, the largest proportion is the trip purpose of going to work, generally above 18%.

The proportion of trip for leisure such as entertainment and shopping is more than 11%. And the proportion of small cities in a developed zone is generally higher than that in a general developed zone (below 8.5%). The developed zone is usually an industrially advanced region which has more employment opportunities. So, the employment rate and the rigidity trip rate for going to work in a developed zone are higher than that in a general developed zone. The trip rate for going to school of small cities in a general developed zone is higher than that in a developed zone.

5. Trip time

Trip time is the one-way time from origin to destination in a trip. The decisive factors are trip mode, trip distance, etc. And trip time is an important index for evaluating the rationality of urban overall layout, trip accessibility, urban road network and public transit function.

Table 6 and table 7 respectively illustrate the statistical data of average trip times and trip time distribution of different trip mode of small cities in a developed zone and a general developed zone.

Table 5. Resident trip purpose structure of small cities belonging to a developed zone and a general developed zone respectively (%)

City	Job	Study	Official business	Shop -ping	Entertain-ment	Visit	Home	Office	Other	Year
Shengze	24.46	11.65	1.43	7.69	3.53	1.86	42.81	1.61	4.96	2005
Nan'an	23.44	11.27	1.54	9.15	4.28	2.17	42.19	1.23	4.73	2006
Fuqing	21.62	14.69	0.87	7.71	5.01	2.4	40.56	0.63	6.51	2003
Anlu	20.9	14.98	1.45	5.12	2.12	3.71	45.92	1.22	4.58	2001
Zhong xiang	18.12	17.39	1.45	5.84	1.77	2.06	48.14	0.7	4.53	2001
Yidu	19.75	18.64	1.49	6.21	2.84	2.65	42.66	0.6	5.16	2002
Dangyang	18.32	17.81	1.57	6.87	1.84	2.23	44.8	0.86	5.7	2004

Table 6. Average trip times of different trip mode (mins)

city	walk	bicycle	bus	taxi	Motor cycle	car	Unit car	other	year
Dangyang	17.79	18.06	23.95	8.42	19.94	22.6	37	18.57	18.39
Yidu	18.42	18.29	21.54	10.17	18.71	21.32	31.5	19.14	18.45
Zhongxiang	19.56	19.15	22.56	14.14	18.31	-	25	17.25	19.22
Anlu	16.97	17.91	27.36	12.09	20.35	-	17.93	20.17	18.37
Fuqing	15.66	17.45	25.49	24.54	21.21	22.77	23.18	18.74	19.26
Nan'an	16.07	16.98	22.05	26.05	23.45	20.37	21.43	21.36	20.80
Shengze	15.08	16.73	25.51	20.14	20.71	19.52	18.55	17.52	18.52

Table 7. Different time distance distribution of comparative cities(%)

Time distance (mins)	Fuqing	Nan'an	Sheng ze	Dang yang	Anlu	Zhong xiang	Yidu
5-10	17.31	13.54	14.66	14.43	15.34	13.44	20.73
10-15	24.39	19.35	27.76	27.09	33.5	22.54	22.78
15-20	20.3	16.8	17.99	16.67	13.6	21.05	20.76
20-25	13.81	20.29	14.55	19.16	20.81	20.37	18.56
25-30	8.38	10.36	10.35	12.73	8.73	11.33	9.74
30-35	5.43	7.38	5.49	2.44	3.8	4.04	2.28
35-40	3.73	3.06	1.92	0.81	0.25	2.48	1.65
40-45	1.93	2.73	2.8	1.72	1.61	1.12	0.89
45-50	0.86	2.05	0.64	0.41	0.28	0.78	0.58
50-55	0.5	1.32	0.75	0.86	0.39	0.7	0.49
55-60	1.25	1.44	1.6	0.16	1.13	0.51	0.56
>60	2.11	1.78	1.49	3.52	0.56	1.64	0.98

5.1 General Characteristic

Trip time cost of different trip mode is a main factor for checking the city transport structure. The investigation has shown that the general property of trip time cost is larger if the city dimension is larger, and the trip time cost is, in general, similar in same dimension city.

The average trip time cost in the seven typical small cities is about 20 minutes, and it is reasonable although it is a bit larger in Nan'an. The average trip time cost is larger than that of the other four cities because the city dimension of Fuqing, Nan'an and Zhongxiang is larger, too.

In the comparison of different trip mode, the trip time cost of non-vehicle mode is less than the average trip time cost, and the trip time cost of walk mode is the minimal one. In the vehicle mode, the trip time cost is general higher than the average except for the taxi trip, and the trip time cost of public transport is the maximum one. It is shownthat public transport service network between the two areas is not reasonable and its service quality is lower.

5.2 Comparison Characteristic

Because of the difference in social economy level, resident income, transport management policy and trip structure, there are some differences of trip time cost in two areas.

1) Difference of trip time cost

In the developed area, such as Fuqing, Shengze and Nan'an, the proportion of vehicle mode is large in those small cities, especially the motor mode. Some walk and bike trips are replaced with motor trips, and the result is that the trip time cost of non-vehicle mode is less than that in generally developed areas. Contemporarily, industry is expanding into the small city in the fully developed area and the integration of rural and urban area is enhanced. But there are shortages of the advanced transport services, such as combined public transport; the motor is the main transport tool in traffic between the urban and the rural. And the result is that its trip time cost is larger than that in the general developed area, such as Zhongxiang, Dangyang. In the developed area, such as Fuqing, Shengze, and Nan'an, the motor service for passengers is serious, and the taxi is servicing mainly between the rural and the urban. Thus, the trip time cost is larger. But, in the generally developed areas such as Dangyang, the taxi is servicing mainly the city area, and the result is that its trip time cost is apparently less than that in the developed area.

2) Difference of tolerance trip time

There is an apparent difference in the tolerance time of one trip. Define the acceptable maximum trip time cost as the trip time of 90% resident(XU,Jiqian,1998). On the basis of definition, tolerance time of one trip in seven cities is showing in the following Figure 2.

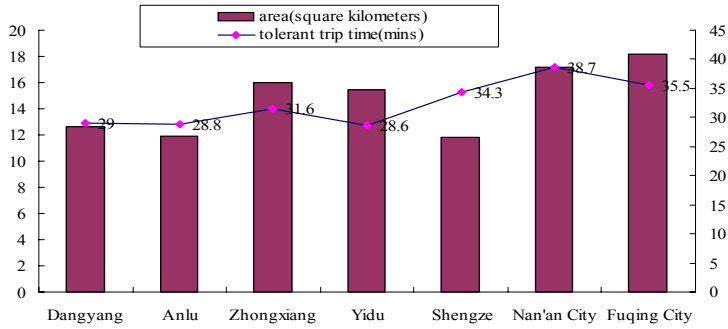


Figure 2. Tolerance trip time comparisons

The influence on the trip time cost is complex. Generally, the trip time cost is more if the trip distance is longer, and the trip time cost is less if the proportion of vehicle trip is larger(LU,Jian,2004). However, the comparison shows that, the trip tolerance time is reversely higher (about 5 minutes) in the developed area where the proportion of vehicle trips is larger, if the city dimension is similar. It is shown that the satisfactory threshold of residents in the developed area is relatively lower than that in centre city.

The grey correlation is used to checkout the above conclusions. First, divide the maximal trip time, urban scale and economy development level into interval values. Then calculate the order difference of rank matrix $\Delta_i(k)$ and the correlation degree of maximal trip time cost, urban scale and economic development level (Table 8). The computing formula is as following:

$$\zeta_i(k) = \frac{\underset{i}{Min}\,\underset{k}{Min} V_i(k) + \rho \underset{i}{Max}\,\underset{k}{Max} V_i(k)}{V_i(k) + \rho \underset{i}{Max}\,\underset{k}{Max} V_i(k)}$$

Correlation Degree $\gamma_i = \frac{1}{n}\sum_{k=1}^{n} \zeta_i(k)$

Table 8. Maximal trip time cost and correlation degree of factors

index	population	area	Urban residents per capita disposable income (10 thousand yuan)	Per capita GDP (10 thousand yuan)
Correlation degree	0.115	0.237	0.756	0.613

The calculation result shows that the order of relevant factors is urban residents' per capita disposable income >per capita GDP>area>population, according to the correlation degree of tolerance trip time. The examination result shows that economy is the dominant factor which causes the diversity of tolerance trip time in these two zones.

The reasonable explanation is that the economy level of small cities in Centre China is low, and there is obvious disparity between urban areas and rural areas. Resident trips centralize in the urban area according to the administrative regions. Contrarily in the developed zone, there is a bit difference between urban areas and rural areas. Residents communicate continually and most trips have broken through the urban limitation.

6 Conclusion and advice

6.1 Conclusion

Economy is the main factor for the variability of resident trips in the fully developed city and the general developed city, if there are no apparent differences in the urban scale and its space patterns.

The average trip times per person·day of small cities in the developed zone is more than that in the general developed city. And its proportion of vehicle trip mode is far higher than that in the general developed area. But its vehicle trip mode is shrinking more and more. In the developed area, the elastic proportion of trip for leisure and amusement is higher, but the average time cost of vehicle trip is lower. And there is some variability in trip time cost between the fully developed area and the general developed area. The most apparent difference is the tolerance trip time, which is 5 minutes longer in the fully developed area than that in the general developed area.

6.2 Advice

Based on the internal relations between urban transportation characteristic and economy, the recommendations are given as follows:

1) As the rapid development of economy and urbanization, urban and rural area will be integrated together and more residents will break through the limitation of urban scope and trip distance in the general developed area of centre China. The advice is that tolerance trip time is programmed to be 40 minutes. In the developed area, the small city will be more rapidly integrated into area economy system, and the trip distance will be longer. And it is essential to enhance the construction of transportation infrastructure in urban and rural area. Long-distance rapid transportation system should also be constructed.

2) In the small city of a general developed zone, the human-oriented design and program for transportation system of walk and bicycle should be enhanced. The non-vehicle trip mode should be retained as a major one and the motorcycle trip mode should be put under control in the developed small city. The trip mode of walk and bicycle should be advocated for the short distance trip.

3) Multi-level and diversity passenger transport trip mode should be constructed to satisfy the demand of elastic trips for leisure and amusement as the economy is rapidly developing. The program and construction of public transport in rural and urban areas should be accelerated. The public transport service should be improved. So, a transport system is constructed, which can satisfy the travel demand in the process of integration of rural and urban area.

4) In the centre China, there is very large room for economy development because of the chance of centre-rising. So, some margin should be left in the city transport bearing capacity in the future program. In the developed area, infrastructure construction is now beginning to take shape, and transport need is relatively large. Thus, transportation demand management should be carried into execution in fully developed small city prior to it in the general developed area.

References

College of Traffic Science and Engineering of Huazhong University of Science and Technology.Report of Transport Planning about Dangyang、Anlu、Zhongxiang、Yidu、Shengze、Nan'an、Fuqing.

DENG,Maoying,and XIE Li(2000).“The Characteristics of the Trips of the Local Residents and the Transport Policy in Guangzhou.” City Planning Review, 24 (11), 45～49.

LU,Jian,and WANG Wei(2004).“Study of Resident Trip Time Consumption Characteristic.” Journal of Highway and Transportation Research and Development, 10.

XU,Jiqian(1998).“Strategy on the Sustainable Development of Urban Passenger Transportation ”. City Research, 6.

THE STRATEGIES OF MULTIMODAL TRANSPORT TRANSFER SYSTEM FOR WORLD EXPOSITON 2010 SHANGHAI BASED ON TRIP CHAIN ANALYSIS

Wusheng Liu[1], Ya Wen[2], Kefei Yan[3],Chaowei Yu[4]

[1]School of Traffic and Transportation Engineering, Tongji University, Siping Road 1239, Shanghai, P.R.C, 200092; PH and FAX (86-21) 65980487; email: lwusheng@163.com
[2]School of Traffic and Transportation Engineering, Tongji University, Siping Road 1239, Shanghai, P.R.C, 200092; PH and FAX (86-21) 65980487; email: yahuwen@hotmail.com
[3]School of Traffic and Transportation Engineering, Tongji University, Siping Road 1239, Shanghai, P.R.C, 200092; PH and FAX (86-21) 65980487; email: yankf@mail.tongji.edu.cn
[4]School of Traffic and Transportation Engineering, Tongji University, Siping Road 1239, Shanghai, P.R.C, 200092; PH and FAX (86-21) 65980487; email: xiaoyuer01225@163.com

Abstract: It's well known that World Expo 2010 will be hosted by Shanghai, which is a metropolis with a population of over 16 million. Approximately 200 countries and international organizations are expected to participate in and more than 70 million people plan to pay visits to World Expo 2010, making it the largest of its kind in the history of World Expo. How to make World Expo 2010 become one successful global event and how to make it achieve the theme "Better City, Better Life" in such a dense traffic and highly populated city are big challenge. Construction of a multimode transport transfer system to transfer visitors from private or inefficient transport modes to efficient ones is being proposed by many experts. The specific strategies of the transfer system are studied in the paper. Firstly, the definition of Expo trip chain is put forward. The transferring characteristics of transport modes are explained by analyzing a stated preference (SP) survey of Expo visitor trip chain. The survey was conducted in two Shanghai airports. Then, experiences of the transfer systems from several previous Expos are studied. In the end, based on the trip chain SP survey and the cases analysis, the strategies for Shanghai multimode transport transfer system are elaborated.

Key words: *strategy, multimodal transport transfer system, world exposition 2010 Shanghai, trip chain, SP survey*

1、Concept of Expo Trip Chain

The term "Trip Chain" is used to analyze series of trips from a new aspect. Different from traditional "trip", trip chain pays more attention to the transferring and connecting process of travel. Trip chain consists of three basic elements: anchors, anchors and trip links. The term "anchors" is used to describe the start and end point of trip chain. The term "stops" means the temporal transferring point between anchors. And the term "trip links" means the travel space among anchors. Considering the travel aim, connecting time and the transferring of different transportation modes, we can define Expo Trip Chain as the sequential series trip aiming at Expo visit, using anchors as the start and end point, using the transport hub as the transferring space between different trip links, with transfer time being not more than 30 minutes. Based on the sequence of different trip links, we can define upriver trip links and downriver trip links: upriver trip links come before downriver trip links. The definition describes the whole process of one Expo travel, express modes transferring by transportation hubs. Therefore, Expo trip chain provides a basis for studying the transferring and connecting of Expo visitors.

2、The Stated Preference (SP) Survey of Expo Trip Chain

In order to find out some characteristics of Expo visitors' transferring, two SP surveys of Expo trip chain were taken separately at Shanghai Pudong and Hongqiao international airports in July, 2005. Part of the survey form is given in table 1.

Table 1. The Expo SP survey form

1. Age: ①□< 30 ②□30～45 ③□46～60 ④□>60

2. Occupation: ①□Government Staff ②□Commercial Staff
③□Student ④□Retired ⑤□Other

3. Your monthly income: ($): ①□<1000 ②□1000～2000 ③□2001～5000
④□5001～8000 ⑤□8001～10000 ⑥□>10000

4. Would you visit EXPO'2010: ①□Yes (A□Specially B□Passing-by) ②□No

5. Which of the following choices would you make if you visitesEXPO'2010:

①□Private car→Parking & Ride→Rail rapid transit (or Bus Rapid Transit) →EXPO site

②□Bus→Rail rapid transit→EXPO site ③□Hotel→EXPO Bus→EXPO site

④□Tour bus→EXPO site ⑤□Others

2.1 Analysis of the Survey Results

By analyzing 1500 samples, 1367 valid records were acquired from the two surveys. The results of trip chain selection are given in table 2, where "total" refers to all visitors.

In table 2, trip chain 1, 2, 3, 4 and others represent the five trip chain in question 5 in table 1 respectively. 7% more Passing-by-visitors than Special-visitors select chain 3. 8% fewer Passing-by-visitors than Special-visitors select chain 1. It should be noted that no transferring occurs in trip chain 3. Comparatively, the number transferring in chain 1 is more than that in chain 3. Therefore, Passing-by-visitors prefer to select the chain with less transferring.

Table 2. Results of visitors' EXPO trip chain survey

chain type / visit type	trip chain 1	trip chain 2	trip chain 3	trip chain 4	others
Special visit	20.39%	37.65%	21.05%	14.38%	6.54%
Passing-by visit	28.59%	39.12%	14.22%	10.81%	7.25%
total	24.30%	38.39%	17.77%	12.66%	6.88%

As a whole, trip chain 2 is the most favorite. More than 38% visitors chose it. Trip chain 1 comes next. About 24% chose it. Thus, high occupancy, rapid public transport is the main mode of trip chains.

2.2 Correlation Analysis of Visitor Characteristics and Trip Chain Selection

(1) Occupation

The correlation analysis of visitor occupation and his/her trip chain selection is shown in table 3. Commercial staff prefer to choose trip chain 1 and 3. Trip chain 2 needs more time and less cost than chain 1 and 3. Students prefer trip chain 2.

Table 3 Relationship of visitor's occupation and his/her trip chain selection

chain type / occupation	trip chain 1	trip chain 2	trip chain 3	trip chain 4	others
Government Staff	14.45%	18.00%	16.99%	16.89%	15.97%
Commercial Staff	56.09%	46.35%	58.69%	39.73%	43.70%
Student	13.88%	24.60%	11.97%	17.81%	15.97%
Retired person	6.52%	7.31%	3.47%	18.72%	19.33%
Others	9.07%	3.74%	8.88%	6.85%	5.04%

(2) Age

The correlation analysis of visitor age and trip chain selection is shown in table 4. Visitors younger than 30 years prefer to choose trip chain 2 and 4.

Table 4 Relationship of visitor's occupation and his/her trip chain selection

chain type / age	trip chain 1	trip chain 2	trip chain 3	trip chain 4	others
<30	31.71%	42.11%	35.27%	45.95%	39.60%
30～45	49.71%	41.04%	46.51%	38.92%	48.51%
46～60	14.00%	12.01%	15.89%	12.97%	10.89%
>60	4.57%	4.84%	2.33%	2.16%	0.99%

(3) Monthly Income

The correlation analysis a of visitor's monthly income and trip chain selection is shown in table 5. Visitors whose income are below 1000 prefer to choose chains other than chain 1 and chain 3. And many of those whose incomes are between 5001~8000 choose chain 1. In general, visitors with lower incomes are more inclined to choose chains with less cost. And those with higher incomes are more inclined to choose chains with less journey time.

Table 5 Relationship of visitor's occupation and his/her trip chain selection

chain type / monthly income	trip chain 1	trip chain 2	trip chain 3	trip chain 4	others
<1000	12.17%	21.43%	8.84%	20.22%	17.89%
1000～2000	10.09%	14.65%	11.24%	19.66%	11.58%
2001～5000	35.91%	44.14%	42.17%	39.89%	29.47%
5001～8000	21.96%	14.10%	12.45%	11.80%	17.89%
8001～10000	8.61%	2.38%	13.65%	5.06%	7.37%
>10000	11.28%	3.30%	11.65%	3.37%	15.79%

Form above analysis, it is easy to see that monthly income has a greater influence on visitors' selection than age and occupation, and age is the less important factor.

3. Brief Summary of Passengers Transferring System of Previous Expos

3.1 Traffic Transferring System (including P+R) of previous Expos

A brief summary of traffic transferring system (including P+R) of previous Expos is listed in table 6 below.

Table 6 Comparison of traffic transferring system (including P+R) of previous Expos

city	transfer mode	The condition of P+R parking lots	The arisen problem in use
Osaka (1970)	Railway Special line bus Group bus Private car	Passageway and parking lots were set at four entrances. Some distance existed between the entrance and parking lots, and buses were used to transfer. Parking lots were 0.75 million square meters large, having more than 23 thousand parking lots.	Chaos occurred when the number of ingoing visitors increased rapidly. When the number of visitors arrived at the peak of 0.85 million on Sep.5, 5000 visitors remained at the EXPO site until the next day, in despite of a lot of means such as increasing the railway trains and trip bus had been implemented. In order to prevent an accident, administrators had to close the Expo provisionally.
Hanoverian (2000)	Railway Other public traffic mode Private car Expo bus Other	Almost 35 thousand parking lots were set in the suburb, and 25 thousand parking lots around the Expo site.	Because of the total number of visitors was less than the forecast, and the share rate of cars had been overestimated (51% anticipated), the use rate of the 35 thousand suburb P+R lots was only 1%, the administrators decreased and even exempted the cost of parking in order to encourage parking lot use, then the using rate of car rose from 25% to 34%.
Aichi (2005)	Railway Underground Shuttle bus Public bus Private car	In Nagakute's site, 3 main entrances were designed, and 1 was at Seto's. There are 6 private car parking lots around, and having 10.6 thousand parking lots.	The large-scale parking and ride mode have been implemented successfully (a 20.2% visitor flow had actually used the system)，the P+R have been ratified, but there also, a lot of visitors used the civilian parking lots(7.6%) and deliver –receive mode(3.8%), which caused the jam condition. Some problems occurred on the transferring system between the magnetic extension line and underground. On Sep.4 2005, when the magnetic line had an accident, there were 8000 visitors stuck , this caused chaos.

A previous Expo paid attention to the usage of public transit, and regarded the urban public transit system, especially urban rail transit system, as the important modes for passenger traffic. The characteristics of an urban traffic transferring system are as follows:

(1) Rail transit played a leading role and bus transit was complementary mode in main passenger traffic aisle

During many previous Expos, public transit mode played a very important role. Rail becomes the main trip mode among many kinds of trip chains; it not only connects each junction of transfer point, each P+R lots and enters directly in Expo area, but also sets up the frame of the entire urban transfer-and-feeder system. Rail

transit completes main passenger traffic flow tasks, and feeder bus system makes up for its disadvantage. For example, during Osaka Expo, the share of public transit reached 75.8 %, among which rail transit reached 62.8%; during Aichi Expo, public transit accounts for 64.5% (including P+R) and rail transit shares 49.5%.

(2) Providing special service feeder bus lines during Expos

Special service feeder bus lines were seen at some Expos such as Osaka and Aichi. The special bus lines for Expos often circulated around or through each transfer junction, P+R point and Expo site. During Aichi Expo, share rate of Expo feeder bus was 20.2%，and 17.2% for Osaka Expo.

(3) Indispensable P+R

Several Expos took place in developed countries where cars are popular. Although their Expo sites weren't located in the center of an urban area like Shanghai, high strength passenger traffic compelled planners to treat cars cautiously. These Expos nearly planned and built large scaled P+R lots. Aichi Expo transferred 20.2% of car passengers on six P&R lots; 23000 P+R lots around Osaka Expos fields also succeeded very well; Hanover Expo set about 35000 P+R lots in downtown because in the trip survey 51% of visitor preferred to choose cars.

3.2 Brief summary to traffic transfer system at the suburb of city

Following is the brief summary of traffic transfer system in the suburbs of a previous EXPO:

Table 7. Comparison of traffic transfer mode at the suburb of city

city	transfer mode	comparison of operation results
Osaka (1970)	Railway, Bus, Air transportation, Waterway, transportation	Railway carried most visitors. There were two lines connecting the Expo site directly. A big transfer station was located east of the Expo site. There were special line buses moving between the cities of the center area and the Expo site. Osaka airport had been expanded, and the Osaka wharf and the Kobe wharf have been built or rebuilt. Watercraft replaced hotels to provide abode. 69 thousand of visitors arrived at Osaka wharf by 304 ships. On the other hand, visitors who arrived at the Kobe wharf increased. There were 70 thousand on March, 160 thousand on August.
Hanoverian (2000)	city light rail, rail	A high-efficiency city light rail system made a close connection between Hanoverian airport, city and Expo site.
Aichi (2005)	Railway,Air transportation, Ordinary public traffic mode	Because the Expo site was quite out-of-the-way, a public traffic system had not been developed fully. The visitors who arrived in center international airport had to make 2 or 3 transfers to get to the Expo site. It is quite convenient to transfer by electric car from Nagoya. It is quite convenient to transfer by railway.

During previous Expos, visitors arrived at airports, railway stations and other external traffic hubs, then by rail transit or other traffic modes in cities they could get to Expo site, so the connection between external traffic and city traffic had been realized.

4. The 2010 Expo multi-modal traffic transferring tactic based on trip chain analysis and experiences of previous Expos

4.1 The Mode Split and Tactic Selection near the entrance of Expo site

After some research, the aim of 2010 Expo traffic is quite clear: using the massive capacity rapid rail traffic as the main mode, utilizing the flexibility of ground transfer and the special line bus, using waterway bus to overcome the defect of Expo site. According to <World Expo 2010 Shanghai China Registry Report (Abstract 5)——Expo Site Plan>, rail transit will carry 50% visitors, ground traffic (including special line bus, travel bus and conventional public transportation) will carry 35%~40%, waterway will carry 5%, other transportation (VIP car, walking, bicycle) will carry about 5%~10%. Based on the experience from previous Expos and the results of trip chain surveys mentioned above, the split of traffic modes and tactic of every trip chain will be further analyzed below.(P.S.: only ground traffic modes were considered in the trip chain investigation. If a waterway can carry 5% of the total visitors, the final share of every mode can be reached by multiplying the investigation results with 0.95)

(1) Rail traffic

According to the trip chain SP survey, 58.9% of the respondents are inclined to choose rail traffic, which is 10% higher than expected. From the comparison of previous Expos, the rate of rail traffic at Osaka, Aichi and Hanover were all among 50% to 65%, which means the share of rail traffic near the Expo site may increase if Shanghai's rail traffic can provide enough high capacity in 2010. Therefore, for the rail traffic (including maglev), it is necessary to improve the capacity to more than 50% of the total traffic, even 60%. The construction of rail and maglev (such as Hu-Hang maglev line and the extension line from Long yang Lu to EXPO area) lines connecting EXPO site and other places should be accelerated. The planning and traffic management of every rail line and the station within the area should be carefully considered.

(2) Ground Public Transportation

From the SP survey, the share of Expo special line bus (conventional bus) is 17.77%. The rate of tourist buses carrying visitors from the Yangtze River delta and outer suburbs of Shanghai is 12.66%. The sum of these two modes achieves 30.43%. Thus, the two modes are also very important. The position design of Expo special line bus, the tourist bus routine for entering the area, and the management of parking reservation near the entrance of EXPO area should be studied thoroughly.

(3) Waterway Traffic

Waterway traffic is new. Waterway is expected to carry 5% of the visitors. The management of waterway buses and the connection among the dock and other buses should be further studied.

(4) Other Transportation Modes

Besides the four trip chain mentioned above, 6.54% of the respondents from the trip chain survey choose other modes such as VIP cars, walking and bicycle to reach the Expo site.

4.2 The Tactic Selection about the Multimodal Transfer System in the Area of Shanghai City during Expo 2010

In 4.1, modal split and basic tactic of the trip chain near the entrance of Expo site were analyzed. In order to achieve the aim, the connection between the upriver trip links and the downriver trip links can never be over-emphasized besides the necessary intensive construction of transportation infrastructure near Expo site.

According to the analysis of trip links and transfer stops of trip chain, the transfer tactic about every traffic mode in the urban center and periphery can be elaborated as below:

(1) The Urban Center

Upriver and middle-river trip links occurred in this area. About 60 percent of the visitors will be transferred here. It is very important to see whether upriver visitors can be transferred to the downriver trip links or not, because this will decide the realization of mode split aim (90% public transportation) in downriver trip links. Following is the connection and transfer planning that should be regarded in the urban center:

① Urban Rail Transit

On one hand, more urban rail transit lines should be built, on the other hand, the lesson of management of Maglev in Aichi Expo should be learned. The transportation capacity between Maglev and other vehicles should be well matched in order to avoid visitors staying on platform too long.

② Public Transportation

Two problems should be paid attention to: the first problem is the setup of Expo special line buses. Special lines of Expo buses should be set up in hotel and inhabitant convergence areas. The second problem is that the bus arrivals should be on time. As the upriver trip links of urban rail transit, buses are the main body that joins the upriver trip links of rail transit.

③ Private Cars

Learning experience from Expo 2000 Hanover Germany, all kinds of infrastructure of Park and Ride (P＋R) and parking guidance system should be built according to the actual conditions so that Expo can offer convenience for the transfer of private cars and taxi, and it is the basis that limits private cars entering Expo site.

(2)The Periphery of Urban Center (the suburb)

On one hand, the visitors from the Shanghai exurb and external cars should be excluded. By setting up Expo specialized lanes and P+R hubs in urban center periphery, these visitors can be transferred by Expo bus running in Expo specialized lanes. In this way, the most upriver trip links will be transferred to other traffic modes. On the other hand, for visitors from air, railway, road and waterway, the transferring from the urban periphery to hotels in urban area should be well planned. And experiences from Aichi Expo and Osaka Expo can be used.

4.3 The main countermeasures that ensuring the strategies to be implemented

(1) Park and Ride of car (P&R)

, 24.3 percent of the people surveyed expressed that they will park their cars in the rail stations and enter the garden of EXPO by rail transit though they have their own cars. Therefore, the rail station should connect directly with city expressway, main roads and distributing center when planning so that cars can enter or leave the garden of EXPO conveniently, and the time of transfer will be reduced.

(2) Transfer between normal bus and rail transit

Because passenger flow is centralized and some passengers aren't familiar with the city traffic, when transferring between normal bus and rail transit, we should plan legible and intuitionist transfer sign and information services, and build waiting rooms for those passengers who can't leave on time for some reason. Furthermore, we also should plan reasonable channels that connect with those rail stations.

(3) Transfer between other traffic model and rail transit

According to the characters of passengers who transfer between walking and rail transit, we should plan much more space for walking passengers, offering walking facilities, with the countermeasures ensuring passengers safety. Moreover, we should design some bicycle parks around rail stations.

(4) The connection between other traffic models and special line bus of EXPO

A special line bus of EXPO should be a big bus, and we should offer enough parking and scheduling space when planning and do some jobs about propagandizing special line bus of EXPO, build some induction signs in order to reduce transfer time. In addition, most connection stations of EXPO bus should be located around some hotels or transportation hubs, so we will optimize traffic organization at these areas, offering special bicycle routes and etc.

References

Gregory L, Newark, Pnina. O. Plaut.(2005).“Shopping Trip Chaining Behaviors at Fringe Malls in an Emerging Economy.” *TRB 84th Annual Meeting.* Washington, D. C.

Jianqing Chen.(2003).“The countermeasures of scheming tour establishments in Expo 2010 in shanghai city.”*Journal of Shanghai University (Society Science Edition)*, 18(3),12-16.

Mcguckin, Zmud, Nakamoto.(2005).“Trip Chaining Trends in the U.S- Understanding

Travel Behavior for Policy Making."*TRB 84th Annual Meeting*. Washington, D.C.

Ping Xiong, Yongping Liu.(2006)."The analysis of the transportation planning and implementing effect of Aichi Expo in Japan."*Traffic & Transportation*, 16(6),1-5.

Harmonious Urban Development and Strategic Transportation Planning in China

CHIAO Kuo-Ann[1], ZOU Xi[2], YANG Fei [3]

China is currently experiencing rapid economic growth. A byproduct of this economic expansion is an unprecedented investment in the country's transportation infrastructure. However, there are two fundamental problems with the transport system that is presently being developed. First, it is largely highway oriented, which impacts future land use patterns, the environment, congestion, and issues of equity since many do not presently own automobiles. Second, the planning of transportation investments is not being undertaken in an integrated fashion that incorporates the input and objectives of all stakeholders in a given region. As a result, opportunities to enhance mobility, support economic development, and achieve sustainability objectives are squandered.

Because the expansive economic growth in China is being accompanied by significant transport infrastructure investment, this is a pivotal moment for many Chinese cities. The decisions made regarding these investments have long-term implications for the environment, development patterns, equity issues, and general social well being of China. Moreover, once transport development decisions are made they are reversible only at extremely high cost. Thus, it is critical that the right investments are made at this juncture, where the character of China's transport infrastructure remains in the balance. The decisions made today will determine whether or not China achieves its goal of creating a harmonious society supportive of the environment, the economy, and the social welfare of all her people.

In the context of China's expansive growth, urban transport officials and professionals need to transform their thinking and problem-solving methods in order to effectively manage a complex transportation system. They should be able to collect needed data and information, use appropriate planning and analytical methods, and develop appropriate policies and planning procedures. This will enable them to solve the traffic problems of today as well as to plan for the needs of a future given limited resources.

Chinese leadership recently proposed the overall goal of building a harmonious society by 2020, a society characterized by, "democracy and the rule of law, fairness and justice, sincerity and amity, full of vitality, stability and order, and in which people and nature live in harmony." As the crucial link of economic activities, urban transport development in China must embrace this objective and coordinate with the stakeholders from all affected strata. Operating through appropriate channels, they

[1] Kuo-Ann Chiao (焦国安), New York Metropolitan Transportation Council (NYMTC), 199 Water St, New York, NY 10038; PH: (212) 383 7212; FAX: (212) 383 2418; e-mail: kchiao@dot.state.ny.us

[2] Xi Zou (邹熙), STV, Group, Inc., 225 Park Avenue South, New York, NY 10003;PH: (212) 777-4400; FAX: (212) 529-5237; e-mail: gavin.zou@gmail.com

[3] Fei Yang(杨菲), NYMTC, New York 10038, PH: (212) 383 7229; FAX: (212) 383 2418; e-mail: fyang@dot.state.ny.us

must reflect the interests, needs and objectives of all parties, and form consensus on urban transport plans, standards and courses of action. (1)

However, urban leadership, vested with decision-making powers should understand that transport development is not only the responsibility of individual government agencies, but is also the common responsibility of all relevant parties in a region. Transport issues must be coordinated through appropriate investigation and consultation to reach an optimal policy decision that will support economic development, while taking into account people's living environment. It is only by transforming the thinking process and accommodating input from various stakeholders, that a truly collaborative transport planning process can be established.

This paper sets forth some approaches that decision makers can use to ensure that planning is done in a coordinated manner and that long term objectives are achieved. They include:

- regional development and shared vision,
- a new management model of urban transportation planning,
- regional transportation planning and financial needs,
- regional transportation planning and decision-making process, and
- regional transportation planning and sustainable development.

Finally this paper provides suggestions for Chinese urban transport planning and development that could help to achieve the goal of harmonious urban development.

1 Regional Development and Shared Vision

A vision is the long-term picture which people look forward to achieving. It includes goals, missions and core values. A shared vision is the common mission, values and expectations of achieving a future situation held by everyone in a given region or organization. A shared vision not only arouses people's common expectations and provides great impetus among a population, but also facilitates coordination among individuals or localities when their interests conflict. This helps to defuse conflicts and promote harmonious development.

The existence of a shared vision among stakeholders can inspire the constructive participation of members of a society and coordinate regional development. In urban and regional development, especially in transportation issues, there will be a conflict of interest between social strata, the local community and various government agencies. However, it these stakeholders embrace a shared vision of harmonious development it could facilitate a collaborative planning process.

Transportation development programs should be considered from the perspective of regional socio-economic progress, and be based upon a shared vision generated by stakeholders in the region. Transportation impacts various social goals and serves as a link between economic and social activities. However, transportation usually represents conflict of interests from various involved parties that make it an issue critical to the general public, interest groups and various government agencies. It will be very difficult to reach a balanced and harmonious solution if any of the involved parties is ignored.

Traffic congestion and associated pollution lower quality of life and hamper economic growth. Inappropriate transportation development will not only fail to obtain proper returns for large-scale investments, but also create some negative long-term side effects. Thus, it becomes a great challenge for Chinese transportation policy makers and professionals to maintain the contribution of the transportation industry to social and economic growth while reducing its negative impact on other socio-economic objectives. Making transportation system the vehicle for promoting harmonious social development is a common aspiration of all stakeholders.

The Communiqué of the Sixth Plenum of the 16th Central Committee of the Communist Party of China(2) asserted that the party "must transform the government system, integrate its resources, and improve its efficiency and functions. This transformation will be undertaken under the leadership of the Party committee, with the government facilitating cooperation among other organizations and public participation. Service should be the goal of the transformed government system."

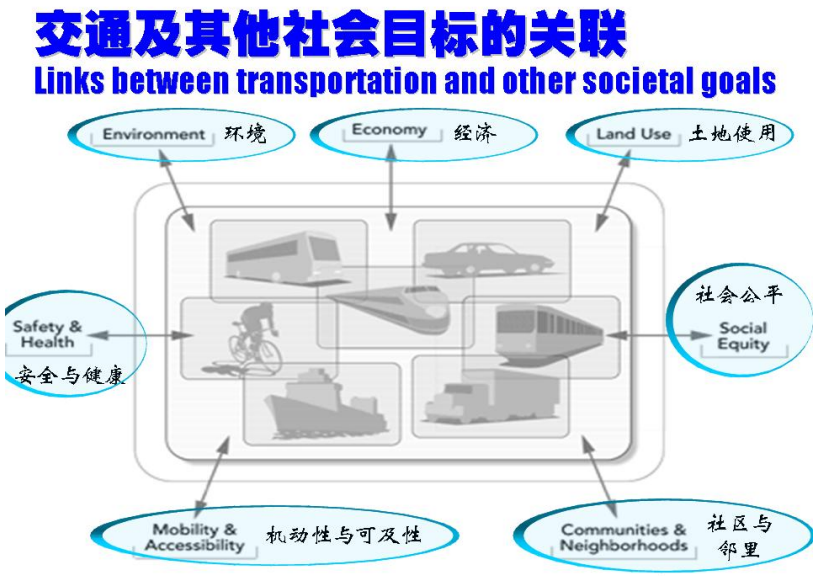


Figure 1 Links between transportation and other societal goals(3)

This proposed government management model in China not only responds to the goal of establishing a harmonious society, but is also in line with the inherent characteristics of transportation planning, construction and management. It also complies with the needs of transportation planning and management, and is conducive to the establishment and implementation of a shared vision. From this perspective, urban transportation planning and management should not be the sole responsibility of traditional transportation planning and management agencies. All stakeholders should

be involved in building a shared vision, a standardized mechanism for coordination, and a process of finding the best options.

The establishment of a shared vision and coordination mechanisms that regulate urban transport planning, construction and management is essential for building a harmonious society. Transportation and other closely related regional issues such as population growth, economic development, air quality, land use, social justice, mobility and accessibility, and health and safety must addressed in view of regional strategies and through a continuous, comprehensive, and cooperative (3C) planning process. Transportation decision-makers and professionals will work with various stakeholders with regional perspectives, using system engineering methods to decide the direction and priorities of transportation investment and realize the region's shared vision.

2. Introduction of A New Transportation Management Model

A new transportation management model requires that government leaders, transportation decision-makers and professionals change their way of thinking and update their problem-solving methods. The major elements of the new transportation management model include:

- engagement of a collaborative planning process,
- collection and analysis of transportation related data,
- application of appropriate planning and system analysis methods, and
- development of proper strategic plans, policies and planning procedures.

This will enable the decision-makers and transportation professionals to manage a complex transportation system so as to be able to solve the traffic problems of today and to envision for the future with limited resources.

In the new management model, the government departments and transportation professionals must cooperate closely with the related participants, including not only government agencies, which provide places for residences, schools, offices, industrial production, logistics and parks, but also the private sector and civil organizations.

Currently in China, transportation-related matters are dealt with by different government agencies at various levels. The distribution of the functions between departments has not yet adapted to the new requirements of a service-oriented and public-involved management model. Some distribution of functions is still under the shadow of the planned economy era. Meanwhile, the government has no suitable planning framework for relevant government agencies and civil society groups to communicate with each other on regional issues and shared-visions. Without the proper coordination mechanism, it is hard for participants at all levels to share resources and information.

The realization of the new transportation management model is a step-by-step process. First, there is a need for legislation to regulate investment in urban areas. Requirements and standards for urban transportation planning, coordination and management need to be formulated. Second, government should provide funding for

metropolitan areas to set up a "regional transportation planning organizations". Such organizations would be responsible for: (1) coordinating regional transportation planning, construction and management; (2) training project management engineers and planning professionals; (3) collecting, analyzing transportation related data and providing public and professional information services; (4) developing planning tools for cities and the regional transportation planning study.

The transportation planning organizations must be regional in scale in order to properly address issues that arise in the process of urbanization and urban economic development, which closely integrate cities and their surrounding area.

3 Regional Transportation Planning and Financial Management

Transportation finance is one of the major issues in regional transportation planning. Currently, in many underdeveloped regions in China, obtaining proper funding is still a major challenge. Due to traditional modes of thinking Chinese government officials still concentrate mainly on getting funding, rather than using funds efficiently, improving system effectiveness, and evaluating the impacts of transportation system on issues beyond promoting economic growth. The current multi-channel funding mechanism in China for transportation programs and projects may contribute to this mode of thinking.

Different levels of economic development in a given region of China usually determine the amount of available transportation funding. They may include direct investment of central and local governments, commercial bank credit, issuance of stocks and bonds, and the sale of infrastructure management to foreign companies. These complex transportation funding mechanisms are without consistent requirements and regulation, which prevents proper coordination among transportation programs and projects to reach regional goals.

The new transportation management model proposed above, implemented through regional transportation planning organizations, would be helpful in solving these problems.

The “Priorities in Transportation Finance in the 11th Five-Year Plan"[(4)], issued by Ministry of Communications of the People’s Republic of China, points out that in the "11th Five-year Plan" period funds from the central government fall far short of meeting transportation infrastructure needs. More than two trillion Chinese Yuan of transportation development funds are expected to be raised through the financial market. Therefore, the transportation industry should further improve the financial management system, explore and establish sustainable funding mechanisms and systematic and flexible methods for optimizing the resources allocation.

It requires a shared regional vision as the foundation for integration of financing, planning, construction, funding management, project evaluation and project coordination into an overall strategic plan. The proposed regional transportation planning organizations will provide a platform for such coordination and decision-making. Transportation and other closely related issues, such as population growth, economic development, air quality, land use, social justice, mobility and accessibility, and health and safety, must be integrated and coordinated through regional planning

organizations. Only in the context of systematic planning and coordination, will issues such as rezoning, relocation of residents and the compensation of environmental impact of transportation projects be realized in a fair and orderly manner.

Feng Zheng-Lin, the Deputy Communication Minister, pointed out recently,[5]"In the economic environment at home and abroad, constraints on resources are increasingly obvious. Developing a recycling economy and a resource-saving and environment-friendly society to achieve sustainable development will cause an increase in logistic and transportation development costs. It will require the transport industry to take more effective measures to increase revenue, cut costs, save energy, and promote the maximum social and economic benefits." A common perception is that the more roads built, the better the economy will be. It makes people concentrate on the investment of new infrastructure and overlook the overall efficiency of the transportation system, system management and maintenance.

In addition to capital investment and dominant transportation costs, transportation development also causes a lot of hidden and long-term costs. They include, for instance, the negative impact on the environment, quality of life and the social cost of traffic congestion and traffic accidents. Improper transportation development will not only fail to enhance the efficiency of the transportation system, but also have some long-term negative side effects, and thus, limit the return of large-scale transportation investments. The land used by transportation infrastructure in most cases is among non-recyclable and non-renewable resources (due to the high cost of renewing and reuse).

In the early stage of economic development, the costs of land and other resources are relatively low and quantities are large. At that time people are more interested in macroeconomic indicators such as the GDP growth and do not sufficiently understand the hidden costs which gradually emerge. However, at a certain stage of economic development, the amount of available land becomes limited, the cost increases, the need to improve the quality of life lifts, the awareness of environmental issues increased, and transportation systems near capacity. All these factors expose the hidden costs of transportation system.

The solution to this problem requires sustainable transportation development in the early stage which reduces negative long-range effects and enhances the coordination among various projects, thus yielding a production-addition effect. The comprehensive planning and assessment of transportation investment can not be accomplished by transportation, construction, environmental protection agencies or financial departments alone. It requires a regional transportation planning organization to coordinate and address these issues within a new transportation management model.

4. Regional Transportation Planning Procedures and Transportation policies

Generally speaking, a transportation plan is selected among different alternatives by planners, who use scientific methods to find the shortest travel time and the lowest logistics expenses. However, in the process of transportation decision-making, various approaches need to be explored to prevent future problems. Such considerations include: transportation system efficiency, mobility and accessibility, reliability, and

environmental impacts of traffic on communities. Since transportation is closely related to other social goals, transportation planners and traffic engineers are not the only participants in the transportation decision making process. Other participants may also include economists, environmental planners, logistics experts, managers and users of public transportation, freight managers, shippers and retailers, managers of cargo and passenger terminals and airports, community managers, and business district managers.

The transportation policy decision-making process, as illustrated in Figure 2, includes [6].

1. Research and assess current transportation system；
2. Establish a long-term shared-vision. Characteristics that should be considered to formulate a shared-vision include: projected population growth, projected economic changes, safety, maintenance of transportation facilities, and preserving the human and natural environment;
3. Forecast future transportation needs (air, bus, rail, roads, bicycle, pedestrian and water)；
4. Propose alternative policy approaches；
5. Establish a regional long-term transportation plan；
6. Establish a transportation improvement program；
7. Propose project and non-project solutions；
8. Monitor and evaluate

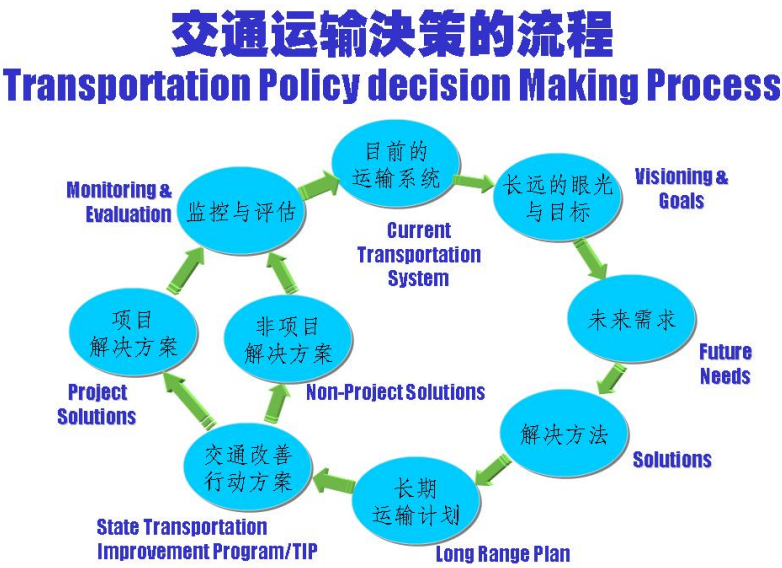


Figure 2 Transportation Policy Decision Making Process [6]

The management and the realization of the transportation decision-making process should be implemented by regional transportation planning organizations. In this regard, international experience will help China to formulate regional transportation planning organizations. In the United States, for example, the 1973 Federal-Aid Highway Act required that a Metropolitan Planning Organization (MPO) be designated for all urbanized areas with a population 50,000 or more.[7] MPO's are regional governmental associations that provide a platform for policy implementation, coordination, regional planning, and funding allocation to achieve the "3C" (continuous, comprehensive, and cooperative) transportation planning process.[8]

The 3C requirement is consistent with the policy proposed by the Communist Party of China. The Third Plenum of the 16th Central Committee of the Communist Party of China states "adhere to the people-centered concept of comprehensive, coordinated, sustainable development and promote the comprehensive development of the economy, society and people ". Having implemented the 3C transportation planning procedures for 30 years, the United States has a rather comprehensive planning process, which can be used as a reference for the development of Chinese cities. MPO's are the key factor of the implementation of the 3C planning process. Its five core functions include: [3]

1. Establish a setting**:** Establish and manage a fair and impartial setting for

effective regional decision-making in the metropolitan area.
2. Evaluate alternatives**:** Evaluate transportation alternatives, scaled to the size and complexity of the region, to the nature of its transportation issues, and to the realistically available options.
3. Maintain a Long-Range Transportation Plan (LRTP): Develop and update a long-range transportation plan for the metropolitan area covering a planning horizon of at least twenty years that fosters (1) mobility and access for people and goods, (2) efficient system performance and preservation, and (3) quality of life.
4. Develop a Transportation Improvement Program (TIP): Develop a program based on the long-range transportation plan and designed to serve the area's goals, using spending, regulating, operating, management, and financial tools.
5. Involve the public**:** Involve the general public and all the significantly affected sub-groups in the four essential functions listed above.

Transportation planning must be cooperative because no single agency bears full responsibility for the construction, operation, or maintenance of the entire transportation system in a particular region. Urban transportation planning is a process leading to transportation policies and decisions. The MPO will have the responsibility of actively seeking the participation of all relevant agencies, groups and individuals in the planning process.

While promoting public participation, regional transportation planning organizations also need to have an objective understanding of travel demand in the region. Travel demand forecasts analyze the impacts of various transportation policies, programs, and projects on regional travel. Forecasting tools include travel demand models and simulation models that evaluate transportation facilities and operations. Optimized transportation models can be produced for a simple intersection or for a complex urban transportation network. Simulation software is used to evaluate different alternatives of existing programs and possible results from the improvement program. Travel demand modeling and simulation techniques have provided the transportation decision-making process with an important, objective and scientific tool and have been widely used in urban planning.

The New York Best Practice Model (NYBPM), developed by New York Metropolitan Transportation Council, is the first operational tour-based travel demand model in the United States and has been widely applied in the New York region.[9] The NYBPM is a model which has the most complex transportation system, land use patterns and population structure. The model is designed to predict the detailed travel patterns of the region's diverse population, for all travel modes. Important innovations beyond conventional travel demand models are included in the NYBPM such as (1) using tours instead of trips as analysis unit; (2) incorporating a micro-simulation approach to generate forecasts of the discrete choices of individuals as opposed to probabilities that represent market segment shares found in conventional models; (3) Using the contemporary conceptual framework of the daily activity agenda of individuals while accounting for intra-household interactions between members and peoples' constraints on travel in terms of both time and space. The model has become an important

analytical tool for all major transportation investment studies in the New York Metropolitan Region.

5 Regional transportation planning and sustainable development

We will use the example of sustainable development to illustrate how the regional transportation planning organization could help to achieve the overarching goal of harmonious development of the society through the transportation planning process. Transport experts and government policy-makers are faced with the challenge of understanding and implementing a sustainable development, and striking a balance between transport development and sustainable development. This requires visionary decision-makers and properly trained transport professionals who understand how to assess a region's transportation investment program with system engineering methods in order to coordinate sustainable development.

Sustainable Development was a concept introduced in the 1980's. It was first mentioned in the1987 report, "Our Future",[10] released by the World Environment Development Committee that received broad consensus in the international community.

In 1994, the Chinese government promulgated the "China 21st Century Agenda -The White Paper on Population, Environment and Development of China in the 21st Century"[11]. A systematic exposition of the mutual relations among China's economy, society and environment, the paper set forth a strategic development framework for comprehensive, long-term and gradual implementation of sustainable development. Sustainable development refers to meeting current needs without compromising the ability of future generations to meet their demand. It entails the notion of an inseparable system with economic, social and resource development, as well as environmental protection that will accomplish economic development while protecting natural resources and environment, such as atmosphere, fresh water, oceans, land and forest, so that future generations will not be unduly impacted by decisions made today.

Air quality protection is an important component of Sustainable Development. How can we solve the traffic congestion and at the same time ensure that air pollution is not getting worse? Let's review how this is achieved in the New York metropolitan area as an example [12]. The New York Metropolitan Transportation Council has the responsibility of implementing the "Transportation Conformity" process to ensure that Federal funding and approval are given to transportation activities that are consistent with air quality goals. It ensures that these transportation activities do not worsen air quality or interfere with the "purpose" of the State Implementation Plan, which is to meet the national air quality standards. Meeting the national air quality standards often requires emissions reductions from mobile sources.

The transportation conformity process includes the following major components:

- interagency consultation among related agencies;
- regional emissions analysis, project level analysis;
- timely implementation of the transportation control measures (TCMs) in the state implementation plan; and

- certain planning requirements (e.g., fiscal constraint) required by Federal transportation legislation.

Key factors of Transportation Conformity used in the U.S. include:

- linkage of funding to transportation conformity;
- linkage between transportation conformity and transportation planning;
- valuation of options and the trade-offs between various emission reduction strategies; and
- linkage between Federal Clean Air Act and transportation planning, and challenges for transportation officials.

According to the U.S. Clean Air Act Amendment [13], transportation plans, programs, and projects cannot create new national air quality standard violations, increase the frequency or severity of existing national air quality standards violations, or delay attainment of the national air quality standards.

In addition, the U.S. Transportation Conformity requires:

(1) every three years each maintenance area must demonstrate "CONFORMITY" by showing that the mobile sources meet the State Implementation Plan budget for the current and future analysis years;

(2) the analysis must use the Environmental Protection Agency's emissions model and the best transportation data (Vehicle Miles Traveled, speeds, vehicle mix, etc.) available; and

(3) an area not able to demonstrate conformity goes into "LAPSE" and most federal money stops.

Each non-attainment or maintenance area must develop a plan (State Implementation Plan, SIP) detailing how they will maintain good air quality, including inspection/maintenance program, reformulated gas, and various transportation related measures. Based on Vehicle Miles Traveled forecasts, the SIP assigns allowable emission "budgets" (tons/day) for mobile sources for future years (2005, 2007, etc).

The transportation conformity process can help us to achieve the win-win goal of solving traffic congestion and improve air quality at the same time. We will also garner a very important by-product: a system planning process that promotes the efficient use of resources within a region.

6 Conclusion

Transportation is an important factor that impacts the harmonious urban development of the economy, society and environment. Transportation planning, management and development must focus on more than technical work. It should be a continuing, cooperative, and comprehensive (3C) planning process, including a shared-vision established by stakeholders.

City leaders, transportation officials and transportation professionals should establish goal-oriented strategic thinking. Guided by the goal of regional socio-economic

development, we should consider population growth, economic development, air quality, land use, social equity, mobility, health and security in strategically planning transportation development. We should also promote the development of regional transportation planning organizations, encourage the formation of a mechanism for public participation, and advance relevant legislation.

The Central Government's Ministries should play a vital role in this process. They can enhance the central and local laws and regulations to promote the standardization and normalization, provide guidance for regional organizations, provide training and information services for professionals in regional transportation planning organizations, and monitor regional planning and financial management activities.

A national training mechanism for planning, engineering, and management can guide transportation officials and professionals to transform their thinking process, so as to meet the above requirements and be able to face the challenges of traffic congestion problems that accompany rapid economic growth.

Transportation professionals need more expertise and resources to effectively manage and maintain the operation of a transport system, to produce forecasts and to determine the best solutions. Like all investment decisions, data and data analysis tools are the two basic elements of transportation planning and decision-making. This is what China's transportation ministries should focus on in order to create an efficient and sustainable transport system supportive of a harmonious society.

References:

1. Song, Hongling, Jian Zhou. Foster Sharing Concept and Build a Socialist Harmonious Society. Guangming Daily (宋洪玲周健人. 树立社会主义荣辱观与构建和谐社会. 光明日报.) 2006-12-11
2. The Communiqué of the Sixth Plenum of the 16th Central Committee of the Communist Party of China (中国共产党十六届六中全会公报.) 2006-10-11
3. USDOT, Federal Highway Administration. No date. Transportation Planning Process: Key Issues, Briefing Notebook, Washington, D.C. publication number FHWA-EP-03-041 (5/04)
4. PRC Ministry of Communication, Transport Finance "11th Five-Year Plan" Program Focus (中华人民共和国交通部. 交通财务"十一五"重点工作规划纲要.) 2006-03-24.
5. PRC Ministry of Communication, Finance Administration. Funding Assurance Mechanism Will Be Established for Transportation Industry (中华人民共和国交通部财务司. 交通行业将建立完善资金保障机制.) 2006-04-24.
6. USDOT, Federal Highway Administration. No date. A Citizen's Guide to Transportation Decision-making, Washington, D.C. publication number FHWA-EP-01-013 HEPH/3-01 (15M)E.
7. USDOT. Nov 1992. Urban Transportation Planning in the United States, An Historical Overview. Washington, D.C. publication number DOT-T-93-02. p.108
8. USDOT. Nov 1992. Urban Transportation Planning in the United States, An Historical Overview. Washington, D.C. publication number DOT-T-93-02. p.42
9. NY Metropolitan Transportation Council, NY Best Practice Model documents. New York, NY. http://www.nymtc.org/project/BPM/bpmindex.html
10. The World Commission on Environment and Development. Report of the World Commission on Environment and Development: Our Common Future. Geneva, Switzerland. 1987-6. http://www.un-documents.net/wced-ocf.htm
11. China 21st Century Agenda -The White Paper on Population, Environment and Development of

China in the 21st Century (中国 21 世纪议程管理中心. 中国 21 世纪议程——中国 21 世纪人口、环境与发展白皮书.) 1994-3-25
12. NY Metropolitan Transportation Council Air Quality Conformity documents. New York, NY. http://www.nymtc.org/project/conformity/conformity.html
13. U.S. Clean Air Act as amended in 1990. Washington, DC. http://www.epa.gov/air/oaq_caa.html/

TOWARDS A HUB-AND-SPOKE NETWORK: A STUDY ON THE CHINESE MAINLAND HUB AIRPORT PLANNING

Bo ZOU Huapu LU Qiang LIU

Institute of Transportation Engineering, Tsinghua University, Beijing, 100084, China; PH (86) 10-62772615; FAX (86) 10-62795339; email: zoubosoar@gmail.com; luhp@mail.tsinghua.edu.cn; liuqiang03@mails.thu.edu.cn

Abstract: The airlines reorganization will accelerate the switch of the Chinese mainland civil aviation system from a former point-to-point (PP) structure to a hub-and-spoke (HS) network. In such a context, the major efforts of our paper are put into identifying hub airports in mainland China. We first analyze the HS network's competitive advantages and find four desirable hub characteristics. A three-level method is developed to incorporate the required hub features and are applied to the hub airport planning. Roles of the hubs in the planning scheme are mainly in three categories. Further discussion is extended to network adjustment as well as its dynamic evolution.

Key Words: *HS network; PP structure; hub airport planning; Chinese mainland*

1 Introduction

The airline reorganization in mainland China completed in Oct. 2002 is a key step in the switch from a former point-to-point-dominant structure to a hub-and-spoke (HS) network. More than twenty domestic airlines have merged into three groups: Air China, China Eastern, and China Southern. In addition, Hainan airlines, a regional airline previously, became the fourth largest passenger air carrier by combining with several of its former rivals. Meanwhile, airlines have more freedom to reconfigure their routes. Most airlines choose to adopt the HS system, which is most popular among the top 20 airlines worldwide. Related research began in the mid 1980s in the western world; early researchers include Kanafani and Ghobrial (1985), Morrison and Winston (1986), and Oum and Tretheway (1987), who have identified this phenomenon and offered explanations for the increase in HS routing in the era of deregulation. To address the hub airport selection issue, the 'uncapacitated single allocation p-hub median problem' was first proposed by O'Kelly (1987), followed by T. Aykin (1990) and J. G. Klincewicz (1991). Further development of the models were made by J. F. Campbell (1996), and T. Ernst and M. Krishnamoorthy (1996 and 1998). These models are, however, usually mathematically difficult to solve; moreover, many approximations in them render the results sometimes far away from reality.

A relevant study of the HS network shows they are used comparatively less in mainland China, which could be attributed to the fact that the airline industry network in China still to some extent remains in a point-to-point (PP) form. Zhang (2000) discussed some prerequisites of the operation of a HS system. Jin (2001 and 2005) gave an overview of the country's passenger airline network development of the past two decades and constructed an ideal network organization model. In his research the hub airports were selected based on an index evaluation system in which weights were predetermined in a quite subjective manner. Bai (2006) provided a method including hub selection and main routes design. However, the results of the first part were obtained through a similar index evaluation process which only considers the local features, e.g. airport operations.

The recent emergence of complex network has shed some lights on the aviation network research. It has been found that HS structure is a prime example of a 'small-world' network (L. A. Schintler 2001). Involved in this area are primarily physicists, among whom R. Guimera and L. A. N. Amaral (2004) studied the world airport systems and concluded that it embodies the small world properties; Cai et al. and Ganesh Bagler did some statistical research of Chinese and U.S., and Indian airport systems respectively and found similar results. Their outcomes, however, are mainly focused on the network's mathematical and statistical properties. No practical suggestions are made to our knowledge, while applying the complex network theories to aviation network study.

To fill these gaps, an attempt is made in our paper to determine hub airports in mainland China in the context of an HS structure. The rest of the paper is organized as follows: in section 2 we analyze the competitive advantages of such structure by comparing it with a PP network. Hub characteristics are subsequently derived. After exploring the current Chinese mainland airport system and its problems in section 3, a three-level method incorporating the above desirable hub characteristics is given and applied to finding hub airports in section 4. Further implications are discussed in the following section which includes network adjustment and dynamic evolution. In the end, conclusions are given and further studies are directed.

2 Analysis of HS Aviation Network

2.1 Competitive Advantages of HS Aviation Network

A typical HS strategy is that all passengers departing from any non-hub origin and traveling to other cities in the network proceed first to the hub. Similarly, all passengers traveling to non-hub destinations arrive at those destinations from the hub. In order to understand the competitive advantages of the HS system over its PP counterpart, we start with a simple case which contains 9 airports. Under the PP structure, a minimum of 72 flights and 36 aircrafts are needed; if it is transformed into

an HS system with 3 hubs, however, a minimum of only 18 flights and 9 aircrafts are required. The reduction is significant, and therefore would considerably save operations costs for airlines.

In addition, cost advantages are also gained by creating the economy of density which manifests itself mainly in two parts. First, the consolidation of passengers from many routes on a single spoke increase the density of service along each spoke. Although hubbing need not change which airports an airline serves or the number of passengers it transports, hubbing does decrease the number of routes on which an airline operates its planes, therefore increasing density. Secondly, the potential increase in frequency along each spoke may increase demand and therefore, density also.

2.2 Hub Characteristics

The competitive advantages of an HS system are highly dependent on the hub cities. Not just any city can serve as a hub airport. There are four desirable characteristics of a hub city: airport performance, economic support, network connectivity, and central location. The high-frequency flights between hubs requires sufficient traffic volume, which holds the central portion of a hub airport's throughput, a most reliable and significant measure of the airport performance; on the other hand, traffic volume is also closely related to the economic situations of the city where an airport locates. A stronger local economy generates more commercial and business interactions between the city and other areas, which is an important source of air transportation for both passenger and cargo traffic. From the study of Jin and Wang (2005), the role of tourism attractiveness alone in influencing the air passenger traffic is still small; as a result this is not our primary concern in this paper. The connectivity form between an airport and its neighboring counterparts is also important. If many links already exist within the neighboring airports, regional travelers do not have to pass through the studied airport, therefore weakening the hub role of the studied airport. Finally, a central location is certainly preferred since it minimizes the distance traveled on each spoke. The importance of topological and physical travel distance may vary according to the actual stage length, e.g. in short-haul air travel, passengers seem to be more concerned with the number of transfers. Such consideration is reflected in the following section.

3 The Current Airport System in China and its Problems

Since the implementation of the Reform and Open-up policy in the 1970s, the civil aviation industry of mainland China has witnessed a rapid growth. By the end of 2004, there were 1035 domestic flights (those from/to Hong Kong/Macao are not included), and 1551186 km in terms of operational mileage. It ranked the 3rd among 188 IATA members. However, the airline network structure still largely remains a PP

form with too much emphasis on the trunk lines; whereas feeder lines do not receive enough of the attention they deserve (Bai 2006). Although this may be partially due to the difference of living standards and social-economic development between metropolises and smaller cities, and between eastern and western part of the country, the primary reasons originate in the air transportation system itself. Firstly, short-haul flights are unable to attract passengers from surface transportation such as rail and long-distance coach transportation, which can be attributed to the uncompetitive airport accessibility, much higher prices, and the low frequency of scheduled flights. This problem is crucial in that a strong support of short-haul feeder lines is a key point in the success of a real HS structure. Secondly, the phenomenon that the airport infrastructure level is higher than needed widely exists. Table 1 illustrates the class distribution of all the airports in mainland China, from which we find that the distribution is far less than proportional in forming an HS structure. This also illustrates that the air transportation network still maintains PP characteristics. Maintaining too many high-class airports only results in a waste of construction, operation, and maintenance expenses, thus affecting the global efficiency and competitiveness of the HS system.

Table 1 Airport class distribution in mainland China (year: 2004)

Airport Type	4E	4D	4C	3C
Number of Airports	25	34	52	26
Aircraft Size Max	B747	B767, B757, MD82	B737	Smaller than B737

Source: Statistical data on civil aviation of China 2005

With the continuity of national economic prosperity and the central government's policy preference, in the decades ahead, the above regional gaps would be greatly shortened and air traffic in the less-developed regions or cities would definitely experience a more remarkable increase. This, at the same time, provides the Chinese mainland airport system with an ideal opportunity to accelerate its pace towards an efficient HS network. To accomplish it, the first step is to configure the hub airports.

4 Hub Airport Planning

We consider the selection of a hub airport from local, regional, and global levels to incorporate the four desirable characteristics discussed in section 2. In the first stage we consider the local effects of airports by employing the cluster analysis method. Then complex network theories are applied to measure the extent of connectivity and central location of each airport. The final planning scheme is based on all the former analysis. Details are given as follows.

4.1 Local Effects

In the four characteristics, airport performance and economic support are

classified as the local factors. To more accurately represent their effects, the two factors are divided into four elements: passenger throughput, cargo throughput, local GDP, and non-agricultural population, which are defined as the input variables for cluster analysis. Relative data was collected for the year 2004. In the cluster analysis we use the following Minkowski distance to measure the similarity between each of the two airports,

$$d_{ij} = [\sum_{k=1}^{n} (x_{ik} - x_{jk})^2]^{\frac{1}{2}} \qquad (1)$$

where x_{ik} and x_{jk} represent the kth input variable of the ith and jth airports respectively, and d_{ij} denotes the similarity between the two airports. In gauging the similarity between two sample classes, the Ward method is adopted:

$$\begin{aligned} D_1 &= \sum_{x_i \in G_1} (x_i - \overline{x_1})^T (x_i - \overline{x_1}) \\ D_2 &= \sum_{x_j \in G_2} (x_j - \overline{x_2})^T (x_j - \overline{x_2}) \\ D_{1+2} &= \sum_{x_k \in G_1 \cup G_2} (x_k - \overline{x})^T (x_k - \overline{x}) \end{aligned} \qquad (2)$$

in which there are

$$\overline{x_1} = \frac{1}{n_1} \sum_{x_i \in G_1} x_i \qquad \overline{x_2} = \frac{1}{n_2} \sum_{x_j \in G_2} x_j \qquad \overline{x} = \frac{1}{n_3} \sum_{x_k \in G_1 \cup G_2} x_k$$

We define the distance between two sample classes is

$$D(G_1, G_2) = D_{1+2} - D_1 - D_2 \qquad (3)$$

More details can be found in Yang et al. (2004), and we opt to form 4, 6, and 8 clusters successively. Part of the results is given in table 2. Simply from the local effects, we find that the airports of Beijing, Shanghai, and Guangzhou are in a well distinct cluster, although the airport of Guangzhou is a little different from the other two in certain cases. The airports of Shenzhen, Tianjing, Chengdu, Chongqing, and Hangzhou are very similar. In the 3rd cluster the cities where airports are located are not necessarily provincial capitals; instead, many in it are economically dynamic cities in the eastern part of China.

4.2 Regional Effects

An airport's regional effects capture the effect of its regional network connectivity, which can be measured by the extent of passing through that airport for a regional trip. In the complex networks theories, a clustering coefficient is used to assess the regional effects of each vertex in the network. Since the Chinese airport network has proven to exhibit properties of a small world, one form of the complex networks, the clustering coefficient is also employed in our study. We first simplify the airport network by transforming it into an undirected graph G (V, E) with each vertex representing an airport and every edge corresponding to a route (two directions are regarded as one in the graph) between two airports. The definition of an airport's clustering coefficient is as

$$C_v = \frac{M}{C_n^2} \qquad (4)$$

where M is the number of the real connections within airport v's direct neighborhood Nv consisting of n airports, and C_n^2 is the total number of all possible connections among its neighboring airports. Obviously we have $C_v \in [0,1]$. If we define variable δ_l^v as

$$\delta_l^v = \begin{cases} 1 & \text{if edge l is linked with v} \\ 0 & \text{otherwise} \end{cases} \qquad (5)$$

M can be calculated as

$$M = \sum_{l \in E; x,y \in N_v} \delta_l^x \cdot \delta_l^y \qquad (6)$$

Figure 1 shows the clustering coefficients of the first 27 airports with values in an increasing order. It should be noted that for the airports who have only one direct link, its clustering coefficient is defined as zero. Such airports are excluded from our consideration. For the rest of the airports, evidently a smaller clustering coefficient means a higher regional importance (effects) because in such a case the communication among the neighboring airports is more likely to pass through the studied airport first. Therefore in figure 1 we use the term 1-clustering coefficient to

represent the regional effect of the airports. The dominance of Beijing, Shanghai, Guangzhou, Shenzhen, and Kunming is significant. The reason for including Kunming is that many Provincial tourism flights depart exclusively from Kunming; in other words, the airport network in Yunan already exhibits some features of an HS structure. Similar phenomenon can be found in Xinjiang, where Urumchi plays an equivalent role (Jin and Wang 2005).

No. of Clusters	1st Cluster	2nd Cluster	3rd Cluster
4	Beijing, Shanghai*	Guangzhou, Shenzhen	Chengdu, Kunming, Xi'an, Hangzhou, Qingdao, Dalian, Nanjing, Chongqing, Wuhan, Shenyang, Harbin, Changsha, Fuzhou, Zhengzhou, Wenzhou, Jinan
6	Beijing, Shanghai, Guangzhou	Shenzhen, Tianjing, Chengdu, Chongqing, Hangzhou	Fuzhou, Quanzhou, Shijiazhuang, Zhengzhou, Harbin, Changchun, Dalian, Shenyang, Jinan, Qingdao, Wuhan, Nanjing, Yantai, Ningbo, Wenzhou
8	Beijing, Shanghai#, Guangzhou	Shenzhen, Tianjing, Chengdu, Chongqing, Hangzhou	Fuzhou, Quanzhou, Shijiazhuang, Zhengzhou, Nanjing, Harbin, Wuhan, Changchun, Dalian, Shenyang, Jinan, Qingdao, Yantai, Ningbo, Wenzhou

Table 2 Cluster Analysis Results

*the passenger and cargo throughput of Shanghai are the sum of the two airports (Hongqiao and Pudong). Such consideration is also applied to the 6 and 8 cases.
in the 8 case, Shanghai is classified in an individual cluster which is near to {Beijing, Guangzhou}. We regard the two as one cluster for the sake of simplicity.

4.3 Global Effects

In the four characteristics stated in section 2, central location is used to measure the global effects of an airport. It can be reflected by the number of the shortest routes in which the studied airport is either the origin/destination, or an intermediate stop. In effect, the larger the number, the more important the role the airport plays in shortening travel distance over the entire network. Such a concept in complex theory is defined as 'betweenness, and is calculated as the fraction of shortest paths between vertexes pairs that pass through the vertex of interest. Let S_{ij} denote the set of the

shortest routes and the airport v's betweenness is defined as

$$B_v = \sum_{i,j \in E} \frac{\sum_{l \in S_{ij}} \delta_l^v}{\left| S_{ij} \right|} \qquad (7)$$

We employ the Floyd algorithm (Wang 1993) to find the shortest route between any two airports in the network. Although in many instances the high speed of air transportation causes passenger more concern about transfer than actual stage length, for the long-haul trips the impact of actual distance on passengers' choice is more perceptible. As a result, we reconstruct a weighted graph in which the weight of each edge represents the physical distance between two airports. Each airport's 'betweenness' is calculated both in the weighted and unweighted graphs. The major results are shown in figure 1.

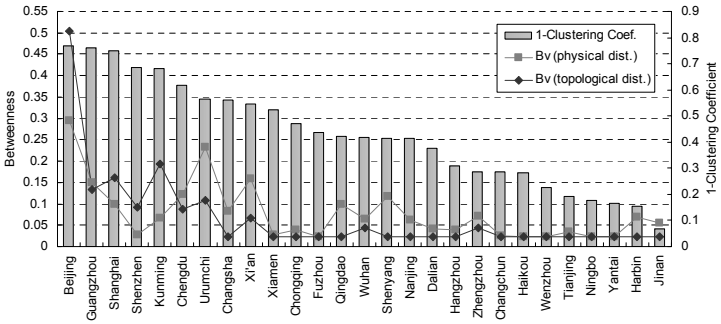


Figure 1 Clustering coefficients and betweenness of the selected airports

Clearly, Beijing Capital Airport owns the highest values of 'betweenness' with either topological or physical distance. In the topological structure, the 'betweenness' of Beijing is followed by Kunming, Shanghai, Guangzhou, Urumchi, Shenzhen, Chengdu, Xi'an, and Wuhan, successively. While the actual distances are taken into account, the decreasing order becomes Urumchi, Xi'an, Guangzhou, Chengdu, Shenyang, Shanghai, Qingdao, Harbin, and Wuhan. The comparatively high degree of 'betweenness' of Urumchi, Xi'an, Qingdao, and Shenyang in the latter case might be attributed to their geographic locations. For example, if passengers from Chongqing fly to Hohhot, they have to make one transfer since no direct flight exists. There are several transfer choices such as Beijing or Xi'an. Their positions are 'equal' providing that only topological connections are considered. However, the actual trip length via Xi'an is much shorter because the two segments of the travel are almost in a line.

Similar circumstances can be found in trips from southeast cities to the northeast part of the country where there are few direct air transportation services. Geographically, Qingdao provides an ideal location for one-stop transfers.

4.4 Planning Scheme of National Hub Airports

If we comprehensively take into account the local, regional, and global effects of all the airports, we find the airports in Beijing, Shanghai, and Guangzhou are obviously predominant of all the airports. Shenzhen Airport is next. What follows are Chengdu, Tianjing, Chongqing, and Hangzhou in terms of local effects, Kunming, Chengdu, Urumchi, Changsha, and Xi'an for the regional effects, and Urumchi, Xi'an, Chengdu, Kunming, Shenyang, Qingdao, and Wuhan if we consider global effects. The airport of Chengdu is somewhat less influential than those in Beijing, Shanghai, Guangzhou, and Shenzhen, yet its global effects are much stronger than any of the rest of the airports. Although Tianjing, Chongqing, and Hangzhou show strong local effects largely due to their economic support, they are not included as primary hubs in that in their neighborhoods, primary hubs already exist which are Beijing, Chengdu, and Shanghai. For the same reason the airport of Shenzhen is not identified as a primary hub. Kunming, Urumchi, and Xi'an show their regional and global competence as well; however, their local effects are not as strong as the other airports and thus might rather serve as a hub whose main function is to provide a place for transfers. Furthermore, the airports of Shenyang, Qingdao, and Wuhan are similar in terms of all the three aspects, by holding comprehensively secondary positions; in addition, their geographical locations are rather independent of the above airports: Shenyang is in the northeast; Qingdao is located in the centre of the northern coastal area; the distances between Wuhan and its nearest airports mentioned above, e.g. Shanghai, Guangzhou, Xi'an, and Chengdu, almost equal to each other and are approximately 800-1000 km.

Our above considerations can be summarized as a preliminary planning scheme of the hub airports on mainland China. These are shown in figure 2. The four national primary airports are: Beijing Capital Airport, Shanghai Airport (Hongqiao and Pudong are regarded as one), Guangzhou Baiyu Airport, and Chengdu Shuangliu Airport, who are responsible for the major transfer tasks in North, East, South, and West China. Secondary airports include Urumchi, Xi'an, Chongqing, Kunming, Wuhan, Tianjing, Shenyang, Qingdao, Hangzhou, and Shenzhen. In a geographic sense they are evenly distributed; however, their roles are not exactly the same. Tianjing, Chongqing, Hangzhou, and Shenzhen serve as auxiliary hub airports of their neighboring primary hubs respectively. In addition to accommodating 'original and terminal' traffic, their function could be extended to sharing part of the passenger and freight transfer traffic to avoid potential congestion at their nearby primary hubs. For the other airports, Urumchi, Xi'an, Shenyang, Wuhan, Qingdao, Kunming, they are

expected to act as rather independent regional hubs.

5 Further Discussions

5.1 Network Adjustment

The HS structure departs from centralizing traffic and promoting efficiency of aircrafts use. Those objectives are realized by increasing flights frequency between hubs and strong support of feeder lines. These would, however, very possibly result in potential congestion especially during peak hours at primary hub airports. In addition to adopting some operational strategies such as 'arrival-departure mix' to form flights waves in order to alleviate this problem (Jin and Wang 2005), it is important for planners and decision-makers to note that the starting point of implementing a HS network is to enhance global efficiency by evading excessive direct links among medium- and small-size cities or towns, rather than to deny the coexistence of PP structure particularly between large cities or province capitals where traffic volume is substantial. Even when the origin and destination are not hubs, city-pair flights are still encouraged. On the other hand, the planning of four secondary hub airports (Tianjing, Hangzhou, Shenzhen, and Chongqing) is, in effect, also an effort to mitigate potential congestion at primary hubs by switching part of the transfer traffic to its nearby partners.

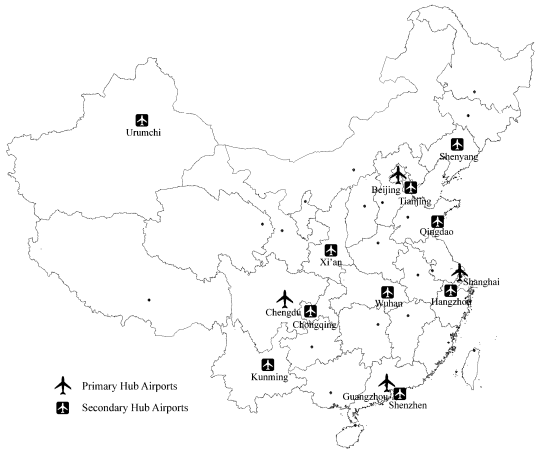


Figure 2 Hub airports in mainland China

5.2 Dynamic Evolution

With network development, hub airports will continually evolve. New hubs will appear and old ones will gradually lose their influence, or primary hubs will be

replaced by secondary ones due to economic, geographic, and political changes. An interesting issue is the role of Guangzhou Baiyun Airport. First, the nearby Shenzhen airport bears a closer similarity to Baiyun Airport than any of its 'parallel' airports to the rest three primary hub airports in terms of airport performance, economic support, location, and airline connectivity, which poses considerable challenges to Baiyun Airport. Secondly, for international transfers, Hong Kong is definitely better than Guangzhou because there are many more international scheduled flights there; on the other hand, many mainland cities have connected to Hong Kong by direct flights, the number of which is still rapidly increasing. As to domestic hub-and-spoke air travel and regional transfer service within the mainland, nevertheless, Hong Kong can not substitute for Guangzhou because of the administrative restraints at least in the short term. In the long run this trend would be very possible if restraints no longer exist and Hong Kong and the mainland are more closely integrated. Furthermore, the realization of 'three exchanges' associated with closer social and economic interactions between the two sides of the Taiwan Straits in the long run would definitely extend the role of Taipei Taoyuan Airport as a new hub in the southeast China, in addition to its function as the air transportation center of the island.

6 Conclusions

Different from previous studies which are based on either sophisticated mathematical models or simple yet quite subjective index assessment, our research gives a new hub airport planning process. We first find airport performance, economic support, network connectivity, and central location are four desirable hub characteristics. These characteristics have been incorporated into local, regional, and global effects by applying cluster analysis and complex network theories to study each airport in mainland China. The airports in Beijing, Shanghai, Guangzhou, and Chengdu are proposed as national primary hubs to be responsible for national major transfer tasks. Tianjing, Hangzhou, Chongqing, Shenzhen, Shenyang, Qingdao, Xi'an, Urumchi, Wuhan, and Kuming are defined as secondary hubs. Their role is, in general, to accommodate local transfer traffic in addition to foster city-pair flights between itself and other major cities. Tianjing, Hangzhou, Chongqing, and Shenzhen can furthermore serve as auxiliary airports of primary hubs mainly owing to their geographic positions. Part of the transfer traffic can be switched from the primary hubs to these airports, which will help to avoid potential congestion problems at primary hub airports. Along the route where traffic volume is significant, city-pair flights are still encouraged. Hub airports are never static. New hubs can replace old ones; it is also possible for secondary hubs to substitute for primary ones. In particular, Guangzhou Baiyun Airport faces considerable challenges from that of Shenzhen. For foreign/domestic transfers, Hong Kong is the better place, although for domestic and regional transfers, Hong Kong airport is currently powerless due to administrative barriers. With the realization of 'three exchanges', Taipei Taoyuan Airport is very

likely to become a new hub of southeast China in the long run.

Some aspects of our studies deserve further work. In the planning process, the clustering coefficient is used to measure an airport's regional effects. However, the two concepts sometimes do not totally overlap. It might be more precise if certain physical distance limitations were introduced to judge whether the airport which has a direct link with the studied one really belongs to its neighborhood. Moreover, how to add international connections in the consideration would also be an interesting area of future research.

7 Acknowledgement

This paper is supported by the key project of Chinese Ministry of Education (No. 104233). The authors would like to extend their gratitude to ZHOU Qian and PENG Hu for their contribution to improving our work. We also thank the anonymous referees for their constructive and helpful comments.

References

Aykin T. (1990) "On a quadratic integer program for the location of interacting hub facilities." European Journal of Operational Research, Vol. 46, No. 3, 409-411.

Bagler Ganesh (2004) "Analysis of the Airport Network of India as a Complex Weighted Network." arXiv: cond-mat/0409773

Bai, M.G. et al. (2006) "Design and Application of Hub and Spoke Network." System Engineering, Vol. 24, No. 5, 29-34.

Campbell, J.F. (1996) "Hub Location and the p-Hub Median Problem." Operation Research 44, 923-935.

Chi, L.P. et al. (2003) "Structural Properties of US Flight Network." Chinese Physics Letters, Vol. 20, No. 8, 1393-1396.

Ernst, A.T. (1996) "Efficient Algorithm for the uncapacitated Single Allocation p-hub Median Problem." Location Science, 4, 139-154.

Ernst, A.T., Krishnamoorthy M. (1998) "Exact and Exact Algorithms for the uncapacititated Multiple Allocation p-hub Median Problem." European Journal of Operational Research, 104, 100-112.

Guimera, R., Amaral L.A.N. (2004) "Modeling the World-wide Airport Network." European Physical Journal B, 38:381-385

Jin, F.J., (2001) "A Study on Network of Domestic Air Passenger Flow in China." Geographical Research, Vol. 20, No. 1, 31-39.

Jin, F.J., Wang, C.J. (2005) "Hub-and-Spoke System and China Aviation Network Organization." Geographical Research, Vol. 24, No. 5, 774-784.

Kanafani, Adib, Atef A. Ghobrial (1985) "Airline Hubbing – Some Implication s for Airport Economics." Transportation Research A (general), Vol. 19, 15-27.

Klincewicz, J.G. (1991) “Heuristics for the p-hub location problem.” European Journal of Operational Research, Vol. 53, No.1, 25-37.

Li, W., Cai, X. (2004) “Statistical Analysis of Airport Network of China.” Physical Review E 69, 046106

Morrison, Steven, Clifford Winston (1986) “The Economic Effects of Airline Deregulation.” Washington, D.C.: The Brookings Institution.

O’Kelly, M.E. (1987) “A Quadratic Integer Program for the Location of Interacting Hub Facilities.” Journal of Operational Research, E 35, 225-239.

Oum, Tae H., Michael W. Tretheway (1987) “Airline Hub and Spoke Systems.” working paper, Faculty of Commerce, University of British Columbia, Vancouver.

Schintler, L. A., Kulkarni Rajendra (2001) “Measuring Small-World Connectivity in Air Transportation Networks.” Proceedings of the 5th International Conference on Technology, Policy and Innovation: Critical Infrastructures.

Wang, Y.X. (1993) “Operations Research – Planning Theories and Network.” Tsinghua University Press, Beijing.

Yang, H. et al. (2004) “Mathematical Statistics.” Higher Education Press, Beijing.

Zhang, H.M. (2000) “Prerequisite of Hub-Spoke System.” International Aviation, 2000.6, 37-39.

Traffic Distribution Forecast of Urban External Traffic Hub Based on Uncertainty

Shunying ZHU[1], Juxiang GUAN[2], Wuxiong PENG[2], Yuan XU[2]

[1] Professor, Dept. of Traffic Engineering, Wuhan University of Technology, 1040 Heping AVE, Wuchang, Wuhan, Hubei, China. 430063. E-mail: zhusy2001@163.com.
[2] Master Student, Dept. of Traffic Engineering, Wuhan University of Technology, 1040 Heping AVE, Wuchang, Wuhan, Hubei, China. 430063. E-mail: gjx_wh@163.com

Abstract: The urban external traffic hub is the important node of the correlation between urban transportation and regional transportation, the converging-dispersing traffic in a traffic hub gradually increased along with the accelerating regional integrated process and the strengthening of the correlation between city and countryside. So its influence on urban traffic couldn't be ignored. In view of some uncertainty of traffic resistance to road networks, trip production-attraction of traffic zones and trip converging- dispersing of traffic hubs in the future, this paper discusses a traffic hub distribution forecast under such uncertain conditions without an actual OD and establishes a fuzzy linear programming model which takes the least transport fee (time)as objective function, traffic hub's converging-dispersing and traffic zone's trip production-attraction as restrictions. The solution algorithms are also proposed. Compared with certain distribution model, it has been found that this model rather accords with the fact. In addition, the author discusses how to calculate the specific relations between traffic zones and traffic hubs and the resolution of different extents of uncertainty problem.

Key Words: *traffic hub, traffic distribution, fuzzy linear programming, integrate transportation*

1 Background

Traffic hub comes into being where there are two or more transit lines. It is the comprehensive facility of transport organization and management, change and transfer, load and unload, multi-through transport, information circulate and assistant service. In a city, besides the airport, any other traffic hub commonly has two or more. They are distributed in different places and have different functions. Among them competition often exists. Considering that its effect on urban road network and traffic flow couldn't be ignored, different traffic hubs have different effects on introduced traffic and traffic assignment, improper organization could cause partial traffic congestion. So a reasonable traffic distribution of traffic hubs plays a very important role in reducing transport fee (time) (initial traffic resistance below), improving the level of service, heightening the competitive ability of enterprise and planning and managing urban road network in a reasonable way. Until now, the research on traffic hub has focused mostly on the layout, location choice and passenger transfer of traffic

hub (Qian Jia, 2006; A-Long Hu, 2006; Han.Peng 2006), but the research about the distribution of traffic hub is extremely lacking.

In practice, the commonly distribution forecast methods are growth factor method, gravity model, intervening opportunities model and system equilibrium model (Ho, H.W. Wong, S.C., Loo, Becky P.Y. ,2006; Thompson, Russell , Young, William.,1988; Chung and Albert Yuen Tai ,2005; Nanba, Shinobu, Konishi, Satoshi; Nomoto, Shinichi ,2004; Chim, Tat Wing , Yeung, Kwan L.,2004). Growth factor method includes uniform growth factor, average growth factor, Detroit model, factor model and so on. These models require actual full OD, and haven't considered the effects of changing network to trips. In the analysis of new traffic modes, new routes and new price policy, these methods appeared helpless, but they are all simple to calculate, so they could be applied in rough traffic distribution in the near future. Gravity model need part actual OD to demarcate the parameters in it. Compared with the growth factor method, its precision is proved as a result of important traffic resistance—it has considered the effects of changing network to trips, but that yet can't be applied without an actual OD. The system equilibrium model is similar to a gravity model, obtaining the result through solving some nonlinear equations. But in a megalopolis, the model is difficult to solve due to excessive traffic zones. The intervening opportunities model has a better modeling basis than the gravity model, but its parameter is hard to demarcate, so it is not easy to be applied. The models introduced above, except system equilibrium model, all need actual OD; and except growth factor method, all need traffic resistance; each model yet needs trip production- attraction, and regards traffic resistance, trip productions and attractions as certain number.

Due to the randomness and uncertainty in traffic conditions: travel speed on roads, delay at intersections, available capacity and traffic accidents and so on, it is unreasonable to use certain number to express them. Because the society, the economy and the traffic policy in the future all have great uncertainty, apparently, the forecast of trip production-attraction will have uncertainty. In the process of regional integrate development, there are competition and cooperation among towns, the changes of relationship between the satellite town and the central city, sometimes could be strengthened and sometimes be weakened, which leads to the great uncertainty in converging-dispersing (converging initial attraction and dispersing initial production below) of central traffic hubs. Therefore, traffic distribution forecast models, which take the certain math theory as the basis, are difficult to consider with such uncertainty. A new model of that, should consider the uncertainty of the traffic resistance, trip productions-attractions, taking the uncertainty math theory as the basis. The fuzzy math is a member of uncertainty math, and it has already been applied in many traffic fields, such as fuzzy shortest path (Moazeni and Somayeh,2006; Xiao-Yu Ji,Iwamura, Kakuzo, Shao Zhen,2007; Yin-Zhen Li, Gen, Mitsuo,Ida Kenichi,1996; Okada, Shinkoh Gen, Mitsuo ,1994), fuzzy comprehensive assessment (Dong-Mei

Zhao, Jiang Liao, Sheng Li, Minqiang Li,2002; Kai Wang, Qi-Xiang Yu , Ya-Qin Wu, Yan-Fei Chen, Yi-Bo Wang,2002; Hong-Yan Xiang, Shun-Ying Zhu, Hong Wang, Xin-Ping Yan,2005).

Using the fuzzy programming theory, it discusses the traffic distribution of urban external traffic hub under the uncertainty conditions analyzed above without an actual OD. The rest of this paper is organized as follows: Section 2 presents the characteristics of the traffic hub. The model and its solution algorithms are discussed in section 3. Section 4 presents the numerical example. Section 5 contains the results and analysis. Discussions are presented and some recommendations for future research are provided in the final section.

2 The characteristics of traffic hub

As the bridge and ligament of a city's external traffic, the traffic hub plays a very important role in the shaping and developing of a city, and has an intimate relationship with the urban traffic system. The traffic hub is an important component of the transportation system of a country or a region. It is the junction of neighboring routes of different traffic modes' network. It is a whole which is composed of many settled and mobile facilities, and undertakes through traffic, transferring operation, local traffic and external traffic. The passenger and freight volume of the four external traffic hubs, including highway, rail, water transport and airlift, have two types: intermodal transportation and converging-dispersing transportation. Because intermodal transportation commonly serves the outskirt area of a city, and has little effect on urban traffic, we ignore this and mainly discuss converging-dispersing transportation which has had great effect on urban traffic.

If there is more than one traffic hub which provides converging-dispersing transportation, then the competition among them exists And the distribution volume of every hub is influenced by traffic resistance, size, location, and so on. As a monopolistic traffic hub which provides service to a single corporation, its distribution volume depends on the locality between the hub and the corporation, which is easy. This paper discusses mainly the trip production-attraction distribution among all traffic zones and multi-traffic hubs, the sketch map（see Fig.1）.

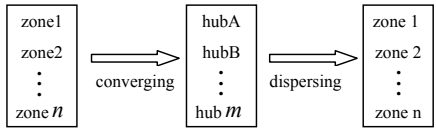


Figure 1. converging-dispersing transportation between all traffic zones and multi-traffic hubs

3 Traffic distribution forecast model of traffic hubs and its solution algorithms

3.1 Basis of the model

In order to simplify the model, traffic distribution among multi-hubs could be forecasted according to the traffic mode, the passenger transport and the freight transport. First, forecast the traffic distribution of highway, rail and water transport respectively; second, combine the passenger and freight volume of each traffic mode together, then get the passenger and freight volume distribution of the traffic hub; in the end, change the passenger and freight volume into an identical unit, and add them together, the traffic distribution of traffic hub will be obtained. The most important question is how to get the distribution of passenger and freight volume of each traffic mode, the following text will mainly analyze how to build a model for it. When traffic hub performs a converging function, the traffic zone is O spot; the traffic hub is D spot. When dispersing, the traffic hub is O spot; the traffic zone is D spot.

Suppose that there are m traffic hubs providing service to n traffic zones. The basic assumptions are made as follow: assumption (1): the trip production-attraction of n traffic zones, the traffic attraction-production of m traffic hubs and the traffic resistance of each traffic zone to each hub are all already known and they are all fuzzy numbers; assumption (2): the distribution volume between traffic zone and the hub is inversely proportional to the traffic resistance between them.; assumption (3): the sum of production-attraction of all traffic zones is equal to the sum of attraction-production of all traffic hubs. The imbalance between them will be explained in the following text.

A variable such as $\widetilde{A}$ stands for the fuzzy number, while A is a certain number. Traffic resistance, trip production-attraction of traffic zones and trip attraction-production of traffic hubs are a fuzzy number, and suppose they are a fuzzy triangle number.

3.2 Traffic distribution model of converging

The converging function of the traffic hub is to converge all traffic zone's trip to all traffic hubs, just as in the left of Fig.1. Take the minimization of transport fee (time) as an objective function, and take the traffic hub's trip attraction-production and traffic zone's trip generation-attraction as restrictions, the model could be built up as follows:

$$\min\ z = \sum_{i=1}^{n} \sum_{j=1}^{m} \widetilde{c}_{ij} \widetilde{X}_{ij}$$

$$s.t \quad \begin{cases} \sum_{j=1}^{m} \widetilde{X}_{ij} = \widetilde{a}_i & (i = 1,2,\Lambda\ ,n) \\ \sum_{i=1}^{n} \widetilde{X}_{ij} = \widetilde{b}_j & (j = 1,2,\Lambda\ ,m) \\ \widetilde{X}_{ij} \geq 0 & (i = 1,2,\Lambda\ ,n; j = 1,2,\Lambda\ ,m) \end{cases} \qquad (1)$$

The model is a typical transportation problem, but it has some differences compared with common problems. Because there are many fuzzy numbers in it, it is difficult to be solved by table algorithm. Here Matlab will be used to deal with the fuzzy numbers, and then the linear program used to solve it. The procedure will be described in the following thesis. The step-by –step procedure is described as follow.

1 Set the max cycle-index k , set iteration k=1.

2 Use random function unifrnd（min, max） which will generate arrays of random numbers between "min" and "max" to generate $\widetilde{c}_{ij}$ 、 $\widetilde{a}_i$ 、 $\widetilde{b}_j$ $(i = 1,2,\Lambda\ ,n; j = 1,2,\Lambda\ ,m)$. For an instance, as fuzzy number （4，6，8）, just take order unifrnd（4, 8）.

3 Judge the balance between supply and demand to see whether it satisfy the equation $\sum_{i=1}^{n} \widetilde{a}_i = \sum_{j=1}^{m} \widetilde{b}_j$, if not, see measures described in next chapter .

4 Bring the random numbers gained in step 2 into（1）, then get the normal linear program, solve it, export and save the results of X_{ij} in a three dimension array $A(n,m,k)$. Set k=k+1 and go to step 2, repeat the upward steps, until reaching the max cycle-index.

5 Seek $\min_k A(i,j,k)$ and $\max_k A(i,j,k)$, $(i = 1,2,\Lambda\ ,n; j = 1,2,\Lambda\ ,m)$, export the finally results,

$$\widetilde{X}_{ij} = （\min_k A(i,j,k)，\ \frac{\min_k A(i,j,k) + \max_k A(i,j,k)}{2}，\ \max_k A(i,j,k)）.$$

3.3 Traffic distribution model of dispersing

The dispersing function of the traffic hub is dispersing all traffic hub trips to all traffic zones, just as in the right of fig.1. Take the traffic hubs as production locus and

traffic zones as attraction locus, then imitate the method of building up the traffic distribution model of converging to obtain a traffic distribution model of dispersing.

3.4 The imbalance of supply and demand among traffic zones and hubs

The method mentioned above has a premise of the balance of supply and demand between traffic zones and hubs, namely $\sum_{i=1}^{n} \widetilde{a}_i = \sum_{j=1}^{m} \widetilde{b}_j$. But in practice, sometimes the imbalance condition might be met. This condition has two types. First, $\sum_{i=1}^{n} \widetilde{a}_i > \sum_{j=1}^{m} \widetilde{b}_j$ in which supply is more than demand, just add a hypothetical traffic hub $j = m+1$, and the trip attraction-production of this hub is $\sum_{i=1}^{n} \widetilde{a}_i - \sum_{j=1}^{m} \widetilde{b}_j$, the traffic resistance from each traffic zone to this hub is 0, $\widetilde{c}_{i,m+1} = 0$. The second type is $\sum_{i=1}^{n} \widetilde{a}_i < \sum_{j=1}^{m} \widetilde{b}_j$, namely, demand is more than supply. Similarly, add a hypothetical traffic zone $i = n+1$, and the trip production-attraction of this traffic zone is $\sum_{j=1}^{m} \widetilde{b}_j - \sum_{i=1}^{n} \widetilde{a}_i$, the traffic resistance from this traffic zone to each hub is 0, $\widetilde{c}_{n+1,j} = 0$. Then the balance condition will be obtained.

4 Numerical example

Shun-Ying Zhu, Hong Wang, Shou-Quan Luo, Xin-Ping Yan (2006) have analyzed the traffic distribution of traffic hubs based on certain numbers, the data taken from the numerical example. There are 10 traffic hubs and 9 traffic zones in a city. Then modify the certain number into a fuzzy number to get the trip production of each traffic zone. The trip attraction of each traffic hub and the traffic resistance between each traffic zone and each traffic hub, see in tab.1. How to modify that: make a fluctuation from 1000 to 5000 person per day of the original trip production and attraction; meanwhile, make a fluctuation from 1to 5 minute to the original traffic resistance, the detailed fluctuation of them made according to the original number. So build up the model as（1）。

Utilize MatLab to make a programme to solve the model, the results shown in Tab.2.

5 Results and analysis

The model proposed in this paper has studied the traffic distribution of traffic hubs under the uncertain conditions without the actual OD. Shun-Ying Zhu, Hong Wang, Shou-Quan Luo, Xin-Ping Yan (2006) have built up the certain model, then make a contrast between its results of numeric example (see in Tab.3) and the results of this uncertain model (see in Tab.2).

Contrast Tab.2 with Tab.3, it's evident to discover that item 0 in table 3 is much more than that in Tab 2. For example, the trip production of zone 5 has been all distributed to hub A in Tab.3, but in Tab.2, it not only distributed to the hub A but also hub F, G, J. The main reason is that the route between zone 5 and hub A is the shortest. This result is similar with the all-or-nothing assignment result, which will make the distribution asymmetrically, and all trip production of a traffic zone was almost assigned to a traffic hub which has the shortest path to this traffic zone. The distribution volume between zone 1 and hub G is 20,000 persons per day, while the trip attraction of hub G in tab.3 is 20,000 persons per day. It indicates that there is no traffic distribute to hub G from other zones, which is often not the truth. It seems more rational in Tab.2, not only from zone 1, but also zone 2, 5, 7, 8. The analysis above shows that when introducing fuzzy number to denote the uncertainty, in other words, considering the randomness and uncertainty of traffic conditions, travel speed on road, delay at intersection, available capacity and traffic accidents and so on, and allowing a little fluctuation to trip production and attraction, the distribution result will be more symmetrical and more rational. And after introducing fuzzy numbers, the distribution volume was a fuzzy number too, using fuzzy triangle number to express itself, as supposed in assumption (2). We can conclude that when the real traffic resistance is low, take the up bound of the fuzzy triangle number, and when it is high, take the lower bound of it.

Analyze data in Tab.2 carefully, it is not hard to discover that the great mass of trip production was concentrated on a few traffic hubs. Due to the limited circumstance, and being unable to carry an investigation on the spot, the traffic resistance of some roads maybe changed heavily, but in this paper, it is just done through a uniform fluctuation of the original data, causing the traffic resistance not to agree with the reality. Therefore, if we can choose the appropriate traffic resistance according to the fact, the result could be more perfect.

Table 1. transport fee, trip production of traffic zones and trip attraction of traffic hubs

time \ hub / zone	A	B	C	D	E	F	G	H	I	J	a_i (ten thousand /day)
1	(15,17.3,19)	(7,8.6,9)	(11,12.8,14)	(9,10.2,11)	(8,9.1,10)	(13,14.0,15)	(8,8.9,11)	(14,15.4,17)	(17,18.3,20)	(16,17.3,19)	(4.2,4.5,4.8)
2	(4,4.6,5)	(9,9.7,11)	(10,11.2,12)	(18,20.1,22)	(19,21.8,23)	(8,8.4,9)	(4,4.5,5)	(17,18.4,20)	(5,5.6,6)	(4,4.7,5)	(1,1.1,1.2)
3	(19,21.6,23)	(13,14.5,16)	(5,5.7,7)	(3,3.2,4)	(15,16.2,17)	(19,21.0,23)	(11,12.4,14)	(7.5,8.3,9)	(15,16.5,18)	(19,21.6,23)	(4.8,5.2,5.6)
4	(20,22.3,24)	(18,20.1,22)	(30,32.7,35)	(29,31.8,33)	(24,26.7,28)	(13,14.1,15)	(24,26.0,28)	(35,36.9,39)	(26,27.9,30)	(16,17.6,19)	(0.7,0.8,0.9)
5	(26,28.8,31)	(33,35.5,38)	(35,37.0,39)	(43,45.9,49)	(45,47.6,49)	(32,34.2,36)	(28,30.3,32)	(41,44.3,47)	(29,31.5,34)	(28,30.5,33)	(0.35,0.4,0.45)
6	(18,19.3,21)	(33,33.6,38)	(33,35.2,37)	(41,44.0,47)	(41,43.6,46)	(25,27.3,29)	(26,28.4,30)	(39,42.4,45)	(27,29.6,32)	(22,24.0,26)	(1.6,1.7,1.8)
7	(35,38.5,41)	(40,43.5,46)	(33,35.6,38)	(42,44.5,47)	(51,55.7,60)	(39,42.3,45)	(35,38.4,41)	(33,34.3,37)	(26,28.3,30)	(36,38.6,41)	(1.9,2.1,2.3)
8	(46,48.2,51)	(33,35.6,37)	(40,42.6,45)	(32,33.7,35)	(23,24.8,26)	(38,40.9,44)	(37,39.9,43)	(42,45.2,48)	(46,49.2,52)	(42,45.5,49)	(3.2,3.5,3.8)
9	(65,71.6,76)	(61,64.1,67)	(51,55.8,60)	(48,51.3,54)	(46,51.1,56)	(65,70.2,75)	(58,62.5,66)	(54,58.4,62)	(62,66.6,70)	(66,71.7,76)	(2.5,2.7,2.9)
b_j (ten thousand /day)	(2.5,3,3.5)	(2.5,3,3.5)	(2.5,3,3.5)	(2.5,3,3.5)	(1.8,2,2.2)	(1.8,2,2.2)	(1.8,2,2.2)	(1.8,2,2.2)	(0.9,1,1.1)	(0.9,1,1.1)	

Table 2. traffic distribution between traffic zones and traffic hubs

time \ hub / zone	A	B	C	D	E	F	G	H	I	J	a_i (ten thousand /day)
1	0	(2,2.25,2.5)	0	0	0	(0,0.1,0.2)	(1.5,1.75,2)	0	0	0	(4.2,4.5,4.8)
2	(0.8,0.9,1)	0	0	0	0	(0,0.05,0.1)	(0,0.1,0.2)	0	0	(0.8,0.9,1)	(1,1.1,1.2)
3	0	0	(2.5,2.75,3)	(0.2.0.35.0.5)	0	0	0	(1.8,1.8,2)	0	0	(4.8,5.2,5.6)
4	0	0	0	0	0	(0.7,0.8,0.9)	0	0	0	0	(0.7,0.8,0.9)
5	(0.2,0.3,0.4)	0	0	0	0	(0.1,0.15,0.2)	(0,0.05,0.1)	0	0	(0,0.1,0.2)	(0.35,0.4,0.45)
6	(1.6,1.65,1.7)	0	0	0	0	0	0	0	0	0	(1.6,1.7,1.8)
7	(0.7,0.8,0.9)	0	(0,0.05,0.1)	0	0	(0,0.1,0.2)	(0,0.1,0.2)	(0.1,0.15,0.2)	(0.9,1.05,1.2)	(0,0.15,0.3)	(1.9,2.1,2.3)
8	0	(0.8,1.15,1.5)	0	0	(1.9,1.95,2)	(0.7,0.81,1)	(0,0.05,0.1)	0	0	0	(3.2,3.5,3.8)
9	0	0	(0,0.05,0.1)	(2.5,2.6,2.7)	(0,0.075,0.15)	0	0	(0,0.05,0.1)	0	0	(2.5,2.7,2.9)
b_j (ten thousand /day)	(2.5,3,3.5)	(2.5,3,3.5)	(2.5,3,3.5)	(2.5,3,3.5)	(1.8,2,2.2)	(1.8,2,2.2)	(1.8,2,2.2)	(1.8,2,2.2)	(0.9,1,1.1)	(0.9,1,1.1)	

Table 3. traffic distribution between traffic zones and traffic hubs

zone \ hub	A	B	C	D	E	F	G	H	I	J	a_i (person/day)
1	0	25000	0	0	0	0	20000	0	0	0	45000
2	67	0	0	0	0	933	0	0	0	10000	11000
3	0	0	30000	3000	0	0	0	19000	0	0	52000
4	0	0	0	0	0	8000	0	0	0	0	8000
5	4000	0	0	0	0	0	0	0	0	0	4000
6	17000	0	0	0	0	0	0	0	0	0	17000
7	8933	0	0	0	0	1067	0	1000	10000	0	21000
8	0	5000	0	0	20000	10000	0	0	0	0	35000
9	0	0	0	27000	0	0	0	0	0	0	27000
b_j (person/day)	30000	30000	30000	30000	20000	20000	20000	20000	10000	10000	220000

6 Discussions

Although every uncertainty of traffic distribution has been considered in this model, and it could reflect the ordinary relationship among traffic sources, traffic facilities and traffic hubs, traffic distribution might also be affected by travel habit and some steady connections. Therefore, the result of this model might have a little deviation from the fact. If there were actual OD, then the modification parameters could be calculated and brought to this model:

$$k_{ij} = \frac{X'_{ij}}{X''_{ij}}$$

$$T_{ij} = X_{ij} \cdot k_{ij}$$

After modification, the restraints of this model would be untenable and the growth factor method is needed to make an iterative calculation.

The model has considered the uncertainty of traffic resistance, trip production-attraction of traffic zone and trip attraction-production of traffic hub simultaneously. In practice, we can choose one or two of them according to some real issues. For example, we only need to use fuzzy number of the last two when emphasizing the connection between the development of social economy and the scale of traffic hub. Or use fuzzy number for the first one in order to analyze the influence of traffic resistance on traffic distribution.

If the uncertainty of traffic resistance, trip production-attraction of traffic zones and trip attraction-production changed significantly, or they would have different extent of uncertainty, then they could be adjusted respectively.

The objective function of this model was minimization of transport fee (time). In practice, we could change it for certain particular cases. Such as in order to emphasize the sustainable development of environment and energy source in eco-transportation system, it is necessary to alter the objective function into minimization of environment pollution or energy resource consumption. Meanwhile, the model has only taken into account the traffic distribution of non-directional service traffic hub, but has ignored the directional service traffic hub existing in some cities. It would be worthwhile to consider how to combine them together in the future.

Acknowledgements

The work described in this paper was supported by grants from the National Basic Research Program of China（Project No. 2005CB724205）and a grant from the Intelligent Transportation System of WuHan Technology University.

Notation

The following symbols are used in this paper:

$\tilde{a}_i$ = trip production of traffic zone i;

$\widetilde{b}_j$ = traffic attraction of hub j;

$\widetilde{c}_{ij}$ = traffic resistance from traffic zone i to hub j;

k_{ij} = modify parameter from zone i to hub j was;

n = the number of traffic zones;

m = the number of hubs;

T_{ij} = final traffic distribution volume from zone i to hub j;

X'_{ij} = traffic distribution volume from zone i to hub j of actuality investigation;

X''_{ij} = traffic distribution volume from zone i to hub j of actuality forecast with this model;

$\widetilde{X}_{ij}$ = traffic distribution volume from traffic zone i to hub j;

References

A-Long Hu (2006). "Study on the layout of urban external traffic transfer hubs". Southeast jiaotong university master degree thesis, China.

Chim, Tat Wing , Yeung, Kwan L. (2004). "Traffic distribution over equal-cost-multi-paths". Proc., IEEE International Conference on Communications, v 2, 2004 IEEE International Conference on Communications, 1207-1211.

Chung and Albert Yuen Tai (2005). "Traffic distribution schemes for multi-homed mobile hotspots". Proc., IEEE Vehicular Technology Conference, v 61, n 4, 2005 IEEE 61st Vehicular Technology Conference - VTC 2005-Spring Stockholm: Paving the Path for a Wireless Future, 2127-2131

Dong-Mei Zhao, Jiang Liao, Sheng Li, Minqiang Li(2002). "Fuzzy comprehensive assessment model of oil-storehouse safety based on fuzzy AHP". Process in Safety Science and Technology Part B, 3,967-972

Han.Peng (2006). "Study on Planning Method for urban public traffic transfer hub". Jilin university master degree thesis, China.

Ho, H.W. Wong, S.C., Loo, Becky P.Y. (2006). "Combined distribution and assignment model for a continuum traffic equilibrium problem with multiple user classes." Transportation Research Part B: Methodological, 40(8), 633-650.

Hong-Yan Xiang, Shun-Ying Zhu, Hong Wang, Xin-Ping Yan(2005)."Fuzzy comprehensive assessment of short-term traffic flow forecasting". Journal of Wuhan University of Technology (Transportation Science and Engineering), 29(6), 921-924.

Kai Wang, Qi-Xiang Yu , Ya-Qin Wu, Yan-Fei Chen, Yi-Bo Wang(2002). "Application of fuzzy comprehensive assessment in optimal scheme selection of mine

ventilation system based on catastrophe theory". Process in Safety Science and Technology Part B, 3,890-893.

Moazeni and Somayeh (2006). "Fuzzy shortest path problem with finite fuzzy quantities". Applied Mathematics and Computation (New York), 183(1), 160-169.

Nanba, Shinobu, Konishi, Satoshi; Nomoto, Shinichi (2004). "Optimum Traffic Distribution Algorithm for Multiple-Satellite Systems under Power Constraints". Proc., IEEE Journal on Selected Areas in Communications, v 22, n 3, April, 492-500.

Okada, Shinkoh Gen, Mitsuo (1994). "Fuzzy shortest path problem. Computers & Industrial Engineering". 27(1-4), 465-468

Qian Jia (2006). "Study on the layout of comprehensive transportation terminal". Chan' an University master degree thesis, China.

Shun-Ying Zhu, Hong Wang, Shou-Quan Luo, Xin-Ping Yan (2006). "Traffic distribution forecast of urban external traffic hub". Journal of Wuhan University of Technology (Transportation Science & Engineering), 30 (6), 1041-1043.

Thompson, Russell , Young, William. (1988). "Traffic distribution fitting. A systematic methodology". Transportation Research Record, 1194, 87-98.

Xiao-Yu Ji, Iwamura, Kakuzo, Shao Zhen(2007). "New models for shortest path problem with fuzzy arc lengths". Applied Mathematical Modeling. 31(2), 259-269.

Yin-Zhen Li, Gen, Mitsuo, Ida Kenichi (1996). "Solving fuzzy shortest path problems by neural networks". Computers & Industrial Engineering.31 (3-4) 861-865.

A Land Use Transport Modeling Framework and Its Design and Development for the Province of Alberta

Ming Zhong[1], John Douglas Hunt[2], and John Edward Abraham[2]

[1]Dept. of Civil Engineering, University of New Brunswick, GD-128, Head Hall, 17 Dineen Dr., P.O. Box 4400, Fredericton, N.B., Canada E3B 5A3, Telephone: (506) 452-6324 Fax: (506) 453-3568, Email: ming@unb.ca

[2]Institute for Advanced Policy Research, University of Calgary, 2500 University Drive N.W., Calgary, Alberta, Canada T2N 1N4, Phone: (403) 220-8783, Fax: (403) 282-7026, Email: jdhunt@ucalgary.ca, jabraham@ucalgary.ca

Abstract

A land use transport modeling framework called PECAS, which represents Production, Exchange, Commodity Allocation System (PECAS) is introduced here. The virtue of such a system is that it places the transportation planning in the context of the whole economy by considering the fundamental root of transport demand, which is the exchange of goods, services, and labor among various activities, such as industries, services, households, and government. PECAS consists of the following three modules: Activity Allocation (AA), Space Development (SD) and Transport Supply (TS) and is linked to an aspatial regional economic model. This paper describes the design and development of such a model for the Province of Alberta, Canada in terms of its data sources, novelty, and strategies for further calibration. The framework has been applied to a number of cities, states, and provinces in North America, and it is introduced here as a better planning tool to their counterparts in China.

INTRODUCTION

The province of Alberta experiences the fastest economy and population growing within the country in recent years. It had a population of about 1.6 million in 1970, but it increased to over 3.3 million in 2005. It has a GDP of C$215.858 billion in 2005 with per capita of C$66,279, which is the highest in the country. The Calgary-Edmonton Corridor is the most urbanized region in the province and one of the densest in Canada. Measured from north to south, the region covers a distance of roughly 400 kilometers. In 2001, the population of the Calgary-Edmonton Corridor was nearly 2.2 million (72% of Alberta's population). It is one of the fastest growing regions in the country. A 2003 study by TD Bank Financial Group found the corridor is the only Canadian urban centre to amass a U.S level of wealth while maintaining a Canadian-style quality of life, offering universal health care benefits. The study found GDP per capita in the corridor is 10 percent above average U.S. metropolitan areas and 40 percent above other Canadian cities (Wikipedia 2007).

The fast growing economy requires a carefully prepared statewide planning for a sustainable development. A land use transport model based on the PECAS framework is being developed at the University of Calgary in response to such a need. The models based on the framework have been or are being developed for a number of cities, states or provinces in North America (Abraham *et al.* 2005[a], Abraham *et al.* 2005[b], Hunt and Abraham 2005). It is introduced here as a better tool for the planning communities in China, in an effort to keep them up with her rapid urbanization.

The Alberta land use transport (LUT) interaction model is the one of the latest developments of such models in North America based on the PECAS framework. It is a spatial economic simulation system with a focus on transportation planning and land development planning. The PECAS framework draws from previous modeling experience, especially MEPLAN (Echenique *et al.* 1990) and TRANUS (De la Barra *et al.* 1984), and provides significant improvements in theoretical basis and behavior representation (Hunt and Abraham 2005).

This paper is organized as the follows: Section 2 provides an overview of the PECAS framework; Section 3 elaborates the model design and model inputs; Section 4 gives the further strategies for model calibrations, and finally Section 5 concludes the paper.

PECAS FRAMEWORK

The PECAS modeling framework consists of the following three modules: activity allocation (AA), space development (SD), and transport supply (TS), and it is linked with an aspatial economic model. The AA module represents how activities locate within the space provided by developers and how these activities interact with each other at a given point in time. The SD module simulates the actions of developers in the provision of space (land and floorspace) where activities can locate, including the new development, demolition and re-development that occurs from one point in time to the next. The TS module uses its networks to establish congested network times and costs and associated interchange (dis)utilities that PECAS uses in its consideration of the interactions between land use zones. The aspatial economic module forecasts future economy status, which provides the scale of activities considered in the model. PECAS uses an aggregate, equilibrium spatial I/O modeling approach for allocating activities into different zones, but has a micro-simulation based space development sub-model.

The area covered by PECAS is organized into a set of land use zones. Activities locate in these zones and interact between them. Ideally these zones are the same as the transport zones used in the transport model being run with PECAS, or at least are aggregations of whole numbers of adjacent transport zones. The connectivity among the land use zones is then based on the representation provided by the transport model,

PECAS works through time in a series of discrete, fixed steps from one point in time to the next, with the activity allocation module running at each point in time and the space development module considering the period from each point in time to the next. The steps can be of any fixed duration, but 1 year time steps are recommended since they allow an appropriately quick response of land developers in the land development module to the prices established in the activity allocation module.

The theories base for the PECAS framework is described in detail in Hunt and Abraham (2005). An open software implementing PECAS has been developed in Java with many individuals and agencies' contributions.

MODEL DESIGN AND INPUTS

Model design

The design of the Alberta PECAS model was set up with an Excel spreadsheet as shown in the Figure 1. The first row in the spreadsheet lists all the commodities considered in the model, which include goods, services, labor, land and floorspace. The first column lists all the activities producing or consuming commodities, including industries, governments, households, inventory and import/export. *The Industries, Businesses, and Government Goods & Service Make Table* on the top left corner reflects the production patterns of various activities in supplying goods and services, whereas *the Industries, Businesses, and Government Goods & Service Use Table* (in the middle of the diagram) shows the consumption patterns of these activities. Between the two tables, there is *the Import & Export Table* consisting of two rows showing the import and export quantities of listed commodities. The total supply in the region is the sum of all commodities produced or made by various industries inside the region and those imported from outside the region, whereas the total demand is the sum of all commodities consumed by various activities and those exported outside the region. The AA module runs each year to determine the locations of all activities based on the equilibrium condition that the total regional supply equates to the total regional demand for all commodities.

The Labor Make Table (in the center of the figure) shows the labor production patterns by households and *the Labor Use Table* (in the center of the figure) indicates the labor consumption patterns of the activities. *The Household Goods & Service Use Table* shows the commodity and labor consumption patterns of households. All of the above interactions or the exchanges of goods, services, and labor involving shipping or travel will be shown in the bottom box – *Transportation Supply Table*. Land and space are non-transportable goods and they are shown on the right-hand boxes. *The Land Development Table* on the top right corner shows available lands for different development types based on certain zoning regulations. *The Non-residential Floorspace Use Table and Residential Floorspace Use Table* show the space consumption patterns of businesses (including industries, services, government, and institutions) and households respectively.

The diagram is of importance, as it provides a whole picture showing model structure, individual components, and the interactions among them and thus serves as a guide for data collection and model development.

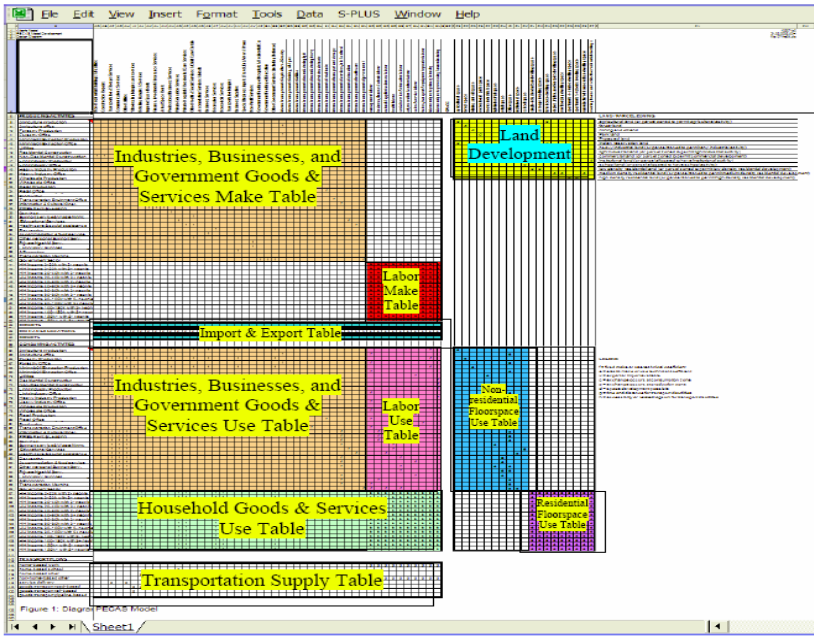


Figure 1 Design Diagram for Alberta PECAS Model

Activities

The activities to be allocated to zones in the Alberta PECAS model include industries, governments, households, inventory and import/export. The industries included are primarily based on the "Small Categories" of North America Industry Classification System (NAICS) and some of them are combined to reflect special characteristics of the Alberta economy. The size of each industry was based on the Input-output table of the province. Many of the industries were divided into two portions – a "production" portion and an "office support" portion to represent very different space consumption patterns of white-collar workers from their blue-collar counterparts in the same industry. The split was done based on the proportion of workers in production vs. office occupations associated with the industry, using the Canadian 1% Public Use Micro-sample (PUMS) data from the Census. The region-wide size of each activity is measured in year 2001 nominal dollars, with the year 2001 size based on the Alberta Input-output table and future year sizes factored up based on an employment projection for the province. Government, import/export and inventory are also included in the regional economic model because of their special roles (e.g., policies, regulations, and control of large surplus or deficit). Totally 32 business activities are included in the model.

Household categories are constructed based on household income and size (with or without children) because it is expected that households segmented in this way would show very different behaviors. The household income level was divided into the following five bins: 0 – 10,000, 10,000 – 30,000, 30,000 – 50,000, 50,000 – 100,000, and 100,000 and more ($/year). The household size was separated into the following two categories: 2 or less people and 3 or more people. The combination of these two factors (e.g., HH Income 10-30k, 3 or more people) results in that total 10 household categories are used in the model.

Commodities

The 26 goods and 29 services commodities, which coincide with the primary output of each industry as defined in the regional Input-output table, are included in the model. Quantities of goods and services commodities are measured in nominal 2001 dollars of production and obtained from the Input-out table.

Ten general labor categories are used in the model and they are based on 1991 Standard Occupation Classification (SOC-91). Quantities of labor commodities are measured in nominal 2001 dollars by using the total amount of "Wages and Salaries", "Supplementary Labor Income", and "Mixed Income" from the Input table. 2001 Canadian PUMS data are used to allocate them to the labors generated by different households. There are some inconsistencies between SOC categories and those used in Canadian PUMS data and therefore some adjustments have to be made to match them.

The non-residential floorspace is measured either by employee-space or square foot. For the blue-collar ("production") portion of the Agriculture, Mining and Oil, and Construction industry, employee-space is used, whereas for their white-collar portions square foot is used. All the other non-residential space types, including office, commercial, education, government and institution, light and heavy industry, are measured by square foot (sqft). The residential spaces are measured by the number of rooms, and the following seven categories defined in the 2001 Canadian Census are used in the model: Single-detached urban, Single-detached rural, duplex, row housing, apartment in low-rise (4 or less stories), apartment in high-rise (5 or more stories), and manufactured homes. To reflect different housing behaviors of owning and renting, these categories are further divided into owned and rent types, which results in a total of 14 residential space categories.

Zones

Totally there are 778 zones used for the Alberta model. These zones are mostly based on census tracts (CTs) or dissemination areas (DAs) for the metropolitan areas and census subdivisions (CSDs) for the rural areas to reflect different treatments to the two distinct features regarding the density of activities. Special considerations are given to Indian and federal reserved lands, summer villages, parks and protected lands. Each of these subcategories within a CSD is treated as a land use zone even though they may not be spatially connected.

Transport Networks

Three transport networks are used in the model: highway, railway, and pipeline. A very detailed highway network is used, which includes all high-level roads (express,

arterials, and collector) and local roads in the province and consists of 350,024 links. There are 4,926 links for the railway network and information regarding the operational status of each link (e.g., railway/sidetrack, abandoned or in-use) is available. The pipeline network is included in the model because of the significance of oil business in the province. The pipeline network is quite dense in the southeast region of the province and indicates that a large amount of oil and gas is transported by the mode.

"Make" relationships

The *make* relationships refer to the production patterns of activities. For example, how much milk and beef are produced by the Agriculture industry in the province is a make relationship.

Make of goods and services by industry

The Output table from the regional Input-output table was directly used to establish the values of the amounts of commodity made by individual activity with the unit of per dollar of production. These values are fixed for year 2001 and future year runs of the model. The number of commodities produced by the industries ranges from 1 to over 20, which shows distinct production patterns of different industries.

Make of labor by households

The make of labor by households is based on the Alberta employment data from 1996 and 2001 Canadian PUMS data. The PUMS households are aggregated to count the number of jobs by the occupation code in each income band and household size. Factors are used to expand the sample for matching the total employment. The PUMS ratios are used to assign a base relationship between household income categories and employment, but the coefficients are made elastic so that labor production can respond to local changes in supply, wages, and transport conditions. This variation includes changes in total labor production (reduced or increased participation in the labor force) and changes in the type of labor produced.

"Use" relationships

The *use* relationships refer to the consumption patterns of activities. It reflects the fact that, in order to produce certain commodities, how much of different input is needed.

Use of goods and services by industry

The Input table was used directly to establish the values of the amounts of commodity used per dollar of production. These values are fixed for year 2001 and future year runs of the model. The table shows very detailed consumption patterns across industries. Most of industries consume over 20 commodities in their production processes.

Use of goods and services by households

Use of goods and services by households is based on the Final Demand table from the Alberta Input-output tables. The total expenditure on categories of goods and services from the Final Demand table are allocated to individual household categories based on the estimated number of households in each category (based on 2001 PUMS data) and corresponding consumption percentages over the total household income, and then divided by the number of households in each category. The resulting pattern of

expenditures is relatively homogeneous with respect to each household category (defined by household income and size).

Use of labor by industry

The use of labor by industry is based on the Alberta employment sample data from 2001 Canadian census. The PUMS individuals were aggregated to count the number of jobs by the occupation code in each industry. As mentioned above, some industries (e.g., "Agriculture" and "Mining and Oil") are split into white and blue-portions as they have quite different labor use. Industry use of labor coefficients are elastic, to allow a given quantity of production to use more or less of a given type of labor in a zone, depending on the local labor market, with coefficients to be established in Stage 2 calibration (see CALIBRATION STRATEGY).

Use of floorspace by industry

Industry use of floorspace coefficients are based on standard floorspace per employee rates. Empirical floorspace use ratios were first used based on the modeling experience elsewhere (PB Consult *et al*. 2005). Floorspace consumption coefficients are elastic, to allow changes in the amount of floorspace used per dollar of production in response to changes in floorspace price. The ratios will be examined and modified during the calibration to match the employment with the industry floorspace use.

Use of floorspace by households

The coefficients for the use of floorspace by households are set to sum to 1, so that the number of dwelling units consumed in a zone will always equal the number of households located in the zone. Dwelling types are substitutable, however, so that households can locate in different types of housing. Initially the housing patterns are assumed to be same for all the zones within the region and consumption patterns for different household categories are based on the census data. The relative preference for different types of housing by different categories of households will be calibrated during stage 2 calibration (see CALIBRATION STRATEGY).

Imports and Exports

Imports and exports interact with the markets outside the region through exchange zones, with an import function and an export function for each commodity in each zone. These are currently set so that if all prices were equal to an expected average price in all zones, the base year total imports and exports would be divided amongst all of the zones equally. A logistic-plus-linear curve (Hunt and Abraham 2005) is used with parameters such that the imports in each zone would go to zero if the prices go to zero. For exports, the elasticity is initially set so that the exports from each zone would double if prices go to zero.

Floorspace short term supply function

The total inventory of floorspace of each type is fixed from the land development module, but the short term supply of floorspace for the spatial IO model (the AA module) is a function of this total inventory. This represents the tendency for some floorspace to remain vacant, with the amount of vacant floorspace increasing as prices decline. Initial parameters controlling the short term floorspace supply function have

been established using expert judgment and average observed prices (e.g., rent of $15/sqft per square foot of office space); further calibration of these parameters will be required during stage 2 calibration.

Floorspace synthesis and zoning regulations

Calculating the non-residential floorspace of each type required in each zone is based on empirical floorspace per employee ratios (PB Consult *et al*. 2005) and 2001 Alberta Place of Work (POW) data. The POW data provides the number of employee by industry by zone. The employee for some industries, such as agriculture and mining and oil, was split into the white-collar and blue-collar portion based on the ratios obtained from the PUMS data and assigned into office and employee spaces respectively. All of the other occupations were combined into groups who share the same space use type (e.g., commercial or industrial). The total number of employee in different occupations (e.g., management or health care) was multiplied with the typical floorspace use ratio (PB Consult *et al.* 2005) to generate the floorspace use inventory. Residential floorspaces were calculated based on housing use coefficients for different household categories from census data for each zone. Some vacancy rates are assumed (e.g., 7%) and added into the current floorspace inventory serving as space capacity used in the AA module, because the land development module will not run until it gets renting price from the AA.

The following two data sets are used to serve as initial zoning regulations: the Alberta land cover data for the rural areas and the DMTI land use data for cities and towns. These data are also used to reconcile with the floorspace quantity generated by POW and population data, and in some cases serve as the indications of floorspace assignment and future zonings.

Land development module

An algorithm used in the Sacramento PECAS model (Abraham *et al.* 2005[a]) will be applied to the Alberta Model for simulating land development. First, the land of the whole region is divided into 30 meters by 30 meters of grid cells, which is about an equivalent size of a parcel. The grid cells with this size are used because the land units being considered should be small enough so that assigning single floorspace types and year-built values to each unit of land can represent the diversity of existing and future coverage.

The development model for each parcel is a logit model containing alternatives for development (or redevelopment) as each future allowed development type, as well as two additional alternatives – the "add-more" alternative and the "do-nothing" alternative. The "demolish" alternative is included when "Vacant" is included as a future allowed development type. The utility function of logit model considers both region-wide rent and local properties. A transition utility function is used to consider whether a change or new development is profitable or not (Hunt and Abraham 2005).

CALIBRATION STRATEGY

A three-stage calibration process is used to develop the values of the parameters in the

model. In the first stage, values are developed for certain 'S1' parameters in each module separately. The intention is that these first-stage values for the S1 parameters will remain fixed as the model development and calibration work progresses. When suitable observations of system behavior are available, statistical methods are used to estimate appropriate values for S1 parameters. In some cases, only a single observation is available, and direct methods are used to provide values. At this point it is not necessary that the modules can be run: the components of the modules are being 'assembled' and the outputs of the modules are not yet being considered. For the purpose of brevity the S1 parameters will not be listed separately in this paper; the methods used to establish their values are described in the above section on model inputs.

In the second stage, initial values are established for all of the parameters that are not S1 parameters, called the 'S2' parameters – considering the fit of each module in isolation. The fit for a given module concerns specified targets for outputs from the module, so the module needs to be run in order for it to reproduce these outputs. Thus, a full set of required inputs for each module needs to be developed, including all those provided by other modules and all those provided exogenously. In order to obtain reasonable values for the S2 parameters, these inputs need to be consistent with the specified targets, representing conditions similar to those that gave rise to the targets.

Calibration software is used to adjust the values of these S2 parameters. The software runs the module repeatedly, experimenting with small changes in the S2 parameter values, to establish a multi-directional dimension of simultaneous change in S2 parameters that is likely to lead to a better fit to the targets. The fit to the targets is measured using a weighted sum of squares of differences. Once the parameter values are established using the search routine, a variety of interactive tools are available for exploring the remaining lack of fit and its relationship to parameters and to other (conflicting) targets. These tools make use of a weight sensitivity matrix, which calculates the change in goodness of fit for each target, and the change in each parameter value, due to small changes in the weight of a target. These tools can suggest changes in the model design to achieve better fit to important targets. The calibration software and the use of the weight sensitivity matrix is described in more detail in Abraham and Hunt, 2000 and Abraham 2000.

In the third stage, the initial values established for certain sets of the S2 parameters are revisited – for all of the modules simultaneously (the spatial IO (AA) module, the space development (SD) module, and the transport supply (TS) module), considering the fit of all modules together, with the full model running, so that inputs to the modules are coming from the other modules in the way they would for a model run. These re-visited S2 parameters are designated 'S3' parameters.

Currently, S1 parameters have been developed based on the individual input datasets mentioned above. Sometimes because of the lack of data, certain arbitrary values or empirical values from previous modeling experiences are used for S1. S2 parameters are being estimated by fitting each individual module against some targets. It is anticipated that, as data is collected and calibration is continued, S1 and S2 parameters will be updated frequently and S3 will be estimated after each module satisfies the converge criteria.

CONCLUSIONS

A land use transport (LUT) model is being developed for policy analysis, in response to the fast growing economy in the Province of Alberta, Canada. The model is based on the full PECAS framework. It is now undergoing further calibration and testing, for use in policy analysis in the near future. The model is featured with a large number of industry, household and commodities categories, micro-simulated land development module, and inclusion of highway, railway, and pipeline networks.

The first-stage model development is currently finished in that the prices of commodities are reasonable and all of the S1 parameters have values that are not expected to require modification. The model will not be ready for policy analysis; however, until the values of the S2 parameters are modified by the calibration software in its comparison of the model's outputs against the targets.

Further design changes in the model may occur to address unexpected difficulties in S2 or S3 calibration (guided by the reports on the weight sensitivity matrix from the calibration software), or if the model is called upon to perform in situations for which it was not designed. Given the complex and interactive nature of the model, the judgment of the ultimate degree of goodness of fit and the adequacy of the performance of the model will always be dependent on the person doing the judging and the way in which the model is being used (Abraham 2000, Abraham and Hunt 2000).

Once the S2 and S3 calibration program described in this paper is complete, the model will be used to analyze various policies, including, but not limited to, future highway construction plans, commuter railway network, pipeline network development, land use development policies, housing affordability plans, and transportation pricing options.

The PECAS framework is introduced here as a better planning tool than traditional 4-step transportation models. The rapid economic growth and urbanization in China may need such a tool to achieve a long-term sustainable development.

ACKNOWLEDGEMENTS

The authors are grateful towards Natural Science and Engineering Research Council (NSERC), Canada for their financial support. The authors also wish to thank Alberta Treasury for the regional Input-output Tables, Alberta Transportation for the place of work (POW) data and Laurie Schretlen at the Maps, Academic Data, Geographic Information Centre (MADGIC), University of Calgary for helps in collecting the GIS data used in this model.

REFERENCES

Abraham, J.E. (2000). "*Parameter Estimation in Urban Models: Theory and Application to a Land Use Transport Interaction Model of the Sacramento, California Region"*, Ph.D. Dissertation, Department of Civil Engineering, University of Calgary.

Abraham, J.E., Garry, G.R. and Hunt, J.D. (2005[a]). "The Sacramento PECAS Model", presented at the *2005 Transportation Research Board Annual Meeting*.

Abraham, J.E. and Hunt, J.D. (2000). "Calibration Strategies for Large-scale Urban Models", *Transportation Research Record* 1722, 9-16.

Abraham, J.E., Weidner, T.J. and Hunt, J.D. (2005[b]). "Calibrating the Oregon Spatial Allocation Models". Presented at *the 52nd Annual North American Meetings of the Regional Science Association International*, Las Vegas NV, USA, November 2005.

De la Barra, T., Perez, B. and Vera, N. (1984). "TRANUS-J: Putting Large Models into Small Computers", *Environment and Planning*, 11B, 87-101.

Echenique, M.H. Flowerdew, A.D.J., Hunt, J.D., Mayo, T.R., Simmonds, D.C., and Skidmore, I.J. (1990). "The MEPLAN Models of Bilbao, Leeds and Dortmund", *Transport Reviews*, 10, 309-322.

Hunt, J.D. and Abraham, J.E. (2005). "Design and implementation of PECAS: A generalized system for the allocation of economic production, exchange and consumption quantities". In: *Foundations of Integrated Land-Use and Transportation Models: Assumptions and New Conceptual Frameworks*, Lee Gosselin and Doherty editors, Elsevier (book chapter).

PB Consult, HBA Specto, EcoNorthwest. (2005). "*Oregon2 Transitional Model, Model Description at Finalization"*, version 1.2, Submitted to the Oregon Department of Transportation.

Wikipedia. (2007). Website: http://en.wikipedia.org/wiki/Alberta (Last accessed August 22, 2007).

EVALUATION OF YANGSHAN INTERNATIONAL HUB PORT COMPETITIVENESS FOR CONTAINER TRANSSHIPMENT BUSINESS

ZHANG, Rong[1] and CHEN, Tao[2]

[1]Associate professor, PhD, School of Transportation Engineering, Tongji University, Tel: 13817667078; email: gxxzr@online.sh.cn

[2]Master Candidate, School of Transportation Engineering, Tongji University, Tel: 13661537113; email: ale922@163.com

Abstract

The international hub port construction of Yangshan is an important part of Shanghai's international shipping center development. Having little container transshipment business is a major problem that the Yangshan port faces for international hub port development. This paper first analyzes the international hub port competition situation that the Yangshan port faced in the East Asian area, finding that Busan port is Yangshan port's major competitor in the container transshipment business development and international hub port construction. The major indicators which affect the hub port competitiveness for container transshipment business are designated. The ports of Yangshan, Busan and Singapore are chosen to set up the Fuzzy-AHP mode to analyze the importance of each indicator, and then the competitiveness of Yangshan, Busan and Singapore ports is evaluated. Last, after analyzing the advantages and shortages of Yangshan port, suggestions are provided for Yangshan international hub port construction.

Key words: *hub port, container transshipment business, competitiveness, fuzzy-AHP model*

1 Introduction

With the development of the international container maritime transportation, East Asia has already become the most booming area for global container shipping services. Based on the booming hinterland, the Shanghai port has become one of the largest container ports in the world. The throughput of Shanghai port is growing fast, and approached 21710 *k*TEU in 2006. It has still kept the third biggest port status in the world. However, the container transshipment business is not as prosperous as the throughput, which has become the major problem of the Shanghai port development. In 2006, transshipment cargo volume from international feeder ports is only 785 *k*TEU, whose percentage of throughput is 3.6%; the transshipment cargo from national feeder ports is only 692 *k*TEU, whose percentage of the throughput is 3.35%. The data of the Yangshan port is a little bigger than the average level of the Shanghai port. The transshipment cargo volume to the Yangshan port from international feeder ports is 170 *k*TEU, whose percentage of throughput is 5.25%; the transshipment cargo from national feeder ports is only 142 *k*TEU, whose percentage of the throughput is 4.39%. However, the transshipment cargo volume is comparatively much larger in other major

container ports in the world, which becomes an important cargo source for those ports' throughput growth. For example, the percentage of transshipment cargo is more than 40% of the throughput in major hub ports in East Asia (Singapore, Hong Kong, Kaohsiung and Busan), and about 20% ~30% in the three major ports in Europe (Rotterdam, Hamburg and Antwerp).

The Yangshan port has already had some of the infrastructures and policies which are necessary for an international hub port. First, the Yangshan port is located on the island far away from the continent, about 104 km away from the main navigation routes, which means the port location of Yangshan is relatively superior. Second, nine berths, whose water depth is more than 16 meters at present, has totally changed the condition of the deep-water berth shortage in the Shanghai port. This indicates that the Yangshan port has the necessary berth conditions of a hub port. Finally, the Yangshan port is the first bonded port in China, which gives it the policy advantage of a bonded area, bonded logistics zone, and an export processing zone. The customs environment is freer and the tax incentives are more attractive, which shows the policy of the Yangshan port is quite suitable for a hub port.

While the Yangshan port is developing the container shipment business and constructing an international hub port, other major Asian container ports such as Hong Kong, Singapore, Kaohsiung and Busan are also competing for or consolidating the status of an international hub port. Their advantage of good port location, modernized deep water terminals, high frequency of shipping lines, a developed distribution system, as well as a free customs environment brings challenges to the Yangshan port for international hub port development. Improving the competitiveness of the Yangshan port, attracting transshipment cargo from national and international feeder ports, and developing the Yangshan port into an international hub port are urgent problems for Shanghai to construct an international shipping center.

The Busan port is the Yangshan port's major competitor for the container transshipment business. First, these two ports have almost the same hinterland, including north China, Korea, Japan and the Far East area of Russia, as well as the container transshipment hinterland including some countries of west ,south and southeast Asia. Second, because of the ports of Hong Kong and Shenzhen, there is no advantage for Yangshan to attract transshipment cargoes from south China,; therefore cargoes from the ports of north China have become the major cargo source of the national feeder port transshipment business. North China ports including Qingdao, Tianjin, Dalian, and Lianyungang provide 50% of the national feeder port transshipment cargo of the Shanghai port. However, the shipping distance for the north China cargo to the west coast of North or South America is comparatively shorter when the Busan port is selected for transshipment. In addition, the Busan port has the

advantage of the clearance efficiency compared to the Yangshan port, and the polices of "Returning the export tax rebate right after cargoes leave the original feeder port" and "Opening the market of national feeder lines to foreign shipping companies" have not been approved or executed, which produce a negative function for shippers or carriers choosing a national trunk port. Therefore, Busan has attracted a large amount of cargo from the ports of north China. In 2004, the ports of north China, including Tianjin, Dalian and Qingdao, provided about 450 *k*TEU to Busan for transshipment, which is about 2/3 of all the transshipment cargo to Busan from China.

This paper selected the Busan port, which is the major competitor to Yangshan, as well as Singapore which is the biggest international hub port, as references. The competitiveness of the Yangshan port is evaluated, and the advantages and shortages of the Yangshan port are analyzed. Suggestions are provided for an international hub port construction of the Yangshan port.

2 Literature Review

The three major trunk routes for container maritime transportation are Trans-Pacific (T/P), far-East-Europe (F/E) and trans-Atlantic (T/A). Since East Asia is the intermediate district connection of the T/P and F/E routes, and the economic prosperity of of the routes,, the demand of container maritime transportation is growing fast; many shipping companies deployed their trunk routes at major ports in this region to meet the demand. In early days, most shipping companies selected only Busan, Kaosiung, Hong Kong and Singapore as hub ports, but with the fast growth in cargo volume of the China market, as well as its port development, it has become more complicated for shipping companies to deploy the trunk routes and select hub ports other than North-America and Europe（Tiwari *et al.*,2003）. Shanghai and Yantian became available to shipping companies for selection as hubs; accordingly, competitiveness among major ports in East Asia has become more serious (Tai and Hwang, 2005).

There are many factors that might affect the competitiveness of a hub port. The reasons for hub port choice are complex and cannot always be fully explained in terms of rational elements (Chou and Chu，2003). Many scholars have extensively studied the factors which may affect the competitiveness of a hub port. Huan Shen（2006） proposed that the five most important factors include port direct hinterland, port water deep and infrastructures, policies, port location and port handling efficiency; Yong-bo LU *et al.* (2002) figured out that the factors which affect port competitiveness include port operation, customer service, port extra environments, the throughput and its development, port facilities and port management; Tai and Hwang （2005）classified factors into internal factors (including port handling facilities and its efficiency, port

charge, customer service, etc), external factors (including cargo sources from the hinterland, efficiency of clearance, port location, frequency of trunk and feeder routes, etc) and operational factors (including saving of operating costs, etc). These studies summarized the major factors which are related to the port operation, facilities and infrastructures, location and customer service, but most of them considered little about the importance of the logistics park and related logistics businesses for international hub port competitiveness. Lu and Yang（2006）analyzed the factors that might affect the competitiveness of port distripark (logistics park) , finding out the respective advantages of the port distriparks of Kaohsiung, Hong Kong and Shanghai, but they did not put the distripark related indicators into the port competitiveness indicator system. 이성우（2005）concluded that port backup areas (distriparks) affect the competitiveness of ports; then he classified the factors into two categories as terminal factors and port backup area factors. A regression analysis model is set up to evaluate the performance of a port. Some other scholars think the port management related factors such as port operation and customer service are most important to hub port competitiveness. Thomson B.J.（1998）stated that the most important factors that affect a port's becoming a transshipment port are the port-related departments' efficiency and port handling facility's efficiency. Brian Slack (1985) thought that the transshipment port decision-makers are influenced more by price, and service considerations of land and ocean carriers. Lobo and Capt.Vivek Jain think that port infrastructures are relatively immobile while competition between hub ports is the competition of customer service quality. They design 54 related indicators and a Gap Analysis model is set up to analyze the port's customer service quality. However, port location and infrastructure as well as policy conditions are also very important to international hub port competitiveness.

Indicators this paper selected are based on the literature review above, as well as discussions with experts, which are classified into following five categories:

- **Port location:** including closeness to main navigation routes, convenience of transportation between port and logistics park, and hinterland scale;
- **Terminal infrastructure and facilities:** including berth condition, marshalling yard area, and port handling facilities;
- **Port operations:** including port charge, frequency of trunk and feeder routes, efficiency of cargo handling, efficiency of clearance, transshipment cargo incentives, and quality of customer service;
- **Port logistics park:** including scale of bonded logistics park/free trade zone, land/warehouse rent, labor cost, and corporate tax incentives;

- **Policy:** including bonded/free trade policy, policy stability, and policy execution.

3 Methodology

The Fuzzy- AHP model is an ameliorated AHP model using the fuzzy theory. The Pairwise comparison matrix is constructed by using a set of triangle fuzzy numbers (*l, m, u*) instead of a single number; then the weights and scores of the indicators are calculated according to the fuzzy algorithm. Considering the illegibility while people are analyzing and thinking, the results of the model are more accurate and reasonable, and match the practice better. The Fuzzy- AHP model is introduced below:

(1) Employing the principles of Fuzzy-AHP: a hierarchy framework of the indicators must first be constructed;

(2) Calculating weight: A pairwise comparison matrix is generated to investigate the relative importance of any two indicators in the proposed hierarchy layers by using the fuzzy numbers as shown in table 1. Table 2 showsa pairwise comparison matrix of the indicator 1 and indicator 2 in the layer *k*, where

$$A_{12}^{k} = (l, m, u),\ A_{21}^{k} = (\frac{1}{u}, \frac{1}{m}, \frac{1}{l}).$$

Table 1. Fuzzy numbers for pairwise comparison matrix

Fuzzy numbers (*l, m, u*)	(9/9,9/9,9/8)	(9/8,9/7,9/6)	(9/6,9/5,9/4)	(9/4,9/3,9/2)	(9/2,9/1,9/1)
importance	Fair	A little More important	Obviously more important	Strongly more important	Extremely more important
score	Fair	A little better	Obviously better	Strongly better	Extremely better

Table 2. Pairwise comparison matrix of two indicators

	Indicator 1	Indicator 2
Indicator 1	1	A^{k}_{12}
Indicator 2	A^{k}_{21}	1

Given: $d'(A_i) = \min V(A_i \geq A_k)$, where

$$V(A_i \geq A_k) = \begin{cases} \dfrac{l_2 - u_1}{(m_1 - u_1) - (m_2 - l_2)}, l_1 \leq u_2 \\ 1 \qquad\qquad , otherwise \end{cases}$$

For $k=1,2,3, \ldots,n, k \neq i$, then the weight vector is given by:

$$W' = (d'(A_1), d'(A_2), \Lambda\ , d'(A_n))^T$$

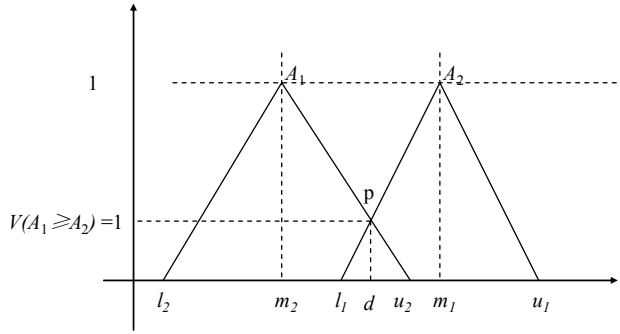


Figure 1. The elements of fuzzy number comparison

Then the weight of indicator i of layer k is:

$$W_i^k = \frac{d'(A_i^k)}{\sum_{i=1}^{n} d'(A_i^k)}$$

Accordingly, the synthetic weight of indicators in the last layer can be calculated:

$$W_N = \begin{pmatrix} W_1 & W_2 & \Lambda & W_N \end{pmatrix},$$

where N represents the index of all the indicators in the last layer.

(3) Score calculation: The same as step (2), a pairwise comparison matrix is generated to investigate the relative scores of any two indicators by using the fuzzy numbers as shown in table 1. N represents the index of all the indicators in the last layer, and then the scores of m objects that relate to N indicators can be calculated as below:

$$S_N^m = \begin{pmatrix} S_1^1 & S_1^2 & \Lambda & S_1^m \\ S_2^1 & S_2^1 & \Lambda & S_2^m \\ \Lambda & \Lambda & \Lambda & \Lambda \\ S_N^1 & S_N^2 & \Lambda & S_N^m \end{pmatrix}$$

(4)The synthetic score vector of m objects can be finally calculated as below:

$$(C_1, C_2, \Lambda\ C_m) = W_N \times S_N^m = (W_1 \quad W_2 \quad \Lambda \quad W_N) \times \begin{pmatrix} S_1^1 & S_1^2 & \Lambda & S_1^m \\ S_2^1 & S_2^2 & \Lambda & S_2^m \\ \Lambda & \Lambda & \Lambda & \Lambda \\ S_N^1 & S_N^2 & \Lambda & S_N^m \end{pmatrix}$$

4 Model Application and Results Analysis

4.1 Hierarchy framework of indicators

According to the 19 indicators above, the hierarchy framework is constructed as in figure 2.

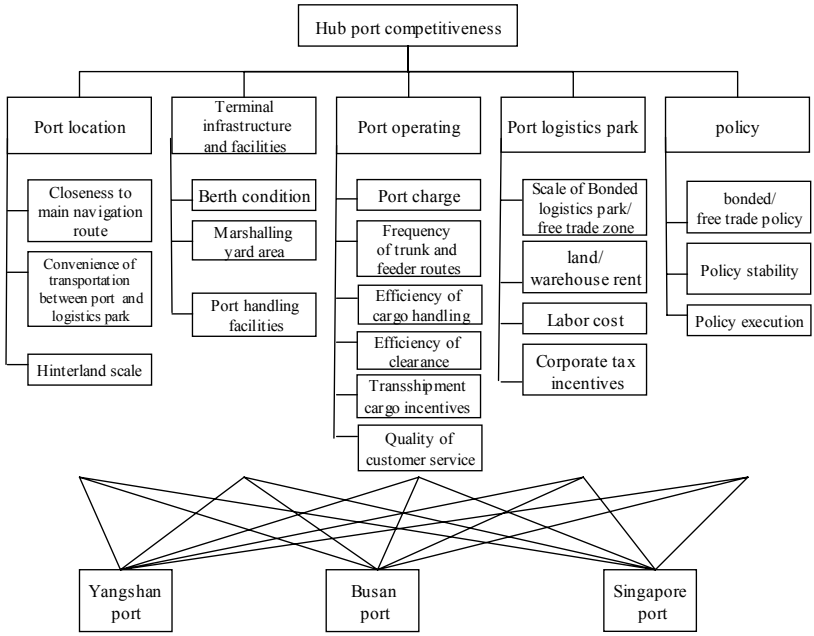


Figure 2. Hierarchy framework of indicators of international hub port competitiveness

4.2 Investigation and Calculation

According to the basic data which are needed for Fuzzy-AHP model construction, fifty related companies as well as ten experts were interviewed in 2005. Questionnaire forms were sent out to these companies and experts. The contents of the questionnaire, including the choice of indicators that probably affect the competitiveness of an international hub port, weights comparison between each two indicators and scores comparison among those three ports. All 60 questionnaire

forms were returned. Major interviewed groups and samples are show in the table 3 below.

Table 3. Details of the interviewed groups

Group	Samples	Proportion
Shipping company	20	33.3%
Logistics company	20	33.3%
Freight forwarders	10	16.7%
Experts	10	16.7%
Total	60	100%

The weights of each indicator as well as the scores related to each indicator of Yanshan, Busan and Singapore ports are calculated as in table 4 according to the basic data investigated.

Table 4. Weights and scores of indicators

Object	Indicators (Weight W^1)	Sub-indicators (Weight W^2)	Yanshan	Busan	Singapore
International hub port competitiveness	Port location (0.253)	Closeness to main navigation route (0.481)	0.191	0.233	0.576
		Convenience of transportation between port and logistics park (0.152)	0.208	0.375	0.417
		Hinterland scale (0.367)	0.474	0.316	0.210
	Terminal infrastructure and facilities (0.063)	Berth condition (0.704)	0.127	0.268	0.605
		Marshalling yard area (0.176)	0.333	0.334	0.333
		Port handling facilities (0.120)	0.334	0.333	0.333
	Port operating (0.230)	Port charge (0.208)	0.474	0.316	0.210
		Frequency of trunk and feeder routes (0.220)	0.191	0.381	0.428
		Efficiency of cargo handling (0.152)	0.292	0.333	0.375
		Efficiency of clearance (0.251)	0.194	0.245	0.561
		Transshipment cargo incentives (0.134)	0.263	0.263	0.474
		Quality of customer service (0.035)	0.292	0.333	0.375
	Port logistics park (0.208)	Scale of Bonded logistics park/free trade zone (0.103)	0.228	0.290	0.482
		Land/warehouse rent (0.416)	0.295	0.529	0.176
		Labor cost (0.223)	0.473	0.312	0.215
		Corporate tax incentive (0.258)	0.286	0.333	0.381
	Policy (0.246)	Bonded/free trade policy (0.509)	0.261	0.304	0.435
		Policy stability (0.119)	0.220	0.390	0.390
		policy execution (0.372)	0.167	0.416	0.417

According to the weights of each indicator and the scores related to each indicator of Yanshan, Busan and Singapore ports, C_Y, C_B, and C_S represent the synthetic score vector of Yanshan, Busan and Singapore ports, which can be calculated

as below:

$$(C_Y, C_B, C_S) = (C_1, C_2, C_3) = (W_1 \quad W_2 \quad \Lambda \quad W_{19}) \times \begin{pmatrix} S_1^1 & S_1^2 & S_1^3 \\ S_2^1 & S_2^2 & S_2^3 \\ \Lambda & \Lambda & \Lambda \\ S_{19}^1 & S_{19}^2 & S_{19}^3 \end{pmatrix}$$

$$= (0.273,\ 0.333,\ 0.394)$$

4.3 Results analysis

The results of the questionnaire survey indicate that among the 5 kinds of indicators that might affect the competitiveness of an international port, the most important indicator is "port location" (0.253); among the three sub-indicators belonging to "port location", "closeness to main navigation route" is the most important (0.481). The port of Singapore guards the throat of the East-Europe (F/E) global trunk routes, and the port of Busan is the key point of the Trans-Pacific trunk routes. Compared to these two ports, Yangshan port has no advantage on the indicator of "closeness to main navigation route". Another important sub-indicator is "Hinterland scale" (0.367) .Yangshan port has an obvious advantage on this indicator, for the Yangshan port has the hinterland of the Yangtze Delta, which is the most developed area of China, and the wide hinterland of middle and west China. Busan and Singapore belong to the transshipment-type port, which has little hinterland. But taking full advantage of their preponderance of closeness to main navigation routes and free port policy, these two ports have attracted a large amount of transshipment cargo from feeder ports in China, which makes China become their indirect hinterland. In addition, because theYangshan terminal is located on an island away from the continent, it has inherent limitations on the sub-indicator "convenience of transportation between port and logistics park". （0.152）

Another important indicator that might affect the competitiveness of an international port is "policy" (0.246). The most important sub-indicator belonging to "policy" is bonded/free trade policy（0.509）. Singapore, as a free port, has an obvious advantage in policy. Yangshan port has no advantage in policy because

(1) Yangshan port is a bonded port but not a free port because the Yangshan bonded port is still "a special area administrated by customs", but not "an area without customs territory". This difference makes the customs procedure much more complicated. For instance, if the Yangshan port was chosen for cargo transshipment, before cargoes leave the feeder port, the shipper must submit cargo information to the customs office of the feeder port. When the cargoes arrive at Yangshan port, the freight forwarder has to submit the cargo transshipment data to Yangshan Customs. After the cargo has been shipped, the shipping agent must submit the cargo information to

customs again. The diversity among the documents submitted to customs may delay the export tax rebate for several months. The same customs procedure at Busan port is much simple. This is also one of the major reasons for shippers in north China choosing the Busan port for cargo transshipment.

(2) The Chinese government has an incentive policy called "export tax rebate", which means that shippers can get some of their paid taxes back from the government only if their cargoes have been exported. The Yangshan bonded port is not "an area without customs territory". Shipping between a national feeder port and the Yangshan port belongs to national transportation and the cargoes can not be considered as exported cargoes when they are shipped on the national feeder line. So the shipper can not get an export tax rebate immediately after the cargoes leave the feeder port. Only after the cargoes have left the Yangshan port can the shipper get the export tax rebate. It is totally different if the shipper chooses the Busan port for cargo transshipment. The shipper can get the export tax rebate immediately after their cargoes leave the feeder port because the shipping line between the national feeder port and the Busan port belongs to the international shipping line, and the cargoes which leave the feeder port to Busan are considered exported cargoes.

In addition, the Yangshan port is the first bonded port, which is a new kind of special area in China. Some of the policies and measures have not been fulfilled; accordingly, "policy stability"(0.119) and "policy execution"（0.372）do not give an advantage for the Yangshan port.

The sub-indicators which belong to " port operating" include "efficiency of clearance"（0.251）, "port charge"（0.208） and "frequency of trunk and feeder routes"（0.220） are relatively important. Compared with Busan and Singapore, the Yangshan port has the obvious advantage of "port charge", but has no advantage in "efficiency of clearance" and "frequency of trunk and feeder routes". Because of the complicated customs procedure, small EDI coverage rate, and little coordination among different departments, the customs clearance time is much longer than at the other two ports. There is small frequency of trunk and feeder routes because Yangshan is a new port which has few shipping lines. In addition, compared with Singapore, the "transshipment cargo incentives" are not enough.

Among the sub-indicators belonging to the indicator of "Port logistics park"(0.208), "Land/warehouse rent"（0.416）is the most important. Land/warehouse rent in the Yangshan bonded logistics park is much cheaper than at the Singapore free trade zone, but much more expensive than at the Busan-Jinhae Free Economic Zone. The Yangshan bonded logistics park has an obvious advantage in "labor cost"（0.223）.

The indicator of "Terminal infrastructure and facilities" is not very important to hub port competitiveness (0.063). Among the sub-indicators, "berth condition" (0.063) is absolutely important. Berth condition includes the number of and the water depth of the berths. Yangshan port has deep-water berths, but the number of berths is not enough because the Yangshan port is under construction.

The synthetic score vectors of Yanshan, Busan and Singapore port competitiveness are 0.273, 0.333 and 0.394, respectively, which means Singapore is the most competitive hub port among these three ports, followed by Busan. The competitiveness of the Yanshan port is the weakest among the three. If the score of the competitiveness of the Singapore port is considered at 100, then the scores of Yangshan and Busan are 69.3 and 84.5.

5 Conclusion and Suggestion

(1) According to the calculations，Singapore is the most competitive hub port among these three ports, followed by Busan. The competitiveness of Yanshan port is the weakest among the three. But as a new port, it has good potential for improving competitiveness. .

(2) Among the 19 major indicators which might affect the competitiveness of an international hub port, the most important indicators are "closeness to main navigation route", "hinterland scale" belonging to port location; "bonded/free trade policy" belonging to policy; "efficiency of clearance", "frequency of trunk and feeder routes" and "port charge" belonging to port operating; "land/warehouse rent" belonging to port logistics park; and "berth condition" belonging to terminal infrastructure and facilities. Yangshan port has advantages in "hinterland scale", "port charge" ,"land/warehouse rent", and "closeness to main navigation route". But indicators of "bonded/free trade policy" and "efficiency of clearance" are three relatively weak indicators for the Yangshan port.

(3) The Yangshan port must first improve the competitiveness for attracting the transshipment cargo source from a national feeder port. "Returning the export tax rebate right after cargoes leave the original feeder port" and ""opening the market of national feeder lines to foreign shipping companies" are considered two major policies for Yangshan to improve the competitiveness of its hub port compared with Busan's. In addition, allowing shipping companies to deploy their shipping lines freely among the different port areas of the Shanghai port is a good idea to attract shipping companies doing container transshipment business at the Shanghai port.

(4) Carrying out a free port policy and improving the customs clearance efficiency are very important for theYangshan port to improve its competitiveness.

(5) The construction of a bonded logistics park just on the north of the Yangshan terminal on little Yangshan Island can change the condition that the logistics park is separated from the terminal in Yangshan port. Then the logistics cost is reduced and the logistics efficiency is improved, which can obviously improve the competition for the Yangshan port to develop the container transshipment business.

(6) The Yangshan port is a new port area of the Shanghai port and its infrastructure construction has not been completed, so the indicators such as "frequency of trunk and feeder routes" and "berth numbers" are not as strong as at the ports of Busan and Singapore. Therefore, while we speed up the construction of Yangshan port, it is necessary for the Yangshan port to build up good relationships and communication systems with other port areas of the Shanghai port, by forming a fluent, low-cost, and flexible transportation system, such as rail and shuttle ships between the Yangshan port and other port areas of the Shanghai port, so that the Yangshan port can be supported by the advantages of the Shanghai port, such as frequency of trunks and feeder routes, berth numbers, and large numbers of cargo sources to improve its international hub port competitiveness.

References

[1] Antonio Lobo and Vivek Jain. (2000). "Port users' perspective of the container transshipment business-Hierarchy of Service Quality Attributes and Dimensions". www.google.com

[2] Brian Slack (1985). Containerization, inter-port competition and port selection, *Maritime Policy and Management*, Vol. 12, No. 4, 293-303

[3] Chin-Shan Lu and Ching-Chiao Yang. (2006). "Comparison of Investment Preferences for International Logistics Zone in Kaohsiung, Hong Kong, and shanghai Ports from a Taiwanese Manufacturer's Perspective". *TRANSPORTTAION JOURNAL*^{TM}.

[4] Chien-Chang CHOU and Ching-Wu CHU. (2003). "A transportation demand split model for international ports in Taiwan area". *Journal of the Eastern Asia Society for Transportation Studies*, Vol.5, 625-637.

[5] Claude Comtois. (1999). "The integration of China's port system into global container shipping". *GeoJournal* 48: 35-42.

[6] Gi-Tae YEO and Dong-Wook SONG. (2005). "The hierarchical analysis of perceived competitiveness: An Application to Korean Container Ports". *Journal of the Eastern Asia Society for Transportation Studies*, Vol. 6, pp. 866 – 880.

[7] Hui-huang TAI and Cherng-chwan HWANG. (2005). "Analysis of hub port choice for container trunk lines in east Asia". *Journal of the Eastern Asia Society for Transportation Studies*, Vol.6, 907-919.

[8] Li-jin LIU and Ruguo FAN (2006). The Application to Supplier Partner Selection of the Fuzzy Analytic Hierarchy Process Based on the Triangular Fuzzy Numbers Comparative Theory, *Logistics Technology*，2006：Vol. 29，117-121

[9] Sheng HUANG (2006). Something about the necessary factors which the international hub port needs: take Kaohsiung, Hong Kong and Singapore for example, *Special Zone Economic*, Vol. 7: 25-26

[10] Thomson, B.J. (1998). Structure changes in the maritime industry's impact on the inter-port competition in container trade, *International Conference in*

Shipping Development and Port Management.

[11] Yong-bo LU et.al. (2002). A Study of Evaluation on Competition Power for Main Container Transportation Harbors in China. *Journal of Northern Jiao tong University*, vol.26, No.5, 102-105

[12] 이 성 우(2005). A Study on Port Performance Related to Port Backup Area in the ESCAP Region. www.google.com

SELECTIVE EVALUATION OF TRAFFIC-TOOLS BASED ON FUZZY ANALYTIC HIERARCHY PROCESS

Jianyou ZHAO[1], Mourong GAO [2] , Meng DU[1]
[1]School of Automobile, Chang'an University, Xi'an 710064, Shaanxi,China; PH 86-29-82334069; FAX 86-29-82334137; email: jyzhao@chd.edu.cn
[2]School of Automobile and Transportation, Shenzhen Polytechnic, Shenzhen 518055, Guangdong, China; email: mourong999@126.com

Abstract： The selection of traffic-tools is a complicated and multi-objective system. The key to it is how to select the best solution among a series of good ones. This paper selects the traffic-tool evaluating indicators according to the principles of the indicators system. Then the value of evaluating factors is determined by using Fuzzy AHP (Analytic Hierarchy Process) against shortcomings of AHP. Finally, This paper analyzes the advantages and disadvantages of every traffic-tool through accounting. The result of this study shows that Fuzzy AHP is effective for traffic-tool selection evaluation.
Keywords: *Fuzzy AHP（Analytic Hierarchy Process）, traffic-tools, evaluation method, selective evaluation*

1 Introduction

Different traffic-tools have different techno-economy characteristics, reasonable range of use and transport function to meet the residents' needs. Secondly, the reasonable matches of different traffic-tools greatly influence the traffic, the improvement of people's living standards and the economy construction etc. Therefore, it's very important to choose traffic-tools reasonably. But the reasonable choice of traffic-tools is a complex system which has many objects and factors. So an effective evaluation method is needed.

AHP is a kind of means to availably handle the complicated problems which can not be completely analyzed by a metered method. AHP? can express the person's subjective judgment in quantity form. Also it can reveal the antinomy existing in the front and back of the subjective judgment to a certain problem made by people. It is a kind of objective method to sort and comprehend peoples' subjective judgment, and a kind of method that combines quantitative and qualitative analysis. Therefore, AHP can be used to solve the problem of traffic-tool selection. But in using traditional AHP it is very hard to test whether the judge matrix is consistent or not. And the standard (CR<0.1) to judge is not scientific enough. Then the consistency of judge matrix is

quite different from the consistency of mankind's thinking. if we use FAHP in programming decision, the problems mentioned above can be solved. Therefore, in this article there is an attempt to use FAHP in the analysis of citizen's traffic-tool selective evaluating.

2 Principle of Indicator System Confirming

The indicators are absolute number, relative number and average number which can synthetically reflect the circumstance of some aspects of social economic phenomenon. The indicator system means a whole constituted by a series of indicators. An evaluating indicator system should objectively consider all the factors and information from each side as completely as possibly. Whether the evaluating indicator is right will influence the trueness of the result, and weather it can reflect the objective realities. It also influences the choosing decision of traffic-tools and the city traffic programming decisions. In order to choose the right evaluation indicator system, the following principles should be obeyed:

(1)Scientifically and Comprehensively: the evaluating indicator must be built up on the basis of science, and can reflect the actual circumstance objectively. At the same time, the evaluating indicator system should reflect every side. Both the factor of function, the service quantity, and the social and economic factors should be considered in city traffic-tool selection.

(2)Combination of Current and Future Development: In the choosing of city traffic-tools, we have to consider not only the current circumstance but also the need of future development, in order to make the transportation programming have longer adaptability.

(3)Combination of the Qualitative Analysis and Quantitative Analysis: Because of the complexity and universality of the system, the evaluating indicator can't be a measurable indicator completely calculated by mathematical formula. For the sake of the overall comparison, the unsure factors must be taken into consideration, such as the environment factor, the policy factor etc. This can be reflected by qualitative indicator.

(4)Combination of Measurability and Comparability: All the evaluating indicators should be measured by direct or indirect method. And the results should be comparable to evaluate whether it's good or not. To the different evaluating indicators, the units should be unified to a certain one before evaluation.

(5)Combination of Personal Benefits and Social Benefits: the evaluation indicator should reflect personal and social benefits completely. Since traffic-tools serve the residents, they are a part of city traffic system. It will influence the city and the societies.

3 The Assurance of the Indicator System

According to the confirmation principle of indicator system and citizen's actual

circumstance for going out, several evaluating indicators are made certain.

(1) Transportation service

The reason for using traffic-tools is to reach the destination quickly, expediently, safely, punctually and comfortably. Hence, the transportation service function is one of the main indicators in the traffic-tool selection. The transportation service function can be measured by a few indicators of underneath: quick (km/h), safety (accident rate), comfort (condition inside the car, the vehicle vibration and noise, seats, time for waiting), convenience (times for transfer, conditions available), and punctuality.

(2) Economy investment

The economic investment is an important factor in the decision-making of traffic-tool selection carried on by the government and residents. To different traffic-tools, the economic investment of the government and the residents varies. It includes the expense of traffic-tools, the transportation infrastructure expense (construction and maintenance expenses), and the operation expense of traffic-tools.

(3)The Energy consume

Because all traffic-tools consume energy and the global energy shortages are becoming serious, energy saving is very important from the point of both sustainable development and economy. It includes running energy consume and other general energy consume (energy consume of support and management, construction and traffic-tool manufacturing).

(4)Use of land

Today the amount of urban land usage is becoming increasingly serious. Land occupation of traffic-tools is quite an important factor which is worth consideration of government sectors. While the occupation of land can be measured by running occupation and parking occupation.

(5)The influence on environment

In traffic sustainable development, the factor of the environment has become one of the most important ones. The influence of environment is not only what the government section values, but also a factor considered by residents. The evaluation indicator of the environment includes the pollution in the environment caused by traffic-tools, noise produced in traffic running and green house effect.

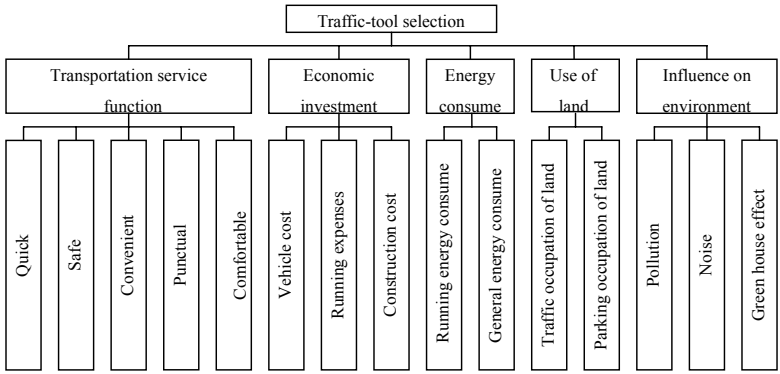


Figure 1. Influence factors of the traffic-tool selection

The above-mentioned indicators contact mutually and control mutually. They constitute a multiple objective and complicated system. Figure 1 is the structure of factors that affect the traffic-tool selection.

4 The Introduction of Fuzzy Analytic Hierarchy Process

4.1 Fuzzy Consistent Matrix and the Relevant Concept

Definition1:

Hypothesis matrix $R=(r_{ij})_{n\times n}$,if fulfils: $0<r_{ij}<1$ $(i=1,2,3\Lambda\ n, j=1,2,3\Lambda\ n)$, then call matrix R as fuzzy matrix.

Definition2:

If fuzzy matrix $R=(r_{ij})_{n\times n}$, if fulfils: $\forall i,j,k$ had or would $r_{ij}=r_{ik}-r_{jk}+0.5$, Then call fuzzy matrix R as fuzzy consistent matrix.

Theorem1:Hypothesis fuzzy matrix $R=(r_{ij})_{n\times n}$ is fuzzy consistent matrix:

(1) $\forall i(i=1,2,3\Lambda\ n)$, then $r_{ii}=0.5$;

(2) $\forall i,j(i=1,2,3\Lambda\ n, j=1,2,3\Lambda\ n)$, then $r_{ij}+r_{ji}=1$;

(3)The element of i'st line and j'st tier sum is *n* in the *R*.

(4)Delete any line and parallelism tier from the *R*, this matrix is still fuzzy consistent matrix.

Theorem2:Fuzzy matrix $R=(r_{ij})_{n\times n}$ is ample and necessary condition of fuzzy consistent matrix, the margin of element in any line and residual line is a constant.

Theorem3:If fuzzy complement matrix $F=(f_{ij})_{n\times n}$ sum by line, sign $r_i=\sum_{k=1}^{n} f_{ik}(i=1,2,3\Lambda\ n)$, the implement is as follows the mathematics transformation:

$r_{ij}=\frac{r_i-r_j}{2m}+0.5$,The matrix that build up from these is fuzzy consistent.

4.2 The Establishment of the Fuzzy Consistent Judgment Matrix

R denotes the comparisons of the importance of related elements at this layer, aiming at a certain element of the layer above. Suppose that there are some relations between the element C in this layer and the $a_1,a_2,a_3\Lambda\ a_n$ in the next layer, then the fuzzy consistent judgment matrix can denote:

C	a_1	a_2	a_n
a_1	r_{11}	r_{12}	r_{1n}
a_2	r_{21}	r_{22}	r_{2n}
a_n	r_{n1}	r_{n2}	r_{nn}

Actual meaning of r_{ij}: when being compared, a_i and a_j have fuzzy relation "…more important than …". Quantity mark degree in table 1.

Table 1. Graduation explanation of the quantity

Mark	Definition	Explain
0.5	of equal importance	Between the two element, same importance.
0.6	a little important	Between the two elements, one less important than another.
0.7	evident importance	Between the two elements, one evident importance than another.
0.8	more important	Between the two elements, one is more important than another.
0.9	extremely important	Between the two elements, one is extremely more important than another.
0.1 0.2 0.3 0.4	Anti- comparison	If compare element a_i with element a_j educe r_{ij} ,then compare element a_i with element a_j educe $r_{ji}=1-r_{ij}$.

4.3 Utilize the fuzzy consistent judgment matrix judgment power of the element

Suppose element $a_1, a_2, a_3 \Lambda\ a_n$ compare in importance chalk up fuzzy consistent judgment matrix $R = (r_{ij})_{n\times n}$, power of element $a_1, a_2, a_3 \Lambda\ a_n$ are $w_1, w_2, w_3 \Lambda\ w_n$, then:

$$r_{ij} = 0.5 + a(w_i - w_j),\ i, j = 1,2,\Lambda\ ,n \tag{1}$$

There into $0.5 < a \le 2.0$, a is a kind of measurement that people knows the difference degrees in the feeling object, But it relation in the evaluate object and. When the quantity or difference degree of evaluate are higher, a can obtain bigger.

When fuzzy judgment matrix is not consistent matrix, the equal mark in equation (1) does not establish strictly, it can get power vector $W = [w_1, w_2, w_3 \Lambda\ , w_n]^T$ by two minimum multiplications. Then solve stipulation programming problem as follows:

$$\begin{cases} \min z = \sum_{i=1}^{n}\sum_{j=1}^{n}[0.5 + a(w_i - w_j) - r_{ij}]^2 \\ s.t.\ \ \sum_{i=1}^{n} w_i = 1,\ w_i \ge 0,\ (1 \le i \le n) \end{cases}$$

5 Applications of the Models

According to the graduation criterion of FAHP law, the relevant experts are invited to differentiate to every indicator of weight, get fuzzy judgment matrix of the second and the third layers.

$$R = \begin{bmatrix} 0.5 & 0.5 & 0.6 & 0.6 & 0.5 \\ 0.5 & 0.5 & 0.6 & 0.6 & 0.5 \\ 0.4 & 0.4 & 0.5 & 0.5 & 0.4 \\ 0.4 & 0.4 & 0.5 & 0.5 & 0.4 \\ 0.5 & 0.5 & 0.6 & 0.6 & 0.5 \end{bmatrix},\quad R_1 = \begin{bmatrix} 0.5 & 0.5 & 0.6 & 0.5 & 0.6 \\ 0.5 & 0.5 & 0.6 & 0.5 & 0.6 \\ 0.4 & 0.4 & 0.5 & 0.4 & 0.5 \\ 0.5 & 0.5 & 0.6 & 0.5 & 0.6 \\ 0.4 & 0.4 & 0.5 & 0.4 & 0.5 \end{bmatrix},$$

$$R_2 = \begin{bmatrix} 0.5 & 0.5 & 0.6 \\ 0.5 & 0.5 & 0.6 \\ 0.4 & 0.4 & 0.5 \end{bmatrix},\quad R_3 = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix},\quad R_4 = \begin{bmatrix} 0.5 & 0.6 \\ 0.4 & 0.5 \end{bmatrix},\quad R_5 = \begin{bmatrix} 0.5 & 0.6 & 0.7 \\ 0.4 & 0.5 & 0.6 \\ 0.3 & 0.4 & 0.5 \end{bmatrix}$$

To judge the formula (1) and the opinion of expert, get $a = 0.8$,then get $w_1 = w_2 = w_5 = 0.25, w_3 = w_4 = 0.125$;under the same circumstances: $w_{11} = w_{12} = w_{14} =$

0.25, $w_{13} = w_{15} = 0.125$; $w_{21} = w_{22} = 0.292$, $w_{23} = 0.416$; $w_{31} = w_{32} = 0.5$; $w_{41} =$

$0.5625, w_{42} = 0.4375; w_{51} = 0.458, = 0.334, w_{53} = 0.208$.

Finally general weighting of each indicator can be achieved as listed in table 2. According to the advantages the of various means of transport, every listed indicator of various means of transport is compared in table 2. Figures in the form: 5 means well, 4 means better, 3 means common, 2 means a little poor, 1 means poor, and 0 means extremely poor.

Table 2.Weight of every index and the indicator of every traffic-tool evaluating

Indicator	Weight	Bus	Light railway	Subway	Car	Motorcycle	Bicycle
Fast	0.0625	2	4	5	4	3	1
Safe	0.0625	3	5	5	2	0	1
It is convenient	0.03125	2	0	0	4	4	4
Punctual	0.0625	1	5	5	2	4	4
Comfortable	0.03125	2	3	4	5	2	0
Vehicle fee	0.073	3	1	1	0	2	5
Operation fee	0.073	3	1	1	0	2	5
Construction fee	0.104	4	1	0	1	2	4
Operate energy consumption	0.0625	2	3	4	0	2	5
Comprehensive energy consumption	0.0625	4	4	4	0	2	5
Traffic occupation of land	0.0703	4	4	5	0	1	2
Parking occupation of land	0.0547	3	4	5	0	1	2
Pollution	0.1145	3	4	4	0	1	5
Noise	0.0835	1	0	1	1	1	4
Greenhouse effects	0.052	2	4	4	0	2	5

The calculation of the comprehensive evaluating indicators of various means of transport can be as follows:

$$T_i = \frac{m\sum_{j=1}^{n} W_j B_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n} W_j B_{ij}}$$

Where： T_i——the comprehensive appraised indicators of various means of transport; W_i——the indicators weight of various means of transport; B_{ij}——the appraisement indicators of various means of transport; m ——the various means of transport number of appraise; n ——the indicator appraised number.

The calculated comprehensive appraised indicators of various means of transportation according to table 2 and formula shown in table 3.

Table 3. Calculation results of every traffic-tool evaluation indicators

Appraise target	Bus	Light railway	Subway	Car	Motorcycle	Bicycle
Comprehensive evaluation indicator	1.083	1.130	1.234	0.388	0.688	1.476
Sequencing	4	3	2	6	5	1

6 Conclusion

Choosing and evaluating the means of transport is a systematic project. There are a lot of factors influencing the choice. To evaluate the means of transport in a correct, all-round way, primary and secondary factors should be distinguished. The fuzzy level analytic approach adopted in this article can well evaluate the means of transport effectively.

References

[1]Alireza Banai-Kashani. Travel demand (modal split) estimation by hierarchical measurement[J]. Journal of advanced transportation. 1984, 18: 1.

[2]Wang, Huiping. The great sustainable city traffic structure development model studying --Take Xi'an as an example [D]. Northwest University. 2004.6.

[3]You, Zhenwei. Research of assigning the mode of the transporting amount - Analyse the application [D] of stratum's procedure law. Taiwan National Cheng Kung University. 1992.6.

[4]Zhang, Jijun. Fuzzy level analytic approach [J]. Fuzzy system and mathematics. 2000, 14(2) : 80- 88.

[5]Zheng, Shuangzhong, Gao, Yongting, Chen ,Baozhi. Research of the fire risk evaluating indicator system of urban architecture. [J] Fire control science and technology. 2001,5: 12-13.

[6]Zhu, Minghe.Xiong, Deqi, Yin ,Peihai. Influence the shipping greasy dirt accident to damage and compensate for every factor fuzzy level analytic approach [J]. Environmental protection of the traffic. 2001, 22(2) : 11- 13.

Predicting the Peaking of Holiday Traffic (Victoria Day Example)

Zhaobin Liu[1], Satish Sharma[2]

[1]Zhaobin Liu, Ph. D., P. Eng., Faculty of Engineering, University of Regina, Regina, SASK, Canada, S4S 0A2. Phone: (306) 585-4958 Fax: (306) 585-4855, Email: zhaobin.liu@gmail.com

[2]Satish Sharma, Ph. D., P. Eng., Professor and Associate Dean, Faculty of Engineering, University of Regina, Regina, SASK, Canada, S4S 0A2. Phone: (306) 585-4553, Fax: (306) 585-4855. Email: Satish.Sharma@uregina.ca

Abstract
In developed and fast developing countries, rising standards of living and the trend towards shorter working hours have significantly changed people's traveling behaviour. Literature reported that, during holiday periods, there were usually substantial increases in traffic volumes on highways. A clear understanding of the holiday traffic characteristics and a reasonable prediction of the upcoming high traffic volumes would significantly benefit both researchers and practitioners who are responsible for the planning, design, operation, and management of transportation networks. However, research focusing on holiday traffic has been minimal to date. The existing traffic prediction methods mainly aim on the flow during regular times. This paper is intended to first investigate the variation characteristics of holiday traffic. Then, the potential prediction errors using popular predicting methods such as time series are discussed. At last, a non-parametric method is proposed to predict the holiday traffic for different types of highways. On the basis of traffic volume data from major highways in Alberta, Canada, it is found that the performances of the proposed method are consistent and reasonable for different holiday periods and various types of roads.

Key words: Holiday Traffic, Peaking, Prediction, Non-parametric Method.

1. Introduction

In developed and fast developing countries, rising standards of living and the trend towards shorter working hours have significantly changed people's traveling behaviour. Literature reported that, during holiday periods, there were usually substantial increases in traffic volumes on highways. For example, in North America, the US Bureau of Transportation Statistics (2003) reported that the number of long distance trips increased by more than 50% during the Thanksgiving period, and about 90% of this travel was via personal vehicles. In European counties such as Germany, it was reported that about 40% of all trips and 50% of all person-kilometres per year are in leisure and holiday traffic, two-thirds of the population goes on holiday trips at least once a year, and 10% of the population regularly goes on holiday trips (Chlond 2006).

Similar findings were also reported in other countries such as Japan (Catbagan and Nakamura 2006), and India. (Muralidhar *et al*. 2006)

A clear understanding of the holiday traffic characteristics and a reasonable prediction of the upcoming traffic volumes would significantly benefit both researchers and practitioners who are responsible for the planning, design, operation, and management of transportation networks. However, research focusing on holiday traffic has been very limited to date. The existing traffic prediction methods mainly aim on the flow during regular times. This paper is intended to first investigate the variation characteristics of holiday traffic. Then, the appropriateness of popular predicting methods such as time series with respect to holiday travel are discussed. At last, a non-parametric method, k nearest neighbour method (k-NN), is proposed to predict the holiday traffic for different types of highways.

2. Study Data

The data employed in this study was collected by permanent traffic counters (PTC) on major highways in Alberta, Canada over the most recent years. Alberta Infrastructure and Transportation (the agency responsible for provincial transportation in Alberta) currently employs approximately 360 PTC sites on its 28,660-kilometer highway system. At each PTC site, there is an automatic traffic recorder (ATR) with magnetic induction loops embedded in the road surface. These recorders were programmed to count the number of vehicles by direction on an hourly basis. Since the directional hourly volumes are the basic element in traffic database, the minimum time interval applied in this study is one hour.

Due to the fact that holiday traffic characteristics are different for various types of roads, the PTCs in the study province were first classified into groups based on considerations of the trip purpose, the trip length distribution, as well as temporal volume variations, as proposed by Sharma et al. (1986). The grouping process resulted in five classifications representing different types of roads, and these road classes were labelled as commuter (CM), regional commuter (RCM), rural long distance (RLD), recreational (REC) and highly recreational (HIGHREC).

From each road group, one representative PTC site was selected for detailed analysis in this study. The locations of these PTC sites are given in Figure 1. The first three digits of each PTC code represent the highway number and the remaining digits identify the counter number. It can be seen that the CM site 016186 is close to the provincial capital – the city of Edmonton, the RCM site 001145 is in the vicinity of the town of Strathmore. The RLD site 016061 is located on the major highway #16 which connects the city of Edmonton and the neighbouring province – British Colombia. The REC site 001025 is on TransCanada Highway #1 between the city of Calgary, the fourth largest city in Canada, and the Banff National Park, a famous recreational site. The HIGHREC site 093001 is near the scenic Lake Louise within the Banff National Park.

The monthly traffic variation patterns of these five selected sites, which are represented by the monthly factors (the ratio of monthly average daily traffic to AADT), are shown in Figure 2. The distribution for the CM site appears to be relatively flat throughout the year. The pattern of the RCM site is similar to the CM

site, with slightly higher summer traffic. The RLD and REC sites, however, show higher peaks in summer and lower troughs in winter. The HIGHREC site has the highest peaks during the summer months.

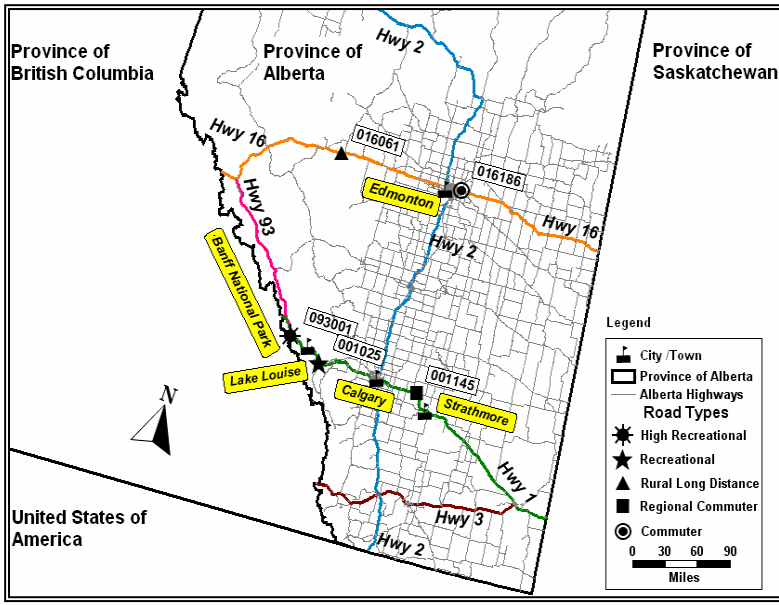


Figure 1 Locations of Selected PTC Sites

3. The Peaking Feature of Holiday Traffic

In order to investigate the variation of holiday traffic as compared to the normal flow, all the daily volumes in a year were plotted for the study sites against the sequence of time. It was found that the holiday traffic peaking phenomena were evident for different road groups, especially for the sites located in rural areas.

Figure 3 (a) shows the daily volume variations over a one-year period in 2002 for the typical HIGHREC site of 093001, which is near the famous scenic Lake Louise in the Banff National Park. This site demonstrates a predominant weekend traffic pattern, where the weekend volumes are considerably higher than the weekday traffic in both directions throughout the year. From close observations on this Figure, it is found that the traffic peaking features on weekend days are significantly different by direction. The westbound traffic, coming from the major population centre (the City of Calgary) and heading to the recreational areas (Banff National Park), always peaks on Fridays. When a holiday is observed, however, a remarkable peak on Friday that is double or triple the normal weekend flow is detected. A possible explanation to the traffic peaking in this direction is that people usually leave their homes and proceed to social and recreational destinations on Fridays to start their vacations, and, hence, substantial

traffic increases are generated. In the theory of travel demand modeling, a trip is usually defined as a one-way movement from the production zone to the attraction zone, or vice versa (Kadiyali 2005). Following the same ideology, the direction that predominantly serves the traffic from the population centre (production zone) to the recreational area (attraction zone) at the beginning of a holiday period is called outbound (OB) hereafter in this paper.

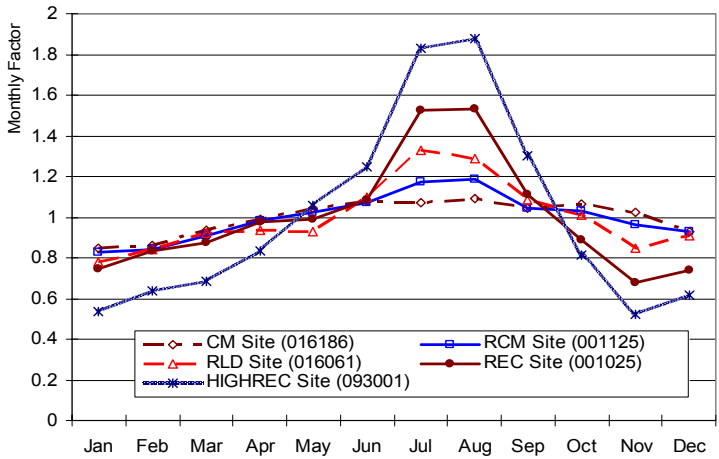
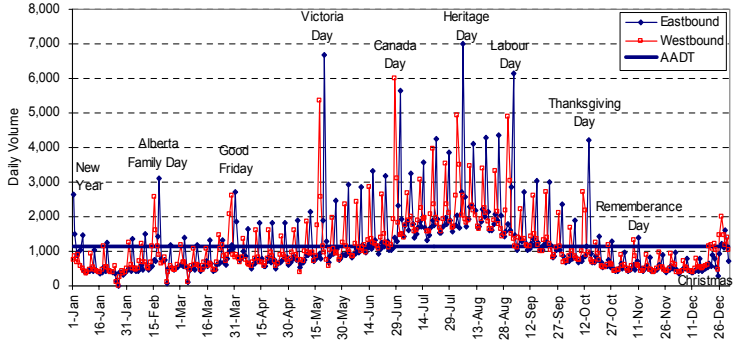


Figure 2. Monthly Traffic Variations of Selected PTC Sites

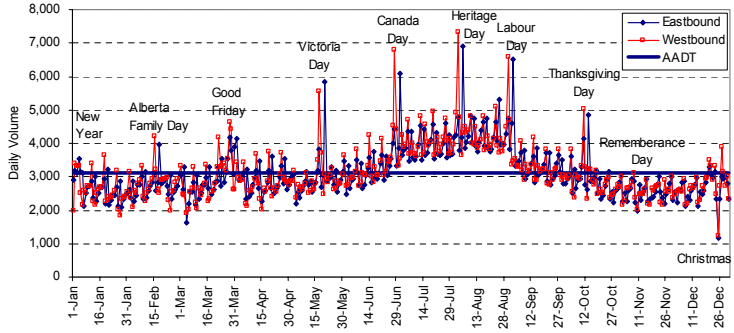
The peaking of traffic volumes can be similarly observed in the eastbound direction, where the traffic moves from the park to the city. The traffic usually peaks on Sundays when there is no holiday. During holiday periods, the peaks are found generally switching to Mondays with a substantial increase, while Sunday volumes becoming lower than those usually observed. This direction's traffic peaking can be most likely explained by the fact that most people return home from vacation sites on Mondays. Hence, this direction that mainly serves the traffic returning from the recreational area (attraction zone) to the population centre (production zone) at the end of a holiday period is called inbound (IB) hereafter in this paper.

The directional peaking of holiday traffic was similarly observed for other sites that are located in rural areas. Taking the typical RLD site of 016061 as an example, Figure 3 (b) shows that the increases of traffic volumes are all substantial during holiday periods at this site. Moreover, there is also one direction that can be named as outbound with typical traffic peaking on Fridays prior to holiday weekends, and the other direction that can be called inbound represented by the high volumes on holiday Mondays.

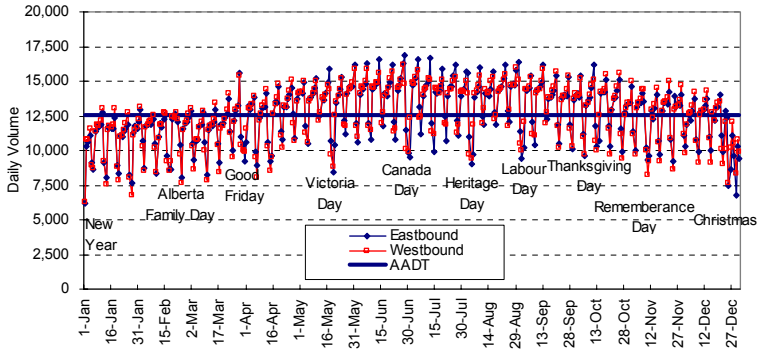
When compared to other types of roads, Figure 3 (c) shows that the typical CM site of 016186 has a very unique daily traffic variation pattern. Because this site predominately serves work-business trips and very little social and recreational travel,



(a) Highly Recreational Road Site 093001, 2002 Data



(b) Rural Long Distance Road Site 016061, 2002 Data



(c) Commuter Road Site 016186, 2002 Data

Figure 3 Daily Volume Variations of Selected Sites

the daily volumes in both directions of this site distribute nearly evenly at the upper and lower sides of the line of AADT throughout the year. Since fewer commuting trips are made when there is a holiday, the traffic in both directions usually decreases significantly on holiday days.

4. Performance of Time Series Models on the Prediction of Holiday Traffic

Since traffic volume generally follows a time sequence, time series models, such as Auto-Regressive Integrated Moving Average (ARIMA) family models, are popularly used in the practice of traffic prediction. The ARIMA models can be simply understood as linear estimators regressed on past values of the modeled time series (the autoregressive terms) or past prediction errors (the moving average terms), and can be applied recursively to produce single or multiple forecasts. For a more detailed discussion, comprehensive time analysis text such as Brockwell and Davis (1996) or Fuller (1996) can be referred.

With regard to the performance of ARIMA models on normal traffic flow, there are many successful examples in the literature (such as Smith and Demetsky (1997), and Williams and Hoel (2003)). However, when dealing with holiday traffic, significant concerns on these models exist. For example, Redfern et al (1993) pointed out that these methods had problems in the presence of non-stationary. The changes of traffic seasonal structure due to various reasons (such as holiday events) would definitely bring problems to the estimation. Kirby and Watson (1997) and Williams (1999) found when trying to apply ARIMA model to the data with summer holiday season included, the ARIMA model did not have good results.

In order to examine the adaptability of ARIMA models in predicting the holiday traffic, these models were applied in this study to predict the inbound traffic of the RLD site (016061) during Victoria Day period in 2003. Every year the Victoria Day in Canada is observed on the Monday preceeding May 25th. In this experiment, the inbound hourly volumes from Thursday prior to the holiday to Wednesday after it were first assumed missing, and then estimated by the ARIMA models using MINITAB Statistical Software 13.1. For each day of the estimating period, the hourly volumes are estimated using data from the same days of the eight previous weeks.

Figure 4 shows the performance of ARIMA models by plotting the actual and estimated volumes. It can be found that, for most days of the predicting period, the differences between the actual and estimated volumes are apparent. On the Victoria Day, especially, the peaking of traffic is missed and the variation of traffic is significantly underestimated. The poor performance of ARIMA models may be attributed to the theory foundation of time series models. These models are linear estimators regressed on past values of a time series. Hence, homogeneous patterns are necessary for making good estimations. In this study, the ARIMA models were trained by eight previous weeks of data when most of the time is non-holidays. Because the holiday traffic patterns can be significantly different to non-holiday patterns, it explains that the ARIMA models are not appropriate in predicting holiday travel.

4. The Proposed Method

A nonparametric regression method, namely, the k-nearest neighbour (k-NN) method, has been successfully applied in forecasting short-term traffic flows. Smith and Demetsky (1997) compared the k-NN method with the historical, time series, and

artificial neural network models. They identified the k-NN method as the one that produced the lowest traffic forecasting errors. An important merit of the k-NN method, as reported by Oswald et al. (2000), is that it maintains the uniqueness of every observation, making it particularly useful in modeling unusual situations. Considering the past successes, the k-NN method is described in this section with regard to the prediction of holiday traffic.

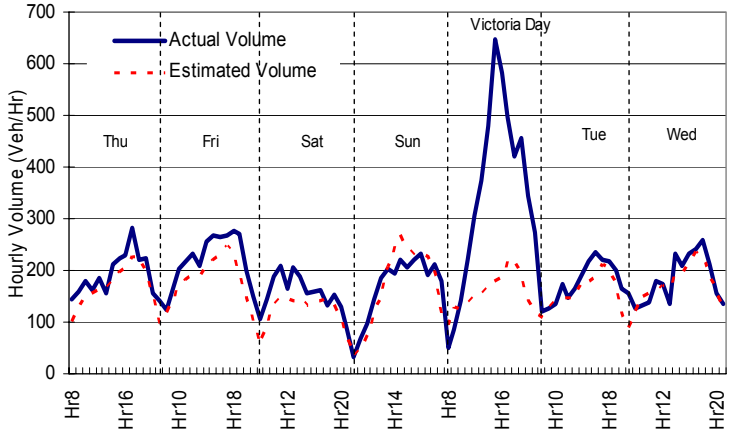


Figure 4 Estimated Volumes v.s. Actual Volumes Using Time Series (ARIMA) Models

5.1 State Vector in the k-NN Method

The k-NN method is a data-driven nonparametric regression technique, also known as memory-based learning or instance-based learning (Hardle 1990). The basic idea behind this method is to match the current input variables with similar historical records (Davis and Nihan 1991). In practice, the input variables are usually multiple and they are typically gathered together into a feature vector to define the state of a system. For the traffic data that are time series in nature, a state vector in k-NN methods usually defines each record with the volume at time lag of t, t-1, t-2, …., etc. According to Smith et al. (2002), a hybrid state vector x(t) is similarly defined in this study:

$$x(t) = \left[V(t), V(t-1), V(t-2), V_{hist}(t), V_{hist}(t+1)\right] \text{ ..... [Eq 1]}$$

where, *V(t), V(t-1)* and *V(t-2)* are the traffic flow at time interval t, t-1, and t-2, respectively. $V_{hist}(t)$ and $V_{hist}(t+1)$ are the historical average volumes at the time-of-day and day-of-the-week associated with time interval t and t+1. In cases of predicting holiday volumes, the historical average is calculated based on volumes that are from the same day and hour during the same holiday period in the past.

5.2 Determination of the k Nearest Neighbours

Once the state vector is determined, the next step is to search for the k nearest neighbours that will be used in prediction. This procedure is a central component of many applications of k-NN method, including knowledge discovery, data mining, pattern recognition, classification, machine learning, data compression, multimedia databases, document retrieval, and statistics (Maneewongvatana and Mount 1999), The 'closeness' between observations in a multivariate space can be commonly measured using the Euclidean Distance, and then the k observations with the shortest Euclidean Distances are recognized as neighbours. The Euclidean Distance is simply the geometric distance in the multidimensional space. In this case, the Euclidean distance ($d(p,q)$) from historical record p to the current condition q can be written as follows:

$$d(p,q) = \sqrt{\begin{array}{l}\left[V_p(t) - V_q(t)\right]^2 + \left[V_p(t-1) - V_q(t-1)\right]^2 + \left[V_p(t-2) - V_q(t-2)\right]^2 + \\ \left[V_{hist_p}(t) - V_{hist_q}(t)\right]^2 + \left[V_{hist_p}(t+1) - V_{hist_q}(t+1)\right]^2\end{array}} \quad \ldots \text{[Eq 2]}$$

5.3 Generation of the Forecast

With the determination of the k-nearest neighbours, the most straightforward approach to generate the forecast, y′(t), for the traffic volume of the next hour is to compute the simple average of the output y(t) values corresponding to the k neighbours. The output y(t) for each state observation is the volume at time interval t+1 relative to the selected neighbour. The formula to forecast y′ (t) is given in Eq 3.

$$y'(t) = \left(\frac{1}{k}\right)\sum_{i=1}^{k} y_i(t) \quad \ldots\ldots\ldots\

5.4 Performance of k-NN method on Holiday Traffic

In order to examine the performance of k-NN method on holiday traffic, this method was applied to the data used in section 4 when ARIMA models did not perform well. The estimated and actual values using the k-NN method are graphically illustrated in Figure 5. It can be found that, regardless of whether the traffic is high or low and normal or abnormal, the estimates using k-NN method fit well the actual hourly volumes in the VIC week.

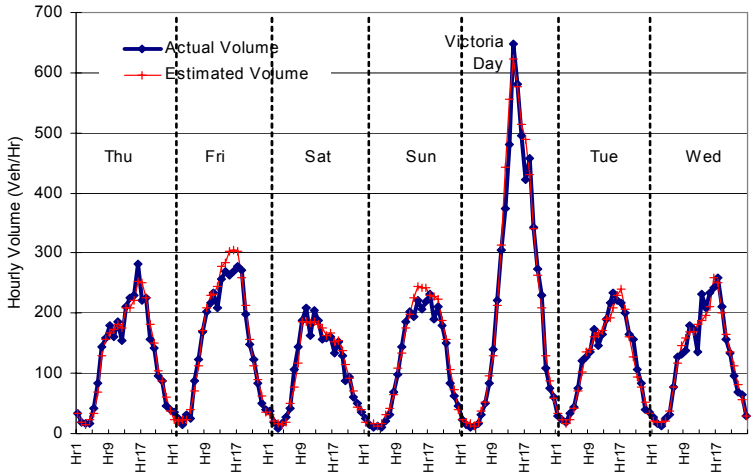


Figure 5 Estimated Volumes v.s. Actual Volumes Using the Proposed Model

With respect to the performances of the k-NN method on other types of roads, this method was similarly applied to other study sites that are representative for four different road groups including CM, RCM, REC, and HIGHREC. Table 1 provides the prediction errors for each direction of the study sites when the traffic data during the same VIC week were assumed missing. It can be found that for CM, RCM, and REC roads, the values of different error measures such as 0-0.4% for the minimum absolute error (MinARE), 6-9% for mean absolute relative error (MARE), 5-8% for 50 percentile error (E50) and 14-20% for 90 percentile error (E90), are very similar to those for the RLD road. The errors seem higher for the HIGHREC site. However, graphical examination indicated that the estimated values were also very close to the actual values. The large estimation errors for this road were likely due to the low volume level on this road (AADT = 2,290) as compared to those on other types of roads whose AADT ranges from 6,000 to 25,000. In short, good performance of the k-NN method is demonstrated for different types of roads.

5. Concluding Remarks

This paper shows that substantial increases in traffic volume can be observed during various holiday periods and for different types of roads. Since the existing traffic prediction techniques, such as time series models, mainly aim on the flow during regular times, they are inappropriate for the holiday traffic. The k-NN method is a data-driven nonparametric regression technique that has been identified to be useful in modeling unusual situations. This study applied the k-NN method to predict the traffic data during Victoria Day period, and the prediction results show that the performances of the k-NN method are consistent and reasonable for various types of roads. Good performances of the k-NN method for the traffic during other holiday periods can also be expected.

Table 1. Prediction Results Using the k-NN Method for Different Types of Roads (Imputation Period: Victoria Day Week)

	CM (016186)		RCM (001145)		RLD (016061)		REC (001025)		HIGHREC (093001)	
	EB	WB	EB	WB	WB	EB	EB	WB	EB	WB
MinARE	0.41	0.00	0.00	0.00	0.00	0.00	0.23	0.26	0.00	0.00
MARE	8.40	6.45	8.82	6.90	9.00	7.90	8.43	7.80	17.42	15.73
E50	7.97	5.57	8.02	6.31	7.09	6.08	6.74	6.14	15.63	14.86
E90	14.89	13.83	16.46	14.77	17.91	16.05	21.67	16.58	34.47	27.17

Note: EB, eastbound; WB, westbound.

6. Acknowledgement

The authors are grateful to National Sciences and Engineering Research Council of Canada (NSERC), Saskatchewan Government Insurance (SGI) and the Faculty of Graduate Studies at the University of Regina for their financial support. The authors would also like to thank Alberta Infrastructure and Transportation for providing the data used in this study.

7. References

Brockwell, P. J. and Davis, R. A. (1996). *Introduction to time series and Forecasting*, Springer, New York.

Catbagan J. L., Nakamura H. (2006). "Performance Measure Evaluation for Japan Two-Lane Expressways", *Transportation Research Board Annual Meeting,* CD-ROM Paper, Washington D.C

Chlond B., Manz W., Last J, Zumkeller D. (2006) "Long-distance travel in a longitudinal perspective – The INVERMO approach in Germany", *Transportation Research Board Annual Meeting,* CD-ROM Paper, Washington D.C.

Fuller, W. A., (1996). *Introduction to Statistical Time Series*, 2nd Edition, Wiley, New York.

Hardle W. (1990). *Applied nonparametric regression*, Cambridge University Press.

Kadiyali L.R. (2005) Traffic Engineering and Transport Planning, Khanna Publishers, 2-B, Nath Market, Nai Sarak, Delhi-110006,

Kirby H. R., Watson S. M., and Dougherty M. S., (1997) "Should we use neural networks or statistical models for short-term motorway traffic forecasting?" *International Journal of Forecasting*, Vol. 13, No. 1, , pp. 43-50.

Maneewongvatana S. and Mount M.D., (1999). "Analysis of Approximate Nearest Neighbor Searching with Clustered Point Sets", *Algorithm Engineering and Experimentation: International Workshop ALENEX'99*, Baltimore, MD, USA, January 15-16.

Muralidhar B., Mathew T. and Dhingra S. (2006). "Development and Analysis of Time-Space Diary for a City of a Developing Country", *Transportation Research Board Annual Meeting,* CD-ROM Paper, Washington D.C.

Oswald K. R, Scherer W. T., and Smith B. L., (2000) "Traffic Flow Forecasting Using Approximate Nearest Neighbor Nonparametric Regression," *Final Report of ITS Center Project, Traffic Forecasting: Non-parametric Regressions*, Center for Transportation Studies, University of Virginia.

Redfern E.J., Watson S.M., Clark S.D., Tight M.R. and Payne G.A., (1993) "Modeling Outliers and Missing Values in Traffic Count Data Using the ARIMA Model," *ITS Working Paper 395*, Institute for Transport Studies, University of Leeds, March.

Sharma S.C. Lingras P.J. Hassan M.U. and Murthy N.A. (1986) "Road Classification According to Driver Population", *Journal of Transportation Research Board:* T*ransportation Research Record No. 1090.* Transportation Research Board, National Research Council, Washington D.C. pp. 61-68.

Smith B.L and Demetsky M.J., (1997) "Traffic Flow Forecasting: Comparison of Model Approaches," *Journal of Transportation Engineering*, Vol. 123 No. 4, , 261-266.

Smith B., Williams B.M., and Oswald R.K., (2002) "Comparison of Parametric and Nonparametric Models for Traffic Flow Forecasting," *Transportation Research Part C,* Vol 10, , pp 303-321.

US Bureau of Transportation Statistics, (2003) "U.S Holiday Travel", *America on the Go*, *Findings from the National Household Travel Survey*, Nov.

Williams B.M, (1999) *Modeling and Forecasting Vehicular Traffic Flow as a Seasonal Stochastic Time Series Process*, Ph.D Thesis, University of Virginia, Charlottesville, Va.

Williams B.M., Hoel L.A. (2003) "Modeling and Forecasting Vehicular Traffic Flow as A Seasonal ARIMA Process: Theoretical Basis and Empirical Results", *Journal of Transportation Engineering*, Vol. 129, No.6, Nov., pp: 664-672.

Research and Application of Cross-Nested Logit Model in Traffic Mode-split

Gu WANG[1] Xiucheng GUO[2] Jie YANG[3]

[1]School of Transportation, Southeast University, P.O. BOX 210096, Nanjing; Email: transyscn@gmail.com
[2]School of Transportation, Southeast University, P.O. BOX 210096, Nanjing; Email: SeuGuo@163.com.
[3]School of Transportation, Southeast University, P.O. BOX 210096, Nanjing; Email: honest-orange@163.com..

Abstract: Due to the simplicity and convenience of logit model, it is widely used in the field of economics and transportation. However, the precondition that the random term should be independently and identically distributed, limits the scope of its application. Focusing on the problem above and based on the good characteristics of logit model, this paper probes into its ameliorating mechanism by way of introducing the Function G(y), and studies the inherent relations between the multinomial logit(MNL) model, the nested logit (NL) model and the cross-nested logit (CNL) model, which reflects the theoretical advantages of the CNL model in dealing with the discrete choice problems. At length, the advantages and defects of the three models are compared through an example of the engineering case. The result indicates that the CNL model is more suitable for the trip mode selection than the others.
Keywords: *Discrete choice models; Multinomial logit model; Nested logit model; Cross-nested logit model; traffic modal split*

1 Introduction

Discrete choice models have been used to represent the selection of one among a set of mutually exclusive alternatives, especially in the transportation field presently. The multinomial logit (MNL) model (McFadden, 1973) is the most widely used choice model due to its simple mathematical structure and ease of estimation. However, the independence of the irrelevant alternatives property of the MNL model produces biased estimates and incorrect predictions as to the alternatives which are highly interrelated.

The most widely known modification of the MNL model is the nested logit (NL) model, which was firstly applied to the selection of the trip modes by Willians in 1977. The NL model is superior in theory since it allows pairs or groups of alternatives to be correlated. However, the remaining restrictions on the equality of cross-elasticity between pairs of alternatives may be unrealistic in many cases. Based on the factors above, the research which probed into the cross-nested logit (CNL) model has carried on.

2 Function G(y) Definition

Discrete choice models are based on the stochastic utility theory which is the measurement of satisfaction. Two basic assumptions have been given as follows: (1) the decision-maker chooses the alternative that provides the greatest utility; (2) the factors related to the decision-maker's utility are jointly determined by the attributes of the alternatives and the attributes of the decision-maker. Utility is decomposed as $Y_i = V_i + \varepsilon_i$, where V_i is the deterministic utility that can be observed, and ε_i is the stochastic variable that can not be measured. The stochastic term ε_i follows the Gumbel distribution.

AS to the logit model, N represents the number of available alternatives, $y_i = e^{V_i}$, and V_i is the deterministic part of the utility function associated to alternative i .

We define $G(y_1, y_2 \ldots y_n)$ as a non-negative differentiable function with the following properties:

1. G is homogeneous of degree $\mu > 0$, that is $G(\alpha y) = \alpha^{\mu} G(y)$;

2. $\lim_{y_i \to +\infty} G(y_1, y_2 \ldots y_n) = +\infty$, for each i=1,2…n;

3. The kth partial derivative with respect to k distinct y_i is non-negative if k is odd and non-positive if k is even; that is, for any distinct indices $i_1, \ldots, i_k \in \{1, .., n\}$, we have

$$(-1)^k \frac{\partial^k G}{\partial x_{i_1} \ldots . \partial x_{i_k}}(x) \leq 0 \tag{1}$$

The homogeneity of G and Euler's theorem give

$$P(i \mid m) = \frac{y_i \dfrac{\partial G_m(y_1 \ldots y_n)}{\partial y_i}}{\mu G_m(y_1 \ldots y_n)} \tag{2}$$

Eq.(2) is the probability of selecting alternative i within the choice set m of a given choice maker. Generally, the stochastic utility function y is a one-order homogenous equation that $\mu = 1$.

3 Derivation of the CNL model

1. As to the MNL model (see Fig.1), $G(y)=\sum_{i=1}^{n}\left(y_i{}^{\mu}\right)$.

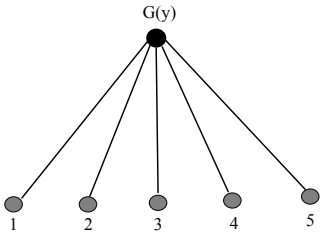


Fig.1 Typical structure of the MNL model

The basic assumption of $y_i = V_i + \varepsilon_i$ is that ε_i follows the Gumbel distribution which is independently and identically distributed (IID). The deterministic utility can be depicted as $V_i = \beta_1 x_1 + ... + \beta_k x_k = \beta' x$, where $x = (x_1,...,x_k)'$ is a vector of variables that relates to an alternative as faced by decision-maker, and $\beta = (\beta_1,...,\beta_k)'$ is a parametric vector to be estimated. For $\mu = 1$, from Eq.(2) we have

$$P_i = \frac{e^{V_i}}{\sum_{i=1}^{n} e^{V_i}}, \tag{3}$$

which is the logit choice probability.

For any two alternatives i and j, the ratio of the logit probabilities is $P_i / P_j = \frac{e^{V_i}}{e^{V_j}} = e^{V_i - V_j}$. This ratio does not depend on any alternatives other than i and j. That is, the relative odds of choosing i over j is the same no matter what other alternatives are available or what the attributes of the other alternatives are.

2. It can be seen in Fig.2, which shows, a typical two-level nested logit structure with $G(y)=\sum_{m=1}^{M}\left(\sum_{k\in N_m}\left(y_k{}^{\mu_m}\right)\right)^{\mu/\mu_m}$, where μ_m is the dissimilarity parameter for the virtual nest m_1, m_2, and $\mu \le \mu_m$. As well as the MNL model, $1 = \mu \le \mu_m$. From Eq.(2) we have

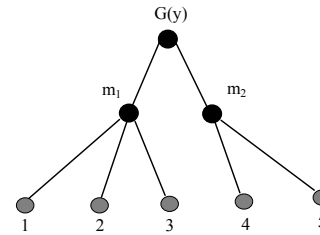


Fig.2 Typical structure of the NL model

$$P(i)=\frac{\sum_{m=1}^{M}\left[\left(e^{V_i}\right)^{\mu_m}\left(\sum_{k\in N_m}\left(e^{v_K}\right)^{\mu_M}\right)^{\frac{1}{\mu_M}-1}\right]}{\sum_{m=1}^{M}\left(\sum_{k\in N_m}\left(e^{V_k}\right)^{\mu_m}\right)^{1/\mu_m}} \tag{4}$$

Where Nm is the number of all alternatives included in the virtual nest m. The function Eq.(4) is expressed as the conditional probability of alternative i in the virtual nest m. The selection probability of the virtual nest is calculated as the MNL model, that is, $P(m)=\frac{\left(\sum_{k\in N_m}\left(e^{v_K}\right)^{\mu_M}\right)^{1/\mu_m}}{\sum_{m=1}^{M}\left(\sum_{k\in N_m}\left(e^{V_k}\right)^{\mu_m}\right)^{1/\mu_m}}$ and $P(i)=\sum_m P(i\mid m)P(m)$. The nested logit model with a virtual nest is appropriate when the set of alternatives faced by a decision-maker can be partitioned into subsets. By inducing the dissimilarity parameter μ_m , it makes improvement on the MNL model.

3. The typical two-level CNL model is showed in Fig. (3), where the alternative 2,3,5 belong to both nest m_1 and nest m_2, therefore, $G(y)=\sum_{m=1}^{M}\left(\sum_{k\in N_m}\left(\alpha_{km}{}^{1/\mu}y_k\right)^{\mu_m}\right)^{\mu/\mu_m}$, $1=\mu$, and $\mu\le\mu_m$. Any alternative k is with different degrees of membership α_{km}, which is non-negative and $\sum_{m=1}^{M}\alpha_{km}=1$. From the function Eq. (2), we have

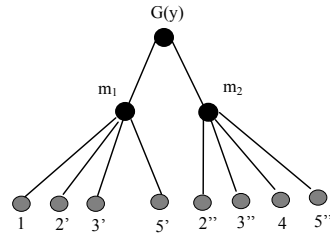


Fig.3 Typical structure of the CNL model

$$P(i)=\frac{\sum_{m=1}^{M}\left[\left(\alpha_{im}e^{V_i}\right)^{\mu_m}\left(\sum_{k\in N_m}\left(\alpha_{km}e^{v_K}\right)^{\mu_M}\right)^{\frac{1}{\mu_M}-1}\right]}{\sum_{m=1}^{M}\left(\sum_{k\in N_m}\left(\alpha_{km}e^{V_k}\right)^{\mu_m}\right)^{1/\mu_m}} \tag{5}$$

In the CNL model, an alternative can belong to more than one nest with different non-zero degrees of membership within a nested structure. The allocation parameter α_{km} represents the effect that the alternative on the lower level contributes to the nest on the upper level. Eq. (5) is a complicated form of the NL model and the MNL model, from which Eq. (3) and Eq. (4) can be derived. The details are not given here.

4 Case comparative research

Considered a group of data given in Table 1, which comes from trancad4.5, MNL.WRK; there are three trip modes: auto, bus and carpool. The parking cost (ParkCost), the bus fee (Buscost), the travel time (Autott、CollTime 、Bustt), the car ownership (Auto2P、Auto1P), the income (Income) and the gender (Gender) are defined as six selection variables. In order to have comparative conditions, an algorithm based on confidence, a region is designed to solve the three models afore-mentioned, and the results are showed in Table 2.

With the execution results of the models shown in Table 2, it is obvious that the complexity of the models augments the running time, from 1 to 5 seconds. The results of R^2-test increase from 0.573485 to 0.575864, and the results of likelihood ratio test go from 1217.23 to 1222.28, which indicate the models are in agreement with the data observed.

Table1. The setting of selection variables for utility function

Utility	Inherent dumb elements		Inherent variables			Partially common variables	Characteristic of the traveler			
Auto	1	0	ParkCost	0	0	Autott	Auto2P	Auto1P	Income	0
Bus	0	1	0	Buscost	0	Bustt	0	0	0	Gender
CarPool	0	0	0	0	CollTime	Autott	0	0	0	0
parameter	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9	β_{10}

Table2. Execution results of the models

Model	Number of data observed	Number of variables	Log-likelihood ratio	Running time(s)	Likelihood ratio test	R^2-test
MNL	966	10	-452.643	1	1217.23	0.573485
NL	966	12	-450.37	3	1221.78	0.575626
CNL	966	11	-450.118	5	1222.28	0.575864

Table3. Parameter calibration for the three models of Auto

Selection variables		The MNL model			The NL model			The CNL model		
		Calibration value	Standard deviation	t-test	Calibration value	Standard deviation	t-test	Calibration value	Standard deviation	t-test
Inherent dumb elements	1	-6.97	0.74	-9.46	-5.75	0.82	-7.01	-5.64	0.60	-9.38
	0									
Inherent variables	ParkCost	-0.30	0.06	-4.68	-0.29	0.06	-4.93	-0.27	0.06	-4.96
	0									
	0									
Partially common variables	Autott	-0.06	0.01	-4.18	-0.05	0.01	-3.57	-0.05	0.01	-4.82
Characteristics of the traveler	Auto2P	4.86	0.39	12.53	4.31	0.44	9.70	3.75	0.34	11.05
	Auto1P	4.10	0.34	12.20	3.74	0.36	10.41	3.11	0.29	10.59
	income	0.02	0.00	5.71	0.02	0.00	5.76	0.01	0.00	5.01
	0									

Table4. Parameter calibration for the three models of Bus

Selection variables		The MNL model			The NL model			The CNL model		
		Calibration value	Standard deviation	t-test	Calibration value	Standard deviation	t-test	Calibration value	Standard deviation	t-test
Inherent dumb elements	0									
	1	-2.56	0.84	-3.06	-1.88	0.80	-2.36	-2.49	0.62	-4.02
Inherent variables	0									
	Buscost	-3.05	0.78	-3.90	-2.96	0.76	-3.89	-1.94	0.53	-3.66
	0									
Partially common variables	Bustt	-0.06	0.01	-4.18	-0.05	0.01	-3.57	-0.05	0.01	-4.82
Characteristics	0									
of the traveler	0									
	0									
	Gender	0.76	0.20	3.72	0.72	0.20	3.62	0.59	0.15	3.92

Table5. Parameter calibration for the three models of Carpool

Selection variables		The MNL model			The NL model			The CNL model		
		Calibration value	Standard deviation	t-test	Calibration value	Standard deviation	t-test	Calibration value	Standard deviation	t-test
Inherent dumb elements	0									
	0									
Inherent variables	0									
	0									
	CollTime	-0.54	0.06	-8.39	-0.43	0.07	-6.11	-0.46	0.05	-8.36
Partially common variables	Autott	-0.06	0.01	-4.18	-0.05	0.01	-3.57	-0.05	0.01	-4.82
Characteristics of the traveler	0									
	0									
	0									
	0									

The parameter calibration results show that the t-test value of the CNL model is |t|>1.96, which means the results can be accepted in 95% confidence intervals. Meanwhile, the standard deviation of parameter calibration for the CNL model is improved compared with the NL model and the MNL model.

5 Conclusion

With the development of computer technology and mathematical algorithms, the application of discrete choice models has not only been limited to the general Logit model, the research on the improvement of the Logit model is developing continuously to adopt its good points and avoid its shortcomings. In this paper, a general idea has been explored to improve the Logit model, which introduces a layered concept for the CNL model. Using this method, any problem caused by the IIA (independence of irrelevant alternatives) property of the MNL mode can be solved. Moreover, by the way of inducting the alternatives cross ratio α_{km} and the nest dissimilarity parameter μ_m, the limitation of the NL model that the result for the same alternative changes when the packet structural is different has been broken through.

References

Ben-Akiva, M. E. and Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand.* MIT Press, Cambridge, Ma.

Ben-Akiva, M. E. (1973). *Structure of passenger travel demand models*. PhD thesis, Department of Civil Engineering, MIT, Cambridge, Ma.

Daly, A.(1987).Estimating tree logit models. *Transportation Research Part B*, 21(4):251-268.

McFadden,d.(1973).Conditional logit analysis of qualitative choice behavior. In: Xaremmbka, P. (Ed), Frontiers in Econometrics. Academic Press, New York.

Papola, A. (2004). Some developments on the cross-nested logit model. *Transportation Research B*, 38(9):833-851.

Small, K. (1987). A discrete choice model for ordered alternatives. *Econometrical*, 5(2):409-424.

Williams, H.C.W.L.(1977).On the formulation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning* A 9 (3):285-344.

Study on Effect of Taxi Quantity to Taxi Split Rate

CHANG,Chaofan1, MAO,Baohua2, CHEN,Tuansheng3

(1,2,3:School of Traffic and Transportation, Beijing Jiao Tong University, Beijing, 100044, China)

Abstract: Determination of Urban taxi quantity plays an important role in promoting the development of urban public transportation and economizing road resources. Rational taxi quantity can also increase the efficiency of taxi running. Firstly, this paper takes taxi unloaded ratio, sharing ratio, average operation speed, average daily operation time and price of taxi as parameters in order to put forward a model of taxi quantity based on a calculating model of unloaded ratio. Using Geijing as a case study, this paper works out taxi quantity of Beijing, analyzes in depth the relationship among unloaded ration, sharing ratio and quantity, discusses the effect of quantity on unloaded ratio and sharing ratio, and confirms the range of sharing ratio for Beijing taxis.

Keywords: Taxi； Quantity；Unloaded Rate；Split Rate；Price

1 INTRODUCTION

Now some big cities are actively developing public transit, and taxi, as an important part of urban public transit, plays a complementary role in conventional public transit, such as bus and subway. With rapid development of taxi in many big cities, taxi quantity goes beyond the value that it is supposed to be, which directly expresses a high unloaded rate of taxi and leads to the waste of road resources and energy. Facing this situation, many cities take measures to restrict providing taxi running cards in order to limit taxi quantity. Because there is not an upper limitation of taxi quantity, and "urban road programming and planning criterion" just provides lower limit, which shows that taxi quantity in big cities is more than 2 vehicles per 1000 persons, and more than 0.5 vehicles per 1000 persons in small cities, and middle cities can get value between these [1], it is difficult to control the development and speed of taxis. Therefore, many unloaded taxis occupy plenty of road resources, increase traffic load, and affect traffic running negatively. Also, taxi rational quantity plays an important role in urban taxi running efficiency, like unloaded rate and split rate.

This paper first educes a calculating model of taxi quantity, and then puts forward

a relative model among unloaded rate, split rate and quantity. Previously, main methods which study taxi quantity and split rate are: (1) By choosing factors affecting taxi quantity, and using nonlinear programming, get a dynamic model of taxi quantity, calculate these factors, and get approximate actual quantity [2]. (2)By calculating unloaded rate and using related data from household surveys, the calculating model of taxi quantity can be obtained [3]. This method is easy and practical, and it's adopted in this paper. (3)Considering social benefits and personal benefits, we choose traveler occupying road area, general traveling fee and congesting degree in public transit system, get taxi traveling proportion of public transit system, and confirm taxi split rate in all-mode traveling configuration [4].

This paper mainly uses calculating method of taxi unloaded rate, combines it with taxi split rate in resident traveling, considers taxi price and taxi traveling fee, and gets taxi quantity and analyzes thoroughly the relations among unloaded rate, split rate and quantity in order to provide a reference for controlling taxi quantity and unloaded rate.

2 CALCULATING METHOD OF TAXI QUANTITY

A household survey plays an important role in city planning and traffic planning. This paper takes advantage of taxi traveling data in a household survey, and these data include resident traveling rate, taxi split rate, taxi traveling fee, taxi travel speed, travel time and price.

2.1 Turnover of Taxi Taking on Resident Travel

Turnover of Taxi Taking on Resident Travel is that taxi loaded traveling distance multiplies quantity of taxi travelers, the formula being:

$$Q = P \cdot T \cdot R \cdot D \tag{1}$$

$$C = M_0 + (D - S) \cdot m, \quad D = S + \frac{C - M_0}{m}$$

where:

Q—Turnover of Taxi Taking on Resident Travel (10000 persons per km);

P—Total Quantity of Resident (10000 persons);

T—Resident Traveling rate (times per day);

R—taxi spilt rate (%);

D—taxi traveling distance (km);

C—Mean Taxi Fee (Yuan per time);

M_0—Taxi Beginning Fee (Yuan);

S—Taxi Beginning Distance (km);

m—Taxi running fee（Yuan per km）。

2.2Taxi Total Loaded Traveling Distance

The number of passengers loaded by taxi every time is uncertain, so it is necessary to use mean passengers loaded by taxi to confirm loaded traveling distance. The formula of taxi total loaded traveling distance is:

$$L = \frac{Q}{A} \tag{2}$$

where：

L—Taxi Total Loaded Traveling Distance (km)；

A—Mean Passengers Loaded by Taxi (persons per taxi)。

2.3 Calculating Taxi Quantity

This paper uses the method of calculating unloaded rate to confirm taxi quantity.

Taxi unloaded rate is the ratio of taxi loaded traveling distance and loaded summing unloaded traveling distance. The formula is:

$$Unloaded\ rate = \frac{unloaded\ traveling\ distance \times 100}{unloaded\ traveling\ distace + loaded\ traveling\ distance}\%$$

And,

$$C = 1 - \frac{L}{TVn} \tag{3}$$

where：

C—Taxi Unloaded Rate (%)；

T—Running Time the Whole Day (hour)；

V—Running Speed (km per hour)；

n—Taxi Quantity。

So the formula of calculating taxi quantity is:

$$n = \frac{L}{(1-C)TV} \tag{4}$$

Considering that travelers' leaving times are different at different times of the day, the quantity calculated by traveling quantity may exceed the day demand and may not meet the night demand. So it is necessary to confirm the period of time in which traveling quantity are the largest through city traveling character [5]. Therefore, the formula obtained by modifying formula (4) is:

$$n = \frac{xL}{(1-C)(yT)V} \tag{5}$$

where:

x, y—Traveling Quantity Taxi Taking on which portion is x in hours of (y*T). It can get taxi quantity through some data like traveling character, taxi price and turnover, which is related with taxi unloaded rate that needs to be confirmed. Before confirming taxi quantity, unloaded rate should be achieved at first, which is complicated to confirm and usually is related with city area, population, road area rate, traffic conditions and governable income of local residents.

3 CASE ANALYSIS

3.1 Introduction of Beijing city

Beijing city consists of 18 areas; its total area is 16410 km^2. In 2005, its total population is 15.38million, the number of motor vehicles is 2583000 [7], and the number of taxi is 66000, which accounts for 2.7% of motor vehicles, and increases 26.6% of that in 2004.

Tab.1 Beijing household survey data

Index	Beijing city
Population	15.38（million people）
Travel Rate	2.29（times per day）
Split Rate	1.65（%）
Travel Fee	20（yuan）
Running Speed	65（km/h）
Mean Loaded Passengers	1.26（persons per vehicle）

Data resource: Household Survey Report in Beijing in 2004.

3.2 Traffic Survey

In 2004, Beijing city administered a small size household survey, which investigated leaving time, place, objective and travel mode, etc. of resident in Beijing [8]. The result shows that travel rate is 2.29 times per day, taxi split rate is 1.65%, taxi travel fee is 20 Yuan, taxi running speed is 65 km/h, and mean load passengers are 1.26 persons per taxi. Details are shown in Tab.1.

3.3Taxi Quantity Analysis

3.3.1 Testing of Actual Taxi quantity

According to household survey, 70% of residents take 1.2 Yuan per km taxi and 30% of residents take 1.6 Yuan per km, so the mean price is 1.32 Yuan per km. Also,

travel quantity between 7:00 and 20:00 accounts for 92.3% of that of the whole day.

The turnover taken on by taxi gotten by formula (1) is:

$$Q = P \cdot T \cdot R \cdot D = P \cdot T \cdot R \cdot (S + \frac{C - M_0}{m})$$

$$= 849.04 \text{ （10000persons • km）。}$$

The total loaded distance obtained by formula (2) is:

$$L = \frac{Q}{A_1} = \frac{849.04}{1.08} = 786.15 \text{ （10000km）。}$$

The taxi quantity obtained by formula (5) is:

$$n = \frac{xL}{(1-C)(yT)V} = 6.86 \text{ （10000veh）。}$$

According to taxi unloaded rate of 51.25% in Beijing, taxi quantity is 68600, which is similar to the fact. So the number of taxi per 1000 persons is 3.6, which is more than average standard, 2 taxies per 1000 persons.

3.3.2 Calculating rational quantity

The rational range of taxi unloaded rate is between 30% and 40%, which means that supply is much more when exceeding this criterion, and means that supply is lower when under this criterion, and travel demand cannot be met. Accordingly, through formula (1), (2) and (5) rational quantity of taxi can be achieved between 47900 and 55800, quantity per 1000 persons is between 2.53 and 2.95.

3.3.3 Taxi Quantity Analysis

In order to confirm taxi quantity, relationship among unloaded rate, split rate and quantity are analyzed for details.

Relationship model among unloaded rate, split rate and quantity for details can be obtained by formula (1), (2), (5),which is:

$$n = \frac{xL}{(1-C)(yT)V} = \frac{x \cdot P \cdot T \cdot R \cdot D}{A \cdot (1-C) \cdot (yT) \cdot V} = \frac{181 \times R}{(1-C)} \tag{6}$$

where：

n—taxi quantity (10000veh)；

C—taxi unloaded rate；

R—taxi split rate (%)。

Through formula (6) the relationship curve between unloaded rate and quantity, and seeing figure 1, under unchanged split rate, taxi unloaded rate increases with the increasing of taxi quantity, and the extent of increase is larger; under unchanged

unloaded rate, taxi split rate increases with the increasing of taxi quantity, and extent of increasing is smaller.

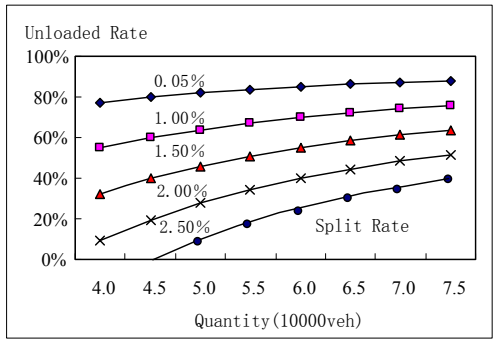


Fig.1 Relationship between unloaded ratio and quantity under various sharing ratio

Two formulas can be proposed by formula (6):

Relationship formula between unloaded rate and quantity is:

$$R = \frac{(1-C)n}{181} \tag{7}$$

Relationship formula between split rate and quantity is:

$$C = 1 - \frac{181R}{n} \tag{8}$$

Formula (7) shows: there is an inverse proportion between unloaded rate and split rate. In some periods, taxi split rate is certain, and under the circumstance of that other factors are fixed, if taxi quantity increased randomly, the unloaded rate must rise, and more social resource is wasted. So many cities take measures to restrict providing taxi running cards in order to limit taxi quantity, save social resource and decrease occupying of roads.

Formula (8) shows: there is a direct proportion between split rate and quantity. However, split rate doesn't always rise with the increase of quantity. Seeing figure 2, when quantity achieves a value of N_0, split rate becomes a certain value of C_0, and keeps invariable. Because when the city develops into some structure and resident economics achieves some degree, city traffic structure is relatively stable, and all transportation modes vary slightly, and resident choosing taxi for traveling does not change greatly because of the increasing of taxi quantity. Demand for taxi of resident also maintains a relatively stable value of C_0, which is not affected seriously by taxi

quantity.

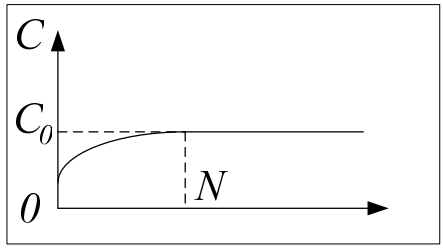


Fig.2 Relationship between sharing ratio and quantity

In fact, confirming formula of taxi split rate should be function consisted of multi-elements:

$$C_0 = f(k_1 \cdot k_2 \cdot k_3 \Lambda\ k_n) \tag{9}$$

In the formula (9), k_n should be related with city area, population, road area rate, resident governable income, GDP, and taxi price etc., and so taxi split rate should be a certain value of C_0 in some period.

Taxi split rate affects taxi running policy, and it determines taxi quantity and unloaded rate to some degree. According to figure 3, quantity obtained are between 47900 and 55800, and split rate is between 1.59% and 2.16%.

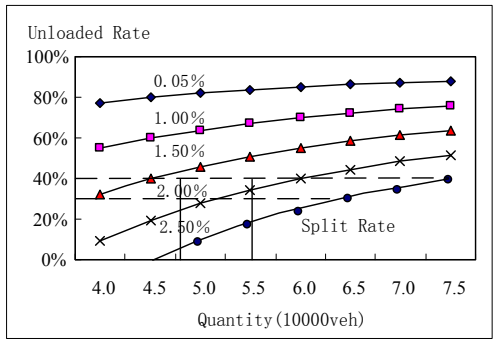


Fig.3 Range of rational unloaded rate

4 CONCLUSIONS

First, this paper studies the confirmation method of taxi quantity, which proposes

a calculating model of taxi quantity through calculating method of taxi unloaded rate. Moreover, according to travel rate, split rate, taxi travel fee, taxi price, running time etc. and combining with value of taxi rational unloaded rate, rational quantity can be calculated.

Also, this paper practically analyzes taxi quantity. According to some results of a household survey, the quantity of Beijing city is between 47900 and 55800, which are 10000 lower than the fact. It further analyzes relationship among unloaded rate, split rate and quantity. Although there is a direct ratio between split and quantity, when split rate achieves a certain value, it will not vary even if quantity varies, which is related to city structure, traffic development level and resident economic level [9].

Finally, this paper analyzes the relationship between split rate and quantity in order to get rational split rate which should be in 1.59% and 2.16%.

The method adopted in this paper is not perfect. It gets confirmation model of taxi quantity just by model of unloaded rate, not considering other complicated factors like city area, city economic level and resident governable income. And it is difficult to confirm rational unloaded rate of taxi, which is obtained in some works and materials, so this method needs to be improved further. Otherwise, taxi split rate will seriously affect the structure of city, and the value of C_0 also is related with some complex factors, which needs to be studied in the future.

REFERENCES

[1] Urban Road Programming and Planning Criterion [S]. GB50220-95. China Planning Press. Dec.1998.

[2] Wang Jin-shan, Ren Bei, Wang Len-fu Xu Chuang. Study on the Model of Taxi Quantity Determine of Hefei City[J].Operation Research and Management Science, Dec.2004,13(6).

[3] Wang Wei, Yang Xin-miao, Chen Xue-wu. Urban Public Transit System Programming Method and Management Technology[M]. Science Press. 2002.

[4] Bian Yang, Wang Wei, Lu Jian, Ma Jun-lai. A Forecasting Method for Urban Taxi Trafi c M ode Split Rate [J]. Journal of Transportation Systems Engineering and Information Technology. April. 2006, 6(2)

[5] Lu Jian, Wang Wei. Confirming method of urban taxi quantity[J]. Journal of Traffic and Transportation Engineering. Mar.2004,4(1).

[6] Beijing Statistical Information Net. Beijing Statistical Yearbook in 2006[Z].

[7] Beijing Transportation Research Center. Beijing Transport Development Annual Report in 2006[R]. Beijing: Beijing Transportation Research Center, 2006.

[8] Beijing Transportation Research Center. Household Survey Report in 2004[R].

Beijing: Beijing Transportation Research Center, 2005 年.
[9] Hyunmyung Kim etc. Effect of Taxi Information System on Efficiency and Quality of Taxi [J]. UCI-ITS-TS-WP, 2004, (11)

INTERMODAL COORDINATION: A VITAL ELEMENT FOR THE SUCCESS OF SHANGHAI MAGLEV SERVICES

Rongfang (Rachel) Liu[1]; Yi Deng[2]; Gao Li[3];

[1.]New Jersey Institute of Technology, Department of Civil and Environmental Engineering, University Heights, Newark New Jersey 07102, Tel: 973-596-5884, Fax: 973-596-5790; Email: rliu@njit.edu. Specially Invited Professor, Beijing Institute of Technology, Beijing China.
[2.]New Jersey Institute of Technology, Department of Civil and Environmental Engineering, University Heights, Newark New Jersey 07102.
[3.]Department of Mechanic and Vehicle Engineering, Beijing Institute of Technology, Beijing China.

ABSTRACT: After three year's operation, the review of Shanghai maglev project is mixed. One of the main critiques is that Shanghai maglev system is not fully utilized. Given the very affordable fare prices and highly congested conditions of other modes, many professionals are puzzled. The purpose of this paper is to call particular attention to intermodal coordination, which maybe one of the key factors that affect the success or failure of Shanghai maglev project. As proved by previous research, the Maglev system is designed to supplement, not replace, the existing transportation systems. The Shanghai Maglev system, as the initial segment of large transportation network systems, is bounded to connect and interact with the existing transportation network. If not supplemented by an integrated intermodal transportation network, especially efficient intermodal transfer stations, the maglev investment may become the biggest waster in transportation history.

KEY WORDS: Intermodal Transfer Penalty, Travel Behavior, Maglev, Operation Planning, Ridership Forecast.

1. INTRODUCTION

Maglev is an advanced transport technology in which magnetic forces lift, propel, and guide a vehicle over a specially designed guideway (FRA, 1997). As the first commercial application in the world, Shanghai Maglev utilizes state-of-the-art electric power and control systems and eliminates the need for wheels and many other mechanical parts, thereby minimizing resistance and permitting excellent acceleration, with cruising speeds on the order of 300 mph or more. This high performance enables a rapid connection between downtown Shanghai and Pudong International Airport.

After three years in revenue services, the review of Shanghai maglev project is mixed. One of the main critiques is that Shanghai maglev system is not fully utilized. Given the very affordable fare prices and highly congested conditions of other modes, many professionals are puzzled.

Along with the investment in guideway related transit systems, intermodal transfer stations are likely to be pursued by local communities. To serve transit users, intermodal transfer stations need to be designed with features and characteristics that are sought by travelers and complementary to known travel behavior. To improve the quality of our transit services and facilities we also need a better understanding of how transfers will affect transit users as well as the performance of the transit network. With limited research on penalty associated with intermodal transfer and lack of consensus on transfer penalty value, its impact on ridership forecast is often ignored or understated. It is especially so when Maglev systems are involved, since the ultra-high speed of Maglev system results in very short travel time.

This effort began with a literature search in order to understand the current state of knowledge and to focus study needs regarding intermodal transfers. The search targeted previous studies concerning intermodal transfer penalties, value of time, and mode choice behavior. After documenting transfer values derived from previous studies, the author focused on the behavioral aspects of transferring within the traditional mode choice context. To paint a richer picture of intermodal transfers in relation to travel demand forecasting, this paper proposes a transfer paradigm, which categorizes those elements associated with transit use and intermodal transfer processes in an integrated framework.

By applying the known function of intermodal transfers, the author demonstrated how the Maglev network and operation schedule could be planned to reduce intermodal transfer penalty to create user-friendly transit networks. To obtain realistic Maglev ridership forecast, it is the author's strong belief that the limited but known value of transfer penalties should be incorporated into the travel demand forecasting model. This paper uses the shanghai Maglev data; however, the underlying principles should be applicable to most other intermodal systems, as well as Maglev.

2. MAGLEV TECHECHNOLOGY

Utilizing state-of-the-art electric power and control systems, this configuration eliminates the need for wheels and many other mechanical parts, thereby minimizing resistance and permitting excellent acceleration, with cruising speeds on the order of 300 mph or more. The new technology is better or competitive in many performance characteristics that include speed, acceleration, travel time, travel cost, comfort, convenience, health, capacity, flexibility, frequency, operational and schedule reliability (weather and equipment delays), accessibility, safety and security, system availability (origin and destination), and energy consumption.

The advantages of high speed Maglev technology are experienced and recognized by more and more people from Shanghai demonstration line. As documented in an earlier study (Liu and Deng, 2004), there are a number of operating advantages which are tested and proved in the Shanghai demonstration project.

2.1 Speed, Acceleration, and Deceleration

As documented in Table 1, the design speed of Maglev is the highest among all existing surface transportation modes. It is about 67% higher than that of the high speed rail, such as ICE train. The operating speed is the major indicator of the traveling speed and is usually lower than the design speed. Similarly The difference is

in the same proportion for Maglev and HSR, namely, the operating speed of the Maglev is about 61% percent higher than that of the ICE train.

If the design and operating speed of each mode play a key role in the travel time comparison, acceleration and deceleration rate may be an even more important factor in terms of safety spacing and average travel speed over certain distances. Because of application of contact less guidance and drive, the accelerating rate, climbing capability, and the curve radius of Maglev system are determined by electrical power, not by friction between wheel and rail like the traditional train. The accelerating rate of Maglev system may reach up to 1-3 m/s2, for example, vehicles in Shanghai line can be accelerated from 0 to 431 km/h within 3 minutes only. The approved climbing capability of Maglev systems may attain up to 10 %, while it is only 4 % for wheel rail systems. The curve radius of Maglev system is smaller than high-speed wheel-rail system at the same speed. As a result, it can shorten the route, use less soil, and reduce the total investment.

Maglev technology also allows flexible route alignments. Maglev can climb 10% grade comparing to a maximum 4% for HSR. Maglev vehicles can negotiate tighter curves (horizontal and vertical) at the same speeds as conventional high-speed rail. Similarly, they can travel through a curve of the same radius at much higher speeds than conventional systems. For example, Maglev is capable of maneuvering curves of 2,350 meter radius at 300 km/h and cants up to 16° (Transrapid International Inc., 2002).

TABLE 1. Specifications of Maglev and HSR Trains

Speed	**ICE**	**MAGLEV**	**Commuter Rail**	**Light Rail**
Experimental Speed (km/h)	407	550		
Design Speed (km/h)	330	550		65
Operation speed(km/h)	280	450	80	55
Average speed (km/h)	200	290	60	30

2.2 Capacity

The capacity of a rail line is decided by the following factors:

- Passenger seats per section,
- Number of sections, and
- Headway.

The upper limits of train length and number of seats or passengers per section are usually decided by the limits imposed by propulsion power. Besides technical specifications, the length of the vehicle selected is closely related to population density and ridership demand. Maglev's capacity is not constant. On the test line in Emsland, Germany, Transrapid vehicles comprise a minimum of two sections, each with approximately 90 seats per section. Depending on the travel demand, Maglev trains may be composed of up to ten sections, two end and eight middle sections.

The headway for Maglev trains is largely decided by the length of sections since its propulsion power is transformed via activating guideway sections. A Maglev guideway section may only contain a single Maglev train at a particular time, which guarantees the safe spacing of Maglev operations. The typical Maglev guideway sections length is from 200 to 2,000 meters. In the case of Shanghai Maglev Project, the average guideway length is 1,250 meters. According to Transrapid International Inc., (2002), the theoretical minimum headway is 5 minutes. However, there has not been a single Maglev system designed to operate in such closely spaced headways. For example, the Berlin-Hamburg test line was designed to run with 20-minute headway. The Baltimore-Washington Maglev project is planning 10-minute headway (Liu, 2001) and 15 minutes for the Shanghai Maglev, in operation since January 2003.

2.3 Safety and Reliability

When discussing transportation strategy, safety and security are both linked. Several studies show that safety and security are two of the most important factors in project evaluation and are of greatest individual importance (Liu, 2003). The factors affecting transportation safety and security are various, among which, the physical structure and guideway security patrols play significant roles.

Maglev's concept has essentially eliminated the safety risks associated with the operation of conventional rail transportation systems. For example, the Transrapid vehicle wraps around the guideway and therefore is virtually impossible to derail. Redundancies achieved through the duplication of components as well as the automated radio-controlled system ensure that operational safety will not be jeopardized. The principle of synchronized propulsion on the guideway makes collisions between vehicles virtually impossible. If two or more vehicles were ever placed simultaneously in the same guideway segment, they would be forced by the motor in the guideway to travel at the same speed in the same direction. The grade separated, flexible route alignment (elevated as well as at-grade guideway) ensures that no other obstacles can be in the way. Energizing only the section of the guideway on which the train is traveling enhances operational safety and efficiency.

Owing to the mechanic and electrical structure of high-speed Maglev system, the dangerous accident of derail, happened in traditional train, may not occur. Eliminating mechanical contacts made it possible to realize calm operation, small vibration, and comfortable ride. The usable rate of 99.83 % and operating punctual rate of 99.57 % in Shanghai line, including normally commercial operation under the 30 m/s typhoon condition, are the best evidence of safety and reliability of high speed Maglev systems.

2.4 Energy Consumption

In terms of energy consumption, Maglev trains are slightly better than conventional trains in general. The energy consumption of the Transrapid system with its non-contact levitation and propulsion technology, highly efficient linear motor (mounted in the guideway) and low aerodynamic resistance is very economical when compared to other transportation modes.

Similar to new generations of automobile engines, Maglev consumes less energy while providing the same output as high-speed railroads. At the same output the super speed Maglev system consumes 20 to 30 percent less energy than the already very 'modest' railroad. In other words, with the same energy input, the performance of Maglev is substantially higher than HSR. The favorable aerodynamic properties and the non-contact technology make the Maglev more economical in energy consumption.

As consumers of energy, the transportation sectors are vulnerable to environmental and global warming concerns and the increasing volatile oil market. Reducing dependency on foreign oil is also an important criterion. The Maglev consumes less energy per seat-mile than conventional commuter trains due to the utilization of lightweight materials and improvement in the advanced technology.

2.5 Maintenance

Owing to uniform force distribution and no abrasion, small maintenance in Shanghai line is to examine and change the electronic module with something wrong, but vast work on replacing and repairing mechanical parts in traditional rail are not required. Along with the development of electronic industry, the reliability of electronic devices will be improved gradually, and the maintenance should become simpler and easier. Little maintenance is a key issue to reduce the cost of operation and investment.

2.6 Environmental Impact

Electrical driving in Maglev systems eliminates emitting of baneful gas and environment pollution. Elevated guideway use less soil land. Magnetic fields in the passenger cabin of Maglev trains are equal to Earth's, smaller than color TV's.

The frictionless operation of Maglev reduces vibration, noise pollution, and maintenance resulting from wear. Comparing the noise levels at different speeds, we have observed that Maglev technology is much quieter than HSR trains in speed of 200-300km/h. The fundamental reason is that Maglev operation does not produce any rolling, gearing, or engine noises. Because those noise sources predominate at low speeds, their absence provides Maglev much advantage in urban areas. Beyond the speed of 250 km/h, aerodynamic resistance becomes the main source of noise with the increase of speed. As speed attains to 400 km/h; Maglev is still quieter than HSR at 300km/h though their difference becomes small. (Transrapid International Inc., 2002).

3. THE ROLE OF TRANSFERS

As an important factor affecting modal choice behavior, transfer penalty has not received its fair share of attention. There is little data available and scarce research about intermodal transfers; even a majority of the transportation profession has realized that this is a very important issue in the context of travel behavior modeling (Stopher, 1967).

Most transit operators and planners agree that the need to transfer and the perception of its inconvenience discourage people from using transit modes. There is a consensus that time spent transferring is more onerous than in-vehicle travel time.

Two studies conducted in different areas and by difference researchers (Liu, et.

al, 1997; Quackenbush, et. al., 1995) arrived at the same conclusion, that is, transfer waiting time is over two times more onerous than in-vehicle time. The marginal rate of substitution of transfer waiting time relative to in-vehicle time is in the vicinity of 2.0 to 3.0. The theoretical rationale for a traveler's higher valuation of waiting time is that such time is difficult to employ productively, is often spent in an unprotected environment, and is typically of uncertain duration, even where the schedule of service is known in advance (Pickrell, 1989).

On the other hand, it is argued that transferring is not all bad. Sometimes, a transfer provides more options and complements a travel process. A personal observation revealed that, in New York City, riders on a local bus often voluntarily transfer to an express bus to minimize their total travel time. Richard Pratt, R. H. Pratt Consultants, Inc., observed that, in Washington D. C., some commuters rode in a reverse direction in the morning, and then transferred to the right train just to get a seat. These phenomena suggest that not all transfers are forced. When chosen voluntarily, transfers may provide added satisfaction and can be utilized to improve travel efficiency.

Using the 1995 Nationwide Personal Transportation Survey, Polzin et al. (1998) reported some interesting statistics related to transfers at the national level. First, the average length of waiting time, including both first and transfer, is 9.6 minutes, which is about 20% of the average total transit travel time including both waiting and in-vehicle time. Second, about 20 percent of all unlinked transit trips are transfer trips. Third, about half of the transit users perceive transferring as a problem or disutility in their travel choices.

In summary, the disutility associated with intermodal transfers and incorporating such measures into travel behavior models is very important and deserves much attention, especially for the transportation project planning.

4. SHANGHAI MAGLEV: A CASE STUDY

To establish a user-friendly transit network, one of the key issues is to ensure the connectivity and easy transferability between different modes. Shanghai has an extensive public transportation system, largely based on buses, and a rapidly expanding metro system. Shanghai has very extensive bus system with nearly one thousand bus lines. The Shanghai Metro (subway and elevated light rail) has five lines (numbers 1, 2, 3, 4, 5) now and are going to expand to 13 lines by 2010.Taxis in Shanghai are plentiful and market competition has driven taxi fare down to affordable prices for the average resident (11 RMB or a little over one US dollar for 3 km).

Bicycling is still acted as the most important commuting vehicle, accounting 25.2% of total trips in 2004. Private car ownership in Shanghai has also been rapidly increasing in recent years. Two railways intersect in Shanghai: Jinghu Railway (Beijing-Shanghai) Railway passing through Nanjing, and Shanghai-Hangzhou Railway. Shanghai has two main railway stations, Shanghai Railway Station and Shanghai South Railway Station. A high-speed railroad to Beijing is also in the works.

4.1 Mode Share along the Maglev Corridor

In cooperation with the Shanghai municipality and the Shanghai Maglev Transportation Development Co. (SMT), German Transrapid constructed the first

commercial Maglev railway in the world in 2002, from Shanghai's Longyang Road subway station to Pudong International Airport. It was completely built on December 31, 2002, and began its commercial operation in May 2004. It has a peak speed of 431 km/h and a track length of 30 km. This line is 30km long, which takes only eight minutes, in which the German TR technology with the maximum operating speed of 431 km/h was used. Maglev connects PVG and Longyang Road Metro Station. PVG is a major international gateway into China, handling 17.15 million passengers on international flights in 2006. It was put into use formally and operates everyday from 7:00 to 21:00The operation interval is 15 minutes. The round trip ticket cost 80 RMB and the one way ticket cost 50 RMB. Passengers with current day flight ticket can buy a one way ticket of Maglev Train at 20% discount.

As shown in Figure 1, there are a number of transportation modes along the Maglev corridor, namely from Pudong International Airport (PVG) Downtown shanghai, Lujiazui. The Red line shows the route for airport shuttle, purple lines is taxi and private route. Blue line is the metro and Maglev. Quite a few choices for airport passengers are listed compared below:

1. Airport Shuttle Bus

Passenger can board the airport shuttle at Lujiazui airport shuttle bus station to take line 5 or line 6 to PVG airport. The headway is 20-30 minutes. Fare is about 15 RMB. Travel time is about 40 minutes, 60 minutes in peak hour.

2. Metro-Maglev

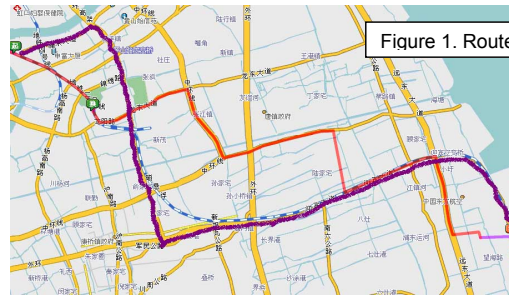


Figure 1. Route Map

There is also option for passenger to take metro and Maglev to PVG airport. Start from Lujiazui metro station, passenger can take subway line 2 from Lujiazui to Rongyang Road, headway 5-7 minutes, travel time 13 minutes. Then transfer to Maglev line to airport, headway 15 minutes, travel time 8 minutes, and cost 40-50 RMB.

3. Taxi

The taxi price starts from 11 RMB for the first 3 km, then 2 RMB per km after. It takes about 40 minutes and 80-90 RMB to the airport.

4. Private car

The travel time is same with taxi. Besides fuel and operation cost, the parking fee is 10 RMB/Hour.

Table 3: Time and Cost Comparison

Mode	Airport Shuttle	Subway-	Taxi	Private car

	Bus	Maglev		
Time (Minutes)	40-60	21(in vehicle)	40-50	40-50
Cost(RMB)	15	44-54	80-90	Fuel+10/hour parking

Applying the classic mode share model dominated by travel time and cost:

$$U = \alpha + \beta 1 \bullet T + \beta 2 \bullet C$$

Where:

$\alpha, \beta 1, \beta 2$ are parameters

T : Total travel time

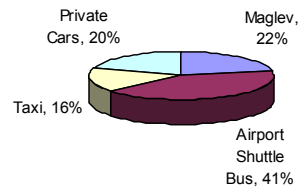
C : Out of pocket cost

Using the multinomial logit model(MNL)

$$P(m_1) = \frac{e^{u_1}}{\sum_{i=1}^{m} e^{u_i}}$$

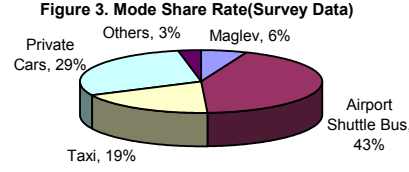
We can easily derive and forecast that ridership for each mode based on the travel time and costs are as the followings. It is understandable how brought an outlook for maglev it is when such rideship levels were projected during the project planning and design stages. By using the data from table 3, the mode share rate can be derived as Figure 2.

Figure 2. Mode Share Rate (From MNL Model)



The reality is that maglev share only takes up merely 5-8 percent depending on the directions, as documented by a survey (Wen, et, al, 2006) as Figure 3. Comparing to a similar survey conducted in 2000, prior to the construction of Maglev, airport shuttle bus, taxi, private car and greyhound. They accounted for 35%, 30%, 20% and 15% respectively. The mode share is fairly consistent, except the increase in private cars and decrease by shuttle buses. The trade off between shuttle bus and private cars may serve as a good indicator of auto ownership by private citizens. However, the overall impact by maglev/metro rail is quite small. It is certainly puzzling given the higher utility rate projected by the classic mode choice formula.

Figure 3. Mode Share Rate(Survey Data)



The Reason for Low Maglev Share

As demonstrated in the transfer paradigm, another important factor, comfort and convenience, comes to play when cost and time is similar or relatively small portion of the overall income.

As we have cautioned the planners before the construction of the maglev, besides the time and cost, transfer is also very important factor in the process of passenger's mode choose. Metro-Maglev is the only mode that needs transfer in the trip. Passengers dislike the time and cost required for transferring, but they also dislike the need for added trip planning, the possibility of a missed connection, the uncertainty of arrival time at their destination, exposure to weather and crowding, the need to find the next vehicle, difficulty of baggage handling and waiting in unfamiliar or hostile surroundings. The passengers to and from airport usually with more baggage then normal passenger, the transfer penalty is even higher for them. Even for travel time, although metro-Maglev in vehicle is low, considering the waiting and walking time with transfer, the total travel time saving is not very attractive for most of passengers. So in Shanghai, passenger with lower Value of Time (VOT), they choose airport shuttle bus for the cheaper price. For the passenger with high VOT, they choose taxi or private car for its convenience.

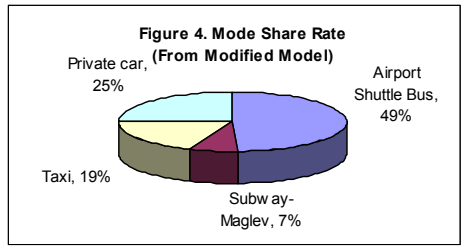
Incorporating transfer penalty incurred by the mode choices along the shanghai maglev corridor, consider transfer penalty, such as walking and waiting time, in the mode split process. We have derived a totally different mode choice scenario as demonstrate in Figure 4.

$$U = \alpha + \beta 1 \bullet (IVTT + OOVTT + TT + TP) + \beta 2 \bullet C$$

Where: TT **:** Transfer Time

TP **:** Transfer Penalty

A close looking at the projected mode share, which is amazingly similar to the survey results, shows that the impact of transfer on the travel demand is very important, especially for urban transportation. So the transfer penalty should be considered in the planning phase to make the projection more accurate.



Possible Solutions

A good intermodal transfer facility can decrease the unpleasantness of the transfer by directly addressing the reasons why travelers avoid transfers. The main reason for shanghai case is the luggage carrying and transfer time.

So possible method to attract the passenger to take Maglev can be

1. Coordination of the Operation of Metro and Maglev

Although metro and Maglev are belong to different authorities at present, their service can be coordinated to lower the transfer waiting time.

2. Luggage Handling

The airport can work with Maglev company to provide extra service for the luggage carrying. If passengers can claim their luggage when they take off the Maglev (the connection area of Maglev and metro), the possibility for them to choose Maglev will be higher.

3. Maglev Headways

The reason is transfer penalty is caused by the waiting time of transfer. So if the operation company can reduce the headway, it can attract more passengers. Besides, it is suggested to add some real-time information board to let the passenger know when the next Maglev will train come and they can make the arrangement and feel better with the waiting.

CONCLUSIONS

Several observations can be made regarding transferring behavior and our ability to understand travelers' desires, forecast travel demand, and design services so as to increase the attractiveness of the public transit travel options.

- Clearly, transferring behavior is a critical component of travel behavior that merits careful consideration in modeling efforts. The paradigm outlined in this paper suggests that a host of conditions, from group size to trip purpose to whether or not one has made that particular trip and experienced the transfer previously, will influence behavior regarding transferring.
- Increased transferring is inevitable if we are to have public transit alternatives play a larger role in our transportation system. Thus, it is critical that we fully understand how transferring is perceived by travelers as we are making significant, not easily reversed, investments in stations and facilities, such as Maglev, which will shape the characteristics of many transfers.
- Additional research into transferring behavior is important to fully understand transferring and correctly characterize it in our travel models.
- Transfer facility design is increasingly a multi-objective process where economic development, joint development revenue opportunities, and other considerations may influence design as much as functionality for travelers. Caution is urged in transfer facility design so as not to compromise the fundamental travel choice objectives of safety, speed, comfort, and convenience as we develop an ever-increasing number of transfer facilities in our urban areas.
- New technologies offering improvements in fare integration, security, passenger information and comfort may enable us to improve the transfer experience relative to historical opportunities.
- Investments in transfer facilities and amenities in transfer environments should be evaluated carefully. All resource investments in transit should be evaluated from the perspective of how the passenger would respond to a given investment in comparison to other investments in improved facilities or services. For example, is a weather-protected shelter at a transfer location of more utility than newer and more reliable or comfortable buses?
- Fundamentally, transferring is an impediment to transit use. Our interest in multimodal and intermodal systems should be driven by desires to serve passenger

needs. Where possible, transferring should be designed to incorporate our most current knowledge of transfer behavior; however, current evidence indicates that designing transit services that do not require a transfer are most likely to be most attractive to travelers.

- Additional research into transfer behavior will not only improve our ability to predict travel behavior, but also promote facilities and service networks that are better designed to meet travelers' needs.

Characterized by self-guided operation, higher speed, lower noise emission, and less land consumption, Maglev technology has wide appeal to various areas with diversified environments. However, the Maglev system is designed to supplement, not replace, the existing transportation systems. The Maglev system, either initial segments or built-out corridors, are bounded to connect and interact with the existing transportation network.

When intermodal transfer penalties are ignored or underestimated, severe modal biases may be created toward the Maglev mode with resultant over-estimated ridership. The goal of this paper is to call particular attention to the impact of transfer penalties on the traveler's behavior, so the on-going Maglev studies will be able to address these issues in the operation planning and ridership forecasting stages.

REFERENCES:

Alger, S.; S. Hansen; and G. Tegner (1974). Role of waiting time, comfort and Convenience in modal choice for work trip. *Transportation Research Record* 534.

Caliper Corporation (1998). Travel Demand Modeling with TransCAD 3.1.

Domencich, T. and D. McFadden. (1975). *Urban travel demand: a behavior analysis*. North-Holland Publishing Company.

Han, A. (1989). Assessment of transfer penalty to bus riders in Taipei: a disaggregate demand modeling approach. *Transportation Research Record* 1139.

Hensher, D. (1972). *The Consumer's choice function: a study of traveler behavior and values*. School of Economics, University of New South Wales, Ph.D. Dissertation.

Horowitz, A. (1982). The subjective value of the time spent in travel. *Transportation Research* 12: 385-393.

KCI/PB (2000). *Baltimore-Washington Maglev Project: Preliminary Operation Plan.*

King, G. (1983). *Mathematical techniques for estimation of the value of travel time savings*. Bureau of Transport Economics Australian Government Publishing Service, Canberra.

Lave, C. (1968). *Modal choice in urban transportation: a behavior approach.* Ph.D. Dissertation, Department of Economics, Stanford University.

Liu, R. (1996). *An assessment of intemodal transfer disutility*. Ph.D. Dissertation, Department of Civil Engineering, University of South Florida.

Liu, R., R. Pendyala, and S. Polzin (1997). Assessment of intermodal transfer penalties using stated preference data. *Transportation Research Record* 1607.

Liu, Rongfang (Rachel), 2001. A Transfer Paradigm: the Impact of Transfer Penalty on Maglev Operation Planning, 80th Annual Transportation Research Board Meeting, Washington D.C., January 2001.

Liu, Rongfang (Rachel) and Guilin Li, 2003. "Intercity Travel Choices Affected by Safety and Security Measures." Transportation Research Board 82nd Annual Meeting, Committee A1C06. Social and Economic Factors, Washington DC. January 15, 2003.
Pickrell, D. (1989). *The treatment of transfers in forecasting ridership on Houston's proposed system connector. Letter to chairman Lanier*, Working Memo, June 26.

Pickrell, D. (1992). A desire named streetcar: fantasy and fact in rail transit planning. *Journal of American Planning Association* 58(2).

Polzin, Steve E., Joel Rey, and Xuehao Chu (1998). *Public transit in America: findings from the 1995 Nationwide Personal Transportation Survey*. Draft Report, National Urban Transit Institute, Center for Urban Transportation Research, University of South Florida, Tampa, Florida.

Quackenbush, K., J. McClennen, J. Gallagher (1995). Transfer penalties in mode choice decisions. *Fifth National Conference on Transportation Planning Methods Applications*, Seattle, Washington.

Resource Systems Group, Inc. (1993). *NJ Transit Stated Preference Survey and Mode Choice Refinement*. Final Report.

Stopher, P. (1967). *Factors affecting choice of mode of transport*. Ph.D. Thesis, University College, London.

Stopher P.; B. Spear; and P. Sucher. (1974). Toward the development of measures of convenience for travel models. *Transportation Research Record* 527.

Transportation Economics and Management Systems, Inc. (2000). *Baltimore-Washington Maglev Preliminary Ridership Forecast.*

Transrapid International Inc., 2002. Comparison of System Characteristics, February 2001

United State Congress (1998). *Transportation Equity Act for the 21st Century*. Congress Record, May 22, 1998.

U.S. Department of Transportation, Federal Transit Administration (1997). *High-Speed Ground Transportation for America.*

Urban Analysis Group (1995). *User Menu for TRANPLAN.*

Watson, P. (1974). *The value of time, behavior models of modal choice.* Lexington Books.

Wen, Ya; Yan, Kefei; Qiao, Xiang-rong; Shi, Jia-wei. The Characteristic Analysis of Passengers' Selection of Ground Transport Mode Connecting Shanghai Pudong International Airport and the Downtown Area. *Transportation Research Board Annual Conference CO-ROM. 2006.*

Research on Basic Concept and Key Technologies of Experimental Traffic System

WU Zhizhou[1], TONG Mei[2]

[1]School of Transportation Engineering, Tongji University, Siping RD 1239, Shanghai, China, 200092; PH: 86(21)65981589; FAX:86(21)65981589; email: wuzhizhou@mail.tongji.edu.cn

[2] School of Transportation Engineering, Tongji University, Siping RD 1239, Shanghai, China, 200092; PH 86(21)65988372; FAX:86(21)65988372; email: tongmei@21cn.com

Abstract: Experimental Traffic Engineering (ETE) is a new research trend of modern traffic engineering for the purpose of the study and development of the complex ITS-related traffic systems. Those traffic systems constructed with the concept of ETE can be called Experimental Traffic Systems (ETS). This paper illustrates the main components of ETS and relevant key technologies. Finally, a case of ETS, SHGTISP (Shanghai-Grid-Based Traffic Information Service Platform) is introduced.

Key words: *Experimental Traffic Systems，Experimental Science，Data Warehouse，Computational Intelligence，Parallel Computing*

1 Introduction

ETS is a general designation of categories of traffic systems constructed on the base of ETE with ITS background. Based on vast amount of real-time data collected from ITS infrastructures, the tasks of ETS are exploring dynamic evolvement of traffic systems by all kinds of technologies for traffic management, traffic control, traffic information services, and computer-aided decision-making support. In general, there are five components in an experimental traffic system, i.e., real-time data collection, online traffic simulation, dynamic parameter calibration, information fusion of historical/real-time data, online analysis and mining of traffic data.

2 Concept of ETS

2.1 Experimental Science and ETS

In the seventeenth century, Francis Bacon introduced the concept of experimentation into scientific research: observing, hypothesizing and verifying the former hypothesis by way of experiments. They form the general chain of an experiment. The character of experimental science is that one or several conditions are assumed to the experimental object for the purpose of experiment.

Claude Bernard had expounded the difference between experiment and observation in “An introduction to the study of experimental medicine”. In his opinion, observation is a passive activity while experimentation is active in the process of exploring the world. Observation made in view of a preconceived idea can be called a passive experiment. The main difference between the two methods is that what

observation gains are nature phenomena without interference while what experiments do will be varied or interfered with by man on purpose. Medicine includes experiments that occur spontaneously and those designed by physiologists.

Homologically, there are passive experiments occurring spontaneously and active experiments designed by traffic system analysts. Hence there are two connotations with ETS: one for explanation of the real world, a passive process to learn based on observed data, the other for experiments or simulation of the real world, a more active method with man-made interference. Considering the complexity, economical and applicable factors in experiments of traffic systems, the practical traffic experiments are always performed using computational experiment, i.e., computer simulation.

Furthermore, as a method to study complex systems, experiments are introduced into many research fields such as psychology, biology, economics, the science of management and so on. The research places emphasis on simulation of researched objects, data collecting, trends observing and summarizing of nature laws. ETS concerns these contents as well.

2.2 Elements of ETS and Their Relations

ETS consists of six major elements: real-time data collection, computational experimental systems including online simulation and online analysis, dynamic parameters verification, data fusion, decision supporting and information dissemination, traffic user (including decision-maker, traffic manager and travelers), and time. Time is the most important element of ETS through which the features of ETS such as dynamic, real-time, online, feedback can be presented.

The relationship of elements is shown as Figure 1.

ITS in Operating module is a compound traffic system composed of one or several ITS sub-systems, take adaptive signal control system, vehicle information collecting system, traveler information system for example, from which real-time traffic data is generated and affected by the computational results of ETS.

Real-time data such as traffic flow and speed, a running trajectory of a floating car, traffic accidents, electronic toll information are collected from relative ITS sub-systems, to serve the traffic users through a common information platform or particular individual system.

The Computational Experimental System module is the most important part of ETS with two major functions: simulation of real world and data mining from huge real world data, which are implemented by online simulation sub-system (OLSS) and online analytical data mining sub-system (OLADM) respectively. OLSS is the computational tool for traffic experiments and OLADM is the analytical tool for demonstrating varied traffic phenomena from the real world or simulated. The two sub-systems support each other and serve the traffic users for decision-making, management, and information service.

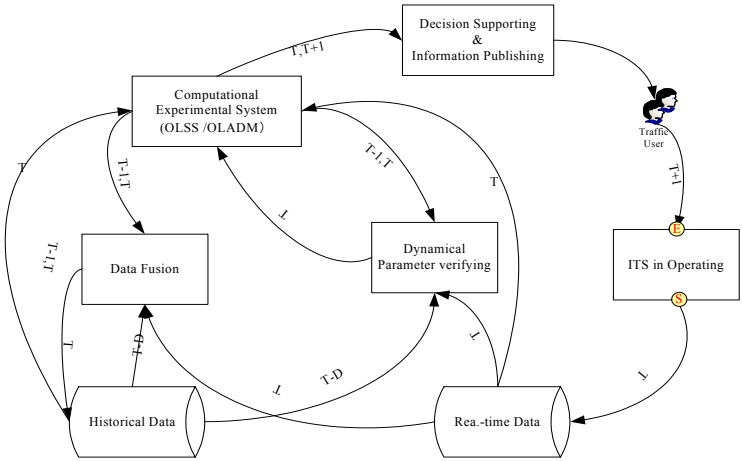


Figure 1. Elements of ETS and their relations

Historical data is a characteristic of ETS, upon which most models of ETS are based. Hence, whether the whole system is precise, applicable and reliable depends mostly on precision of historical data. Before being updated with particular rules for the database, unreasonable/missing data must be corrected/recovered through data fusion based on real-time data, existing historical data and results generated from computational experimental systems.

The Dynamic Parameter Verifying module produces dynamic parameters calibrated for models in computational experimental system. Calibrated from a previous cycle referring to real-time data and historical data, the initial value of the parameters in the current cycle is delivered to computational experimental system.

Time is the glue for all the sub-systems. Referring to Figure 1, starting with ITS in operating, verified dynamically by real-time data generated in time T, historical data in time T-D and data generated by computational experimental system in time T and T-1, the parameters in time T make a reaction to the computational experimental system, forming a computational feedback circuit. Similarly, as the result of data fusion with real-time data in time T, historical data in time T-D and data generated from computational experimental system in time T and T-1, the historical data in time T complete a time feedback cycle. The computational result of computational experimental system in time T+1 can become a computational source of an online analytical system, a decision support system and a traveler information service system, which will affect the operation status of ITS in operating via affecting the behaviors of traffic users. Here to there is a time feedback cycle starting with ITS in operating and ending with ITS in operating.

3 Key Technologies and Methodologies of ETS

3.1 Data Warehouse

Historical data are stored using data warehouse technology in ETS. A data warehouse is a subject-oriented, integrated, time-variant, nonvolatile collection of data in support of management decisions. Unlike that of traditional databases, the establishment and use of data warehouses are all subject-oriented. Subject is a high level standard of data categories. Each subject is corresponding to a specific field. For example, volume and speed can serve as two subjects using data warehouse in the process of analyzing traffic flow characteristics.

One of the key steps in building a data warehouse is the transferring, processing and integrating of data, that is, a fusion of historical data. Raw data must be unified at first and then the data structures should be changed from application-oriented to subject-oriented. For instance, the original records collected from traffic information collecting systems will be stored and managed as fact tables and dimension tables by being cleaned and fusing.

There are several major applications of data warehouse, including Online Analytical Processing (OLAP), Data Ming (DM) and Knowledge Discovery (KDD). OLADM is a crucial sub-system of ETS. Through data presented from varied granularities and dimensions such as time and space, it is more efficient for real-time analyzing and processing. Using OLAP technology, it is easy to perform slice, dice, drill-down, roll-up, pivot on the traffic data. What's more, after matching the analysis results to real-time data, via contrasting, applying rules or reasoning, the consequent meaningful traffic information can well support decision-making, management and information service in the next cycle. Further more, the feature of parallel and real-time processing calls for high computational performance to ETS.

3.2 Computational Intelligence and Data Mining

The definition of Computational Intelligence (CI) was first put forward by James Bezdek. He classified intelligence into three levels, Biological Intelligence (BI), Artificial Intelligence (AI) and Computational Intelligence. CI is at the lowest level and a sub-set of AI. It is a set of algorithms implemented by computer. Up to now the research fields of CI has consisted of neural networks, genetic algorithm, evolutionary computation, immune information processing, agent, artificial life and ecological computation.

Data mining (DM), also called Knowledge-Discovery in Databases (KDD) or Knowledge-Discovery and Data Mining, is the process of automatically searching large volumes of data for patterns using tools such as classification, association rules, clustering and so on. It is a fairly recent topic in computer science but applies to many older computational techniques from statistics, information retrieval, machine learning and pattern recognition, etc.

In ETS, CI can apply to not only traffic data management in the OLSS module, modeling and computation in process of management, but also to analyzing and decision support through data mining with well-organized massive data in OLADM module.

3.3 Parallel Computing and Distributed Computing

Parallel computing is the simultaneous execution of the same task (split up and specially adapted) on multiple processors in order to obtain results faster. The idea is based on the fact that the process of solving a problem usually can be divided into many smaller tasks, which may be carried out simultaneously with some coordination. There are piles of engineering computing and numerical simulation in ETS, and massive data must be processed at the same time, where parallel computing is competent for such an arduous computational task. The key approach of parallel algorithm designing includes task partitioning, communication analysis, task combination and task assignment.

Distributed computing provides a cheaper computing environment for parallel computing than the expensive parallel computers do. Distributed computing is a science which solves a large problem by dividing and giving parts of the problem to many other computers to process and then combining all part solutions to the one for the problem. A Grid is a popular distributed computing architecture, and its kernel conception is resource sharing and collaborative computing, which has the same view with goal of ETS. Grid computing is tightly correlative with service-oriented architecture (SOA). It is a trend to combine standard of grid computing into web service, and the technology is quite mature and widely used in varied industries. It is natural that choosing SOA as basic architecture of ETS as the essence of ETS is a service-oriented data processing platform.

4 ETS Case Study

The Shanghai Grid-based Traffic information Service Platform (SHGTISP) is based exactly on the ETS concept as mentioned above. SHGTISP is an integrated system, which consists of an online mecro-traffic-simulation and a computational support platform. The computational support platform of traffic information service is based on a real-time vehicle information collection system. The purpose of the platform is to simulate the urban road network traffic dynamically using real-time trajectory data of taxis and buses in Shanghai, estimating and predicting road network status, and then providing the decision-makers, traffic managers and travelers with traffic information such as an analysis report of usage of urban transport and roads, an evaluation of public transit and terminal, dynamical trip planning for travelers.

The system architecture of SHGTISP is shown as Figure 2.

The technologies in SHGTISP are based on the Shanghai Grid, a web service oriented architecture, using Globus toolkit as a development platform. Those

traffic-relative models, dynamical OD estimating and predicting, data consistency check, estimating and predicting of traffic states, historical data verifying, management of data warehouse, online analyzing, are all deployed on the grid nodes as web services. Almost all tasks in SHGTISP are accomplished by parallel algorithm, sharing a same database server. A combination of the computational results is also a web service for external application. All the services are organized by a traffic information service managing module, its major function is dispatching and assigning computational sub-tasks parallel to corresponding grid node through the shanghai grid middle ware. Figure 3 is one interface of SHGTISP.

5 Conclusions

Transportation theories absolutely need to be verified by experiments. Experimental transportation becomes possible and feasible due to rapidly developing modern technologies such as information technologies, computer technologies and telecommunication technologies. The implementation of ITS brings about the more and more complicated traffic systems, as well as plenty of data with which we can further explore the world. How to make full use of data to research into the evolutional features and rules of traffic systems, so as to provide better support for traffic evaluation, optimization, management, control and information dissemination, is the main task of ETE. ETS is an application of ETE in modern traffic systems. Now what ETS can be presented with is subject to explanation and observation of data, known as passive experiment. ETS aims to build up the integration of passive experiments and, those that are the most important, active experiments. It will need further research into computational experimental system by way of studying complex system, building a simulation system that can perform as the real world as possible.

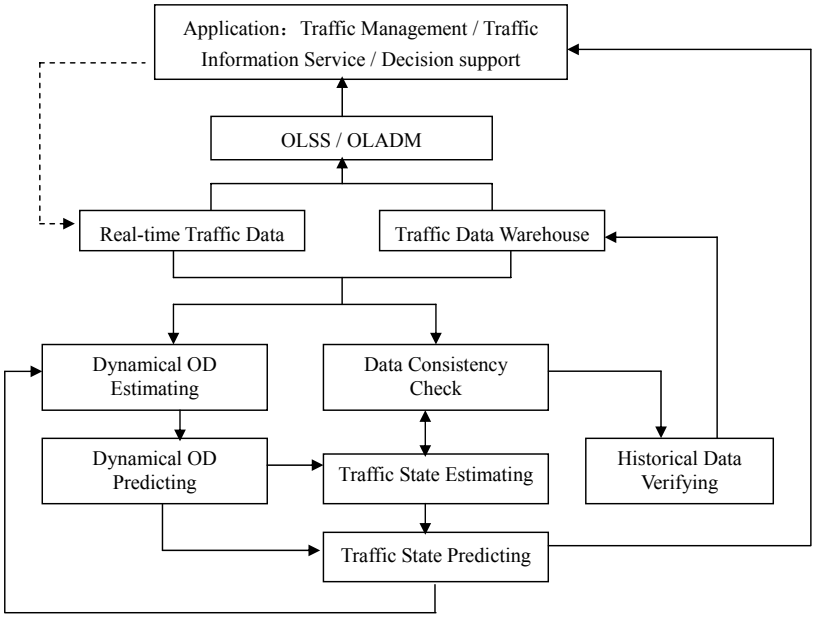


Figure 2. System architecture of SHGTISP ETS

Figure 3. Interface of SHGTISP ETS

Acknowledgement

This research is supported by funding from the National Natural Science Foundation of China (Grant No. 70631002) and the grand project of Science and Technology Commission of Shanghai. (Project No. 05dz15005).

References

Wang Feiyue, *"Computational Experiments for Behavior Analysis and Decision Evaluation of Complex System"*, Journal of System Simulation, Vol 16, No 5.PP893-897, 2004.5

W.H.Inmon, *"Building the Data Warehouse"*, China Machine Press, 2000.5

Yang Xiaoguang and Sun Jian, *"Basic theory and research prospects of Experimental Traffic Engineering"*, in Proceeding of WEC2004, Shanghai, 2004.8

Yang Xiaoguang, Sun Jian, Xu Jianmin,and Yi Sirong, *"Basic Theory (Approaches) of Experimental Traffic Engineering and Information Technology"*, Research on Discipline Development Strategy of Civil Engineering, Science Press, 2005.11

RESEARCH ON APPROPRIATE ROAD GRADE PROPORTION OF URBAN ROAD NETWORK

Haiyuan LI[1] and Wei DENG[2]

School of Transportation, Southeast University, No. 2, Sipailou, Nanjing, P. R. China, 210096; PH: 86-25-83795857; Fax: 86-25-83794102; email: timber2001@163.com

Abstract: One of the direct reasons that cause urban traffic problems is obscure functions and unreasonable road proportion in grades. Therefore, based on the statistic analysis of trip investigation data in several typical cities, the function with double parameters is adopted to fit the relationship between Vehicle Trip Distance (VTD) and trip proportion. Furthermore, the regressing analysis of the fitting parameters is made by three infection factors. In addition, according to VTD-turnover, coupling mechanism, and equilibrium mechanism of traffic supply-demand, equilibrium equations are built to show the equilibrium relationship, between vehicle supply-demand turnovers in each grade of roads. As a result, urban road proportion in grades is deduced and the reasonability of urban road grade scheme is evaluated by an evaluation coefficient. According to urban scales and fields of the fitting parameters, fields of urban road proportions in grades are recommended in different scale cities. Finally, one example is employed to prove the effectiveness of the equilibrium relations and the evaluation measure proposed in this paper.

Key words: *Road Grade Proportion, Supply-Demand Equilibrium, Vehicle Trip Distance (VTD), Transportation Turnover, Evaluation Method*

1 Introduction

Whether the grade structure of a road network is reasonable or not affects the road network playing its role normally and further has an impact on the total capacity of the urban transportation system. The main reason why most cities in China have low efficient road network results from the ambiguity of road function and the unreasonable road grade proportion. Because of the diversity of urban distribution, trip distances, trip demand characteristics, etc, cities with different scale have different road proportions in grades. Even in the cities with the same scale, the optimal road proportions in grades are different due to the great difference of trip characteristics. Currently there are few results about measures of urban road proportions as well as little evaluation of the reasonability of road grade proportions. Fei SHI put forward VTD-turnover measures through the fitting analysis of VTD distribution (Fei SHI,2006). This paper discusses the urban road grading measures from the aspects of the VTD-turnover coupling mechanism and the equilibrium mechanism of supply

turnover and the demand turnover in each grade of roads.

2 Analysis of Vehicle Trip Distance

2.1 Data Collection and Fitting Analysis

Through the analysis of citizens'OD data in more than ten typical cities, we choose the data with automobile trip (bus, car, taxi, motor bicycle, company car, etc). After considering the different speeds of different automobiles with different trip purposes and time periods, as well as the average carrying capacity and conversion factors of different types of vehicles, we gain the VTD distribution. By the fitting analysis of VTD, we find there are some rules of VTD. The characteristics of VTD distribution are similar to Rayleigh distribution. Based on the Rayleigh distribution, we adopt a function with double parameters to fit. When $\beta = 0.5$, it is Rayleigh function. Distribution fitting function and density function of VTD are

$$F(r) = 1 - \exp(-\beta\lambda r^2)/(2\beta) \tag{1}$$

$$f(r) = \lambda r \exp(-\beta\lambda r^2) \tag{2}$$

Where λ and β are double parameters; and r is vehicle trip distance. Figure 1 shows density function curves with different λ and β.

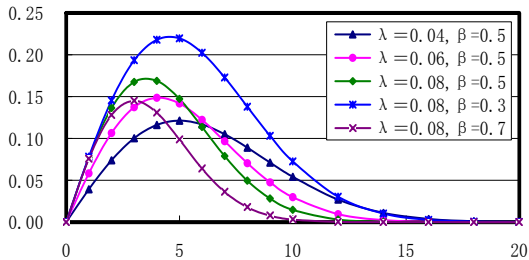


Figure 1. Density function curves with different λ and β

2.2 Regressing Analysis of the Fitting Parameters

We make the regressing multi-analysis according to the variation of λ and β and considering the urban land use scale, urban configuration and passenger transit structure as the three infection factors of λ and β. The result is

$$\lambda = -0.00302x_1 + 0.01903x_2 - 0.00708x_3 + 0.07013 \tag{3}$$

$$\beta = -0.00406x_1 + 0.59302x_2 - 0.14006x_3 + 0.00903 \tag{4}$$

In the function, X_1 is the urban land use the infection factor by the square root of established land area, $\sqrt{S}$; X_2 is the urban configuration coefficient that is the ratio of radii between a standard circle and an actual land circle (excluding waters and other lands). The standard circle has a center that is the city center as well as a radius that

can make a circle covering all the using land. The actual land circle's area is S_0, so the coefficient is $\sqrt{S/S_0}$. If the city has several centers, we can fix the circle's center at the shape's center which is composed of these several central areas. If the city is made up of separated groups, we can compute each group's S and then sum up them. X_3 is the infection factor of urban passenger transit structure expressed as the ratio of public transit trip and private car trip, and appears as "ln".

2.3 Fields of the Fitting Parameters

Approximate fields of the fitting parameters can be determined by research on infection factors of different cities in China. According to the land use scale, city types can be classified into small city (10-20km^2), middle city (30-50km^2), large city (50-200km^2), and metropolis (200-400km^2). Based on the research on passenger transit structure (Shaorong HU,2004; Yuan XU,2001), approximate fields of the fitting parameters can be obtained in different mass cities. See Table 1.

Table 1. Fields of the fitting parameters in different mass cities

City Types	X_1	X_2	X_3		λ		β	
			First Stage	Second Stage	First Stage	Second Stage	First Stage	Second Stage
Small City	3.16-4.47	0.8-1	0-1	-0.7--0.3	0.065-0.076	0.074-0.084	0.325-0.589	0.507-0.687
Middle City	4.47-7.07	0.8-1	0-1	-0.3--0.155	0.057-0.076	0.065-0.078	0.315-0.584	0.477-0.626
Large City	7.07-14.14	0.8-1	0.4-0.845	-0.155-0.4	0.037-0.065	0.040-0.069	0.308-0.518	0.371-0.595
Metropolis	14.14-20	0.8-1	0.4-0.845	0.4-0.65	0.019-0.044	0.021-0.044	0.285-0.489	0.312-0.489

In table 1, first stage is the developing stage of transportation structure; second stage is the stable stage of transportation structure.

3 Analysis of Vehicle Demand Turnover

3.1 Traffic Volume Proportion and Average Trip Distance

In order to solve vehicle demand turnovers in each grade of roads, traffic volumes and average trip distances of the farthest vehicle trips in every grade of roads needs to be determined. Through the analysis of distribution characteristics of VTD, after the fitting parameters λ and β are determined, the distribution density function of VTD can be obtained and the relationship between VTD and vehicle trip proportion can be built. Figure 2 shows the relationship between VTD and vehicle trip proportion. Furthermore, if critical maximum trip distances of the farthest vehicle trips in every grade of roads can be known, according to the relationship between VTD and vehicle trip proportion, traffic volume proportions and average trip distances of that can be calculated.

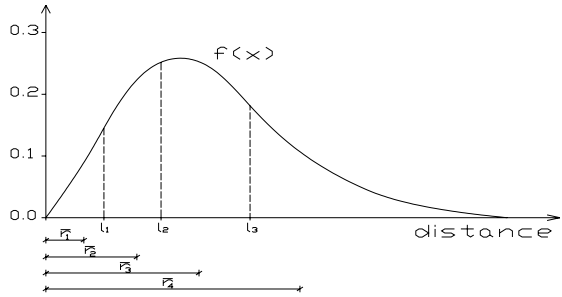


Figure 2. The relationship between VTD and vehicle trip proportion

Then the traffic volume proportion and average trip distance of the farthest vehicle trips on spur track read

$$Q_1^{'} = \int_0^{l_1} f(r)\,dr \tag{5}$$

$$\bar{r_1} = \frac{\int_0^{l_1} f(r) r dr}{\int_0^{l_1} f(r) dr} \tag{6}$$

The traffic volume proportion and average trip distance of the farthest vehicle trips on sub-arterial read

$$Q_2^{'} = \int_{l_1}^{l_2} f(r)\,dr \tag{7}$$

$$\bar{r_2} = \frac{\int_{l_1}^{l_2} f(r) r dr}{\int_{l_1}^{l_2} f(r) dr} \tag{8}$$

The traffic volume proportion and average trip distance of the farthest vehicle trips on arterial read

$$Q_3^{'} = \int_{l_2}^{l_3} f(r)\,dr \tag{9}$$

$$\bar{r_3} = \frac{\int_{l_2}^{l_3} f(r) r dr}{\int_{l_2}^{l_3} f(r) dr} \tag{10}$$

The traffic volume proportion and average trip distance of the farthest vehicle trips on high-speed road read

$$Q_4^{'} = \int_{l_3}^{+\infty} f(r)\,dr \tag{11}$$

$$\bar{r_4} = \frac{\int_{l_3}^{+\infty} f(r) r dr}{\int_{l_3}^{+\infty} f(r) dr} \tag{12}$$

With $\sum_{i=1}^{4} Q_i^{'} = \int_0^{+\infty} f(r)dr = 1$ (13)

Where $Q_1^{'}$-$Q_4^{'}$ are respectively traffic volume proportions of the farthest vehicle trips in every grade of roads; $\overline{r_1}$-$\overline{r_4}$ are respectively average trip distances of the farthest vehicle trips in every grade of roads; l_1-l_3 are respectively critical maximum trip distances of the farthest vehicle trips in every grade of roads; and $f(r)$ is distribution density function of VTD.

3.2 Calculation of Vehicle Demand Turnover

Let Z_i denote turnover of unit traffic volume of the farthest vehicle trips in *i* grade of road, we have

$$Z_i = \sum_{j=1}^{i} z_{ij}, \; i = 1, 2, 3, 4 \tag{14}$$

Where z_{ij} is a partial turnover that Z_i distributes turnovers to j（$j \le i$）grade road.

Vehicle turnovers on roads that are less than *i* grade nearly equal the value of multiplying traffic volume of the farthest vehicle trips in *i* grade of road by average distance of the farthest vehicle trips in *i*-1 grade of road.

Then the turnover of the farthest vehicle trips on spur track reads

$$Z_1 = z_{11} = \int_0^{l_1} f(r) r dr \tag{15}$$

The turnover of the farthest vehicle trips on sub-arterial reads

$$Z_2 = \int_{l_1}^{l_2} f(r) r dr = z_{21} + z_{22} \tag{16}$$

With $z_{21} = \overline{r}_1 \int_{l_1}^{l_2} f(r)dr$; and $z_{22} = \int_{l_1}^{l_2} f(r)(r - \overline{r}_1)dr$

The turnover of the farthest vehicle trips on arterial reads

$$Z_3 = \int_{l_2}^{l_3} f(r) r dr = z_{31} + z_{32} + z_{33} \tag{17}$$

With $z_{31} = \overline{r}_1 \int_{l_2}^{l_3} f(r)dr$; $z_{32} = (\overline{r}_2 - \overline{r}_1) \int_{l_2}^{l_3} f(r)dr$; and $z_{33} = \int_{l_2}^{l_3} f(r)(r - \overline{r}_2)dr$

The turnover of the farthest vehicle trips on high-speed road reads

$$Z_4 = \int_{l_3}^{+\infty} f(r) r dr = z_{41} + z_{42} + z_{43} + z_{44} \tag{18}$$

With $z_{41} = \overline{r}_1 \int_{l_3}^{+\infty} f(r)dr$; $z_{42} = (\overline{r}_2 - \overline{r}_1) \int_{l_3}^{+\infty} f(r)dr$; $z_{43} = (\overline{r}_3 - \overline{r}_2) \int_{l_3}^{+\infty} f(r)dr$; and

$z_{44} = \int_{l_3}^{+\infty} f(r)(r - \overline{r}_3)dr$

Therefore, vehicle turnovers in each grade of roads can be obtained. Let $G_j^{'}$ denote vehicle turnover of unit traffic volume in each grade of roads. We get

$$G_j^{'} = \sum_{i=j}^{4} z_{ij} \ , j = 1, 2, 3, 4 \tag{19}$$

Then, we have

$$G_1^{'} = \sum_{i=1}^{4} z_{i1} = \int_0^{l_1} f(r)rdr + \overline{r}_1 \int_{l_1}^{l_2} f(r)dr + \overline{r}_1 \int_{l_2}^{l_3} f(r)dr + \overline{r}_1 \int_{l_3}^{+\infty} f(r)dr = \int_0^{l_1} f(r)rdr + \overline{r}_1 \int_{l_1}^{+\infty} f(r)dr \tag{20}$$

$$G_2^{'} = \sum_{i=2}^{4} z_{i2} = \int_{l_1}^{l_2} f(r)(r - \overline{r}_1)dr + (\overline{r}_2 - \overline{r}_1) \int_{l_2}^{l_3} f(r)dr + (\overline{r}_2 - \overline{r}_1) \int_{l_3}^{+\infty} f(r)dr = \int_{l_1}^{l_2} f(r)(r - \overline{r}_1)dr + (\overline{r}_2 - \overline{r}_1) \int_{l_2}^{+\infty} f(r)dr \tag{21}$$

$$G_3^{'} = \sum_{i=3}^{4} z_{i3} = \int_{l_2}^{l_3} f(r)(r - \overline{r}_2)dr + (\overline{r}_3 - \overline{r}_2) \int_{l_3}^{+\infty} f(r)dr \tag{22}$$

$$G_4^{'} = \sum_{i=4}^{4} z_{i4} = \int_{l_3}^{+\infty} f(r)(r - \overline{r}_3)dr \tag{23}$$

With $$G^{'} = \sum_{j=1}^{4} G_j^{'} = \int_0^{+\infty} f(r)rdr \tag{24}$$

Where $G_1^{'}$-$G_4^{'}$ are respectively vehicle turnovers of unit traffic volume in each grade of road; and $G^{'}$ is total vehicle turnover, Table 2 shows distribution relations of vehicle turnovers in each grade of roads (Fei SHI,2006).

Table 2. Distribution relations of vehicle turnovers in each grade of roads

Road Grades	Spur Track	Sub-arterial	Arterial	High-speed Road	Σ
Spur Track	z_{11}				Z_1
Sub-arterial	z_{21}	z_{22}			Z_2
Arterial	z_{31}	z_{32}	z_{33}		Z_3
High-speed Road	z_{41}	z_{42}	z_{43}	z_{44}	Z_4
Σ	$G_1^{'}$	$G_2^{'}$	$G_3^{'}$	$G_4^{'}$	$G^{'}$

4 Analysis of Vehicle Supply Turnover

According to the study on vehicle supply turnover on unit length road (Fei SHI,2006), vehicle supply turnovers in each grade of roads can be calculated by multiplying vehicle supply turnovers on unit length road by road lengths in every grade. Vehicle supply turnovers on unit length road in every grade can be given by the following equation (Fei SHI,2006).

$$N_i^{'} = 2N_{0i}k_{1i}k_{2i}k_{3i}k_{4i}k_{5i}n_i / k_{6i} \tag{25}$$

Where N_{0i} is theoretical capacity in *i* grade of road; k_{1i} is average saturation in *i* grade of road and equals to 0.6 in the paper; k_{2i} is lane width factor of *i* grade road; k_{3i} is intersection factor of *i* grade road; k_{4i} is the bicycle interference factor of *i* grade road; k_{5i} is average frequency coefficient of using *i* grade road and is equal to 0.7-0.8 in the paper; n_i is lane number factor of *i* grade road; and k_{6i} is non uniform

factor of two-way traffic in *i* grade of road.

Table 3 shows theoretical capacity per-hour of single lane in each grade of roads. Table 4 shows proposal values of lane width factor. Table 5 shows proposal values of intersection factor. Table 6 shows proposal values of bicycle interference factor. Table 7 shows proposal values of lane number factor. Table 8 shows proposal values of non uniform factor of two-way traffic. Table 9 shows vehicle supply turnovers per-hour per-kilometer in each grade of roads (Fei SHI,2006).

Table 3. Theoretical capacity per-hour of single lane in each grade of roads

Road Grades	High-speed Road	Arterial	Sub-arterial	Spur Track
Theoretical Capacity of Single Lane(pcu/h)	1800	1730	1640	1380

Table 4. Proposal values of lane width factor

Lane Width(m)	2.5	3.0	3.5	4.0	4.5	5.0	5.5	6.0
k_{2i}	0.50	0.75	1.0	1.11	1.2	1.26	1.29	1.30

Table 5. Proposal values of intersection factor

Road Grades	High-speed Road	Arterial	Sub-arterial	Spur Track
k_{3i}	0.9	0.7	0.5	0.4

Table 6. Proposal values of bicycle interference factor

Road Grades	High-speed Road	Arterial	Sub-arterial	Spur Track
k_{4i}	1	1	0.9	0.8

Table 7. Proposal values of lane number factor

Lane Number	1	2	3	4
n_i	1	1.87	2.60	3.20

Table 8. Proposal values of non uniform factor of two-way traffic

Road Grades	High-speed Road	Arterial	Sub-arterial	Spur Track
k_{6i}	1.2	1.2	1.1	1.0

Table 9. Vehicle supply turnovers per-hour per-kilometer in each grade of roads

Road Grades	High-speed Road	Arterial	Sub-arterial	Spur Track
Vehicle Turnover(*pcu·km/h·km*)	3033	2210	1016	334

Vehicle supply turnovers in each grade of road can be determined by the following equation.

$$N_i = L_i N_i^{'}, \quad i=1, 2, 3, 4 \tag{26}$$

Where L_i is total length of *i* grade road.

5 Equilibrium Relationships between Vehicle Supply-Demand Turnovers

According to the equilibrium mechanism of traffic supply-demand on road network, equilibrium relationships between vehicle supply-demand turnovers can be built in each grade of roads.

$$G_i = QG_i^{'} = N_i = L_i N_i^{'}, \quad i=1, 2, 3, 4 \tag{27}$$

Where G_i is turnover of vehicle trips in *i* grade of road; and Q is total traffic demand volume.

Then the equilibrium relationship between vehicle supply-demand turnovers on spur track reads

$$Q\left[\int_0^{l_1} f(r)rdr + \overline{r}_1 \int_{l_1}^{+\infty} f(r)dr\right] = L_1 N_1^{'} \tag{28}$$

The equilibrium relationship between vehicle supply-demand turnovers on sub-arterial reads

$$Q\left[\int_{l_1}^{l_2} f(r)(r-\overline{r}_1)dr + (\overline{r}_2-\overline{r}_1)\int_{l_2}^{+\infty} f(r)dr\right] = L_2 N_2^{'} \tag{29}$$

The equilibrium relationship between vehicle supply-demand turnovers on arterial reads

$$Q\left[\int_{l_2}^{l_3} f(r)(r-\overline{r}_2)dr + (\overline{r}_3-\overline{r}_2)\int_{l_3}^{+\infty} f(r)dr\right] = L_3 N_3^{'} \tag{30}$$

The equilibrium relationship between vehicle supply-demand turnovers on high-speed road reads

$$Q\left[\int_{l_3}^{+\infty} f(r)(r-\overline{r}_3)dr\right] = L_4 N_4^{'} \tag{31}$$

6 Calculation of Road Proportion in Grades

According to the equilibrium relationships between vehicle supply-demand turnovers in each grade of road, lengths of roads in every grade can be respectively calculated. Furthermore, urban road proportion in grades can be deduced.

$$L_4 : L_3 : L_2 : L_1 = G_4^{'}/N_4^{'} : G_3^{'}/N_3^{'} : G_2^{'}/N_2^{'} : G_1^{'}/N_1^{'} \tag{32}$$

Based on the research about vehicle preponderant trip distances in each grade of road (Fei SHI,2006), critical maximum trip distances of the farthest vehicle trips in each grade of roads may adopt the upper limits of vehicle preponderant trip distances in each grade of roads. Table 10 shows upper limits of vehicle preponderant trip

distances in each grade of roads (Fei SHI,2006).

Table 10. Upper limits of vehicle preponderant trip distances in each grade of roads

Road Grades	Upper Limits of Vehicle Preponderant Trip Distances
Spur Track	$L_{1\max} = (d_{1EW} + d_{1SN}) + \max(d_{2EW}, d_{2SN})$
Sub-arterial	$L_{2\max} = (\frac{d_{2EW} + d_{2SN}}{2}) + \max(d_{3EW} + d_{2SN} + \frac{d_{1EW} - d_{1SN}}{2}, d_{3SN} + d_{2EW} + \frac{d_{1SN} - d_{1EW}}{2})$
Arterial	$L_{3\max} = (\frac{d_{3EW} + d_{3SN}}{2}) - (\frac{d_{1EW} + d_{1SN}}{2}) + \min(d_{2SN} + d_{3EW} + d_{4SN}, d_{2EW} + d_{3SN} + d_{4EW})$

In table 10, $L_{1\max}$ - $L_{3\max}$ are respectively the upper limits of vehicle preponderant trip distances in each grade of roads; d_{1EW}- d_{4EW} are respectively east-west average intervals of roads in every grade; d_{1SN} - d_{4SN} are respectively south-north average intervals of roads in every grade.

Compared with proposal values of road density in different scale cities recommended by "urban road transportation planning and designing standard" (GB50220-95), fields of urban road proportion in grades are put forward in different scale cities according to the VTD-turnover method proposed in the paper. Table 11 shows proposal values of urban road proportion in grades in different scale cities.

Table 11. Proposal values of urban road proportion in grades in different scale cities

City Types	High-speed Road	:	Arterial	:	Sub-arterial	:	Spur Track
Small City	-		1	:	(1.3-1.8)	:	(3.2-4.5)
Middle City	-		1	:	(1.2-1.6)	:	(2.9-4.1)
Large City	-		1	:	(1.1-1.5)	:	(2.7-3.8)
Metropolis	(0.1-0.5)	:	1	:	(1.0-1.4)	:	(2.5-3.2)

In table 11, proposal values of urban road proportion in grades are obtained under two conditions. First of all, the blending degree of utilizing land is even or trip generation and trip attraction are uniform. Secondly, the average lane numbers of high-speed roads, arterials, sub-arterials and spur tracks are respectively six lanes, six lanes, four lanes and two lanes.

7 Evaluation of Road Proportion in Grades

In order to evaluate the reasonability of the urban road grade scheme, we can compare the existing road proportion in grades with the road proportion computed by the VTD-turnover measures. Hereby, the results could show the lacking or excrescent extent of roads with evaluating coefficient (Fei SHI,2006).

$$D = \frac{\sum_{i=1}^{n} \left| \frac{x_i - \hat{x}_i}{\hat{x}_i} \right|}{n} \cdot 100\% \tag{33}$$

Where D is the evaluating coefficient; x_i is the proportion in length of existing roads in each grade; $\hat{x}_i$ is the proportion in length by the measure of VTD-turnover; and n is the total types of roads.

If the proportion of roads in one grade is less or more than 40%, it is greatly unreasonable and 25% is the given bound. 15% is the given bound of reasonable and unreasonable and 5% is for extraordinarily reasonable and quite reasonable: then D≤5, extraordinarily reasonable. 5<D≤15, quite reasonable. 15<D≤25, not reasonable, D>25, terribly unreasonable.

8 Empirical Study

We take a round city with the area of 50km^2 (figure coefficient 0.96) as an example, that has a typical transportation structure with $\lambda = 0.08$, $\beta = 0.5$. VTD distribution density function is: $f(r) = 0.08r\exp(-0.04r^2)$. Critical maximum trip distances of the farthest vehicle trips may adopt the upper limits of predominant trip distances. We have l_1=1.46km, l_2=3.53km and l_3=13.8km. Then we compute the proportion of traffic volume, average trip distance and the road proportion in grades by the measures we talked before. The results are shown in table 12.

Table 12. Computing results of the example

Average Trip Distance(km)				Traffic Volume Proportion(%)				Road Proportion in Grades
$\overline{r_1}$	$\overline{r_2}$	$\overline{r_3}$	$\overline{r_4}$	Q_1'	Q_2'	Q_3'	Q_4'	L_3:L_2:L_1
1.0	2.6	5.8	-	8	31	61	-	1:1.6:3.2

The results show that the city has no need for travel demand of high-speed road. It agrees with the city's scale and the requirement of VTD. The proportion of volume in each grade of road presents as a pyramid up side down, whereas the grading proportion of roads presents as a normal pyramid that is the rule. We can give the final result that the existing road proportion in grades (1:1.5:2.5) is quite reasonable through computing the assessing coefficient D = 9.4.

9 Conclusions

Here we may draw the following conclusions.

(1) Through the fitting analysis of citizens'OD data and regressing analysis of the fitting parameters, we get the distribution function and density function of VTD with double parameters and build the relationship between VTD and vehicle trip proportion.

(2) The paper constructs the equilibrium relationships between vehicle supply-demand turnovers in each grade of roads according to the VTD-turnover coupling mechanism and the equilibrium mechanism of traffic supply-demand.
(3) Urban road proportion in grades is deduced and fields of road grade proportion are put forward in different scale cities that provide a valuable reference for determining urban road proportion in grades.
(4) The reasonability of urban road grade scheme is evaluated by the assessing coefficient and evaluation standard is given.

Acknowledgements

This research has been funded by a national project from the Ministry of Science and Technology (Project no. 2006BAJ18B03) and a grant for an excellent doctor dissertation from Southeast University, P. R. China.

References

Anwen ZHENG. (2002). "Research on traffic supply capacity of urban roads." *Journal of basic science and engineering,* Vol.10, No.4.

Chaoqun MAR. (2003). "Research on development level of urban road in China." *Master thesis*, Chang'An University.

Fei SHI. (2006). "Research on grade proportion and layout method of urban road." *Doctor dissertation*, South East University.

Jianjun WANG, Jiping WANG, Zhiqun PENG. (2005). "Discussion on appropriate grade proportion of urban road network." *Urban transport of China*, Vol.3, No.1.

National standard in China. (1995). Urban road transportation planning and designing standard. GB50220-95.

Shaorong HU. (2004). "Research on equilibrium theory of traffic supply-demand on urban road." *Master thesis*, Chang'An University.

Wei WANG, Xiucheng GUO. (2000). "Transportation engineering." South East University Publishing Company.

Xuewu CHEN. (2002). "Research on the sustainable model of urban transportation system." *Doctor dissertation*, South East University.

Yuan XU. (2001). "Research on road network planning in large city—grade proportion of road network and optimization of road function." *Master thesis*, South East University.

Bus Transit Service Optimization – The State-of-the-Art, State-of-the-Practice, and Challenges

Fang Zhao
Lehman Center for Transportation Research, Department of Civil and Environmental Engineering, Florida International University, Miami, FL 33199, USA; PH: (305) 348-3821; FAX: (305) 348-2802; email: zhaof@fiu.edu

Abstract

Transit network design is concerned with laying out transit routes, designing headways, and determining a transit network schedule. A transit route network optimization process attempts to find the route network structure with the minimum number of transfers, maximum ridership coverage, reasonable route directness, minimum access and egress times, minimum waiting and travel times, and minimum operating cost. One of the main difficulties faced by transit network optimization processes is combinatorial intractability. While research has made headways in the area of computing, the problem remains a challenging one. This paper provides a review of some methods that have been proposed to solve large-scale transit network optimization problems. The paper further explores the nature of transit service optimization problems in a large context and discusses the challenges and post research questions given the state-of-the-art and state-of-the-practice.

Introduction

Public transit is an alternative to traveling by private automobiles. This alternative has becoming more promising to many transportation planners and elected officials as congestion continues to worsen in many urban areas and the pressure on the transportation infrastructure continues to grow. Congestion due to increased use of private automobiles is not a problem faced only by cities in developed countries. Urban areas in many developing countries are experiencing a rapid growth in car ownership and congestion, and the infrastructure is often unable to keep pace with the demand. In addition to congestion, increased travel by private automobiles is causing other environmental problems, including pollution (noise, air, and water), urban sprawl, longer commute time, decrease in quality of life, and, by way of construction of new roads and highways, loss of agriculture land, green space, and historical and cultural treasures.

To make public transit an attractive mode for travel requires significant investments and improvements in transit service quality. One option is to build an extensive subway network, which is expensive and infeasible, or ineffective for small or medium sized cities. While some of the larger urban areas are looking to building a network of rail transit, many are finding other ways to attract transit users. These include increasing service frequencies; deploying more comfortable, more convenient,

and cleaner buses; improving bus stop facilities to add comfort, security, and convenience; coordinating transit schedules; improving service reliability with Automatic Vehicle Location (AVL) systems; disseminating real-time travel information to transit users; using more convenient fare collection systems; etc. While contributing to the enhancement of the appeal of transit, these measures do not, in a significantly way, reduce in-vehicle travel time or transfers. A third approach is to optimize the transit system configuration, which includes the network layout, service frequency, and schedule. Optimization of a transit network (TN) not only has the potential in reducing travel time and transfers and improving the overall transit service quality, but also offers opportunities to transit operators to minimize operating costs.

The design of route network, service frequencies, and vehicle timetables for a given budget are the most important design tasks in a transit planning process. They can be formulated as a mathematical program problem, with the decision variables being the route configuration, service frequencies, and schedules. The solution space for such a problem can be extremely large, due to the combinatorial nature of the problem. The mathematical optimization systems derived from realistic combinatorial TN problems are usually at least NP-hard, which refers to problems for which the number of elementary numerical operations is not likely to be expressed or bounded by a function of polynomial form (Garey and Johnson 1979). The NP-hard intractability is due to the need to search for optimal solutions from a huge search space made up by all possible solutions.

In the last several decades, many TN design and optimization approaches have been proposed. Some of them show the potential to solve realistic real-world problems. However, much of the attention has been paid to finding solution strategies, and the issues related to making them practical have not been addressed.

This paper provides a review of the literature and some methods that have been proposed to solve large-scale transit network optimization problems. The paper further explores the nature of transit service optimization problems in a larger context and discusses the challenges and post research questions given the state-of-the-art and state-of-the-practice.

Literature Review

There are several approaches to solving the transit network design problem, mainly mathematical, heuristic, and metaheuristic. Mathematical approaches seek direct solutions to mathematical programming problems. For instance, Wu and Murray (2005) proposed an exact approach to solve multiple-route TN optimization problems. It was applied to solve a dense two-route TN system, with good improvements over the existing system. Recently, Matisziw *et al*. (2006) described an exact mathematical approach for finding the optimal route extensions of an existing TN. The success of mathematical optimization methods, however, has been limited in dealing with large-scale realistic TN optimization problems. Newell (1979) pointed out that TN design optimization "is generally a non-convex (even concave) optimization problem for

which no simple procedure exists short of direct comparisons of the various local minima." Baaj and Mahmassani (1991) observes that mathematical approaches to TN optimization has a number of difficulties, such as non-linearity, multi-objectives, and combinatorial intractability due to the discrete nature of TN systems. Similar observations are also made by Gao *et al*. (2002) and Fan and Machemehl (2004).

To avoid the difficulties of mathematical approaches, the majority of the solution methods proposed have relied on problem-specific heuristic assumptions or design guidelines. Mandl (1979) presented a heuristic algorithm to compare and select better route sets based on reductions in average costs. Baaj and Mahmassani (1991) developed an artificial intelligence (AI) based heuristic approach, which involved the evaluation of multiple objective functions and feasibility constraints at various solution search stages. This method was extended by Shih and Mahmassani (1994) and applied to transit network design with schedule coordination at transfer centers. While heuristic methods are always able to find solutions, these solutions are usually not optimal or even close to optimal solution because the solution space is incomplete. Search methods guided by heuristic rules also limit the possibility of find an optimal solution.

A third approach to transit network optimization is by using metaheuristic methods that do not rely on specific guidelines but, instead, employ general-purpose search methods, also referred to as weak methods. One example of metaheuristic method is the bi-level programming model developed by Gao *et al*. (2002). The model and algorithm are illustrated through a small example. Agrawal and Mathew (2004) implemented a parallel version of a genetic algorithm (GA) and applied it to a large-scale transit network but with a relatively limited number of origin-destination (OD) pairs. Zhao (2006) described a greedy-type search method for optimizing both transit network layout and headways based on a user cost function. Zhao and Zeng (2006) proposed a global search method based on combined GA and simulated annealing (SA) method. In both (Zhao and Zeng 2006) and (Zhao 2006), a static, daily passenger trip demand is assumed and schedule design is not considered. One advantage of the metaheuristic approach is that it does not make assumptions that are based on personal experiences or preferences. When properly defined, the solution space is less likely to exclude potential optimal solutions. The search methods are generic and not domain or geographically specific, therefore have a better chance of finding at least local optimal solutions than the heuristic methods.

Reviews of various earlier works in the TN optimization field may be found in (Van Nes *et al*. 1988; Fan and Machemehl 2004; and Zhao and Gan 2003).

Formulation of Transit Network Optimization Problems

A TN optimization problem can be stated as the determination of a set of transit routes, the corresponding route vehicle headways, and vehicle timetables, given the street network and the transit demand of the transit service area, subject to constraints on operating budgets, routes, headways, and timetables, to minimize the

total user cost of the TN service area. Mathematically, the optimization process can be stated as:

Minimize: $u_{\text{cost}}(\boldsymbol{T}, \boldsymbol{h}, \boldsymbol{S}, \boldsymbol{O}), \quad \forall \boldsymbol{T} \in \underline{\boldsymbol{T}}, \boldsymbol{h} \in \underline{\boldsymbol{H}}$ and $\boldsymbol{S} \in \underline{\boldsymbol{S}}$ (1)

Subject to:

Route headway, route length and total TN vehicle fleet size constraints:
$h_{\min}^{(q)} \le h_q \le h_{\max}^{(q)}$, $l_{\min}^{(q)} \le l_q \le l_{\max}^{(q)}$, ($q$ = 1, 2, ... , T) and $\sum_{q=1}^{T} v_q \le v_{\max}$,

Route load factor constraints: $L_q \le L_{\max}^{(q)}$, (q = 1, 2, ... , T) (2)

Route directness constraints: $d_q \ge d_{\min}^{(q)}$, (q = 1, 2, ... , T (3)

Certain route geometry or topology constraints: $g_q(\boldsymbol{T})$, (q = 1, 2, ... , T) (4)

where $\boldsymbol{T}$, $\boldsymbol{h}$, and $\boldsymbol{S}$ are, respectively, the unknown route network, vehicle headway vector, and vehicle timetable matrix; $\boldsymbol{O}$ represents a given transit demand origin-destination (OD) matrix; $\underline{\boldsymbol{T}}$, $\underline{\boldsymbol{H}}$, and $\underline{\boldsymbol{S}}$ are the search spaces for unknown variables $\boldsymbol{T}$, $\boldsymbol{h}$, and $\boldsymbol{S}$; h_q, $h_{\min}^{(q)}$, and $h_{\max}^{(q)}$ represent the route headway and the minimum and maximum route headway constraints on route q; T is the total number of routes in the network $\boldsymbol{T}$; l_q, $l_{\min}^{(q)}$, and $l_{\max}^{(q)}$ are the route length and the minimum and maximum route length constraints; v_q and $v_{\max}$ are, respectively, the route vehicle fleet size and the maximum allowed total vehicle fleet size of TN; L_q and $L_{\max}^{(q)}$ are the vehicle load factor and the maximum load factor constraint; d_q and $d_{\min}^{(q)}$ are the route directness and minimum route directness constraints.

The total user cost function may be the summation of individual transit users' travel times between all the OD pairs in the transit service area,

$$u_{\cos t}(\boldsymbol{T}, \boldsymbol{O}, \boldsymbol{h}, \boldsymbol{S}) = \sum_{i,j=1, i\neq j}^{N} \left[\sum_{s=1}^{o(ij)} u_{ij,s}(\boldsymbol{T}, \boldsymbol{h}, \boldsymbol{S}) \right], \tag{5}$$

where $u_{ij,s}(\boldsymbol{T}, \boldsymbol{h}, \boldsymbol{S})$ (s = 1, 2, … , $o(ij)$) is the travel time of the s^{th} trip between nodes n_i and n_j. Denote $t_{ij,s}^{(k)}$ (k = 0, 1, 2) as the travel time of a k-transfer trip (i.e., a trip that may be accomplished with k vehicle transfers) between a demand node pair n_i and n_j. This function can be extended to include out-of-pocket costs such as fares, although it will require a value-of-time factor to convert money units to time units.

The user cost function can also include terms that implicitly maximize the ridership. One approach is to penalize for any unserved OD trips with a large cost value. Therefore, trade-offs have to be made to provide the best services to passengers while

serving as many users as possible (Zhao 2006). The minimization of a function of both operator cost and user cost is described in (Zhao and Zeng 2007).

The route geometry/topology constrains $g_q(\boldsymbol{T})$ given in (4) can be imposed by the planner to allow certain desired features for the system under design: (a) routes with fixed paths; (b) route that must start from, and/or pass through, and/or end at certain locations; (c) route with fixed segments; (d) route with or without repeating street nodes. If all routes are subjected constraint (a), it becomes a TN schedule (headway and/or timetable) optimization problem.

In the route directness constraint (3), route directness is measured using the following formula:

$$d \equiv d(\boldsymbol{r}) \equiv \sum_{i=1}^{r-1} \sum_{j=i+1}^{r} \frac{u_{ij}^{(S)} / u_{ij}^{(R)}}{(r^2 - r)/2} \tag{6}$$

where r is the number of nodes on route $\boldsymbol{r}$, $u_{ij}^{(R)}$ is the user travel cost between nodes n_i and n_j along the route, $u_{ij}^{(S)}$ is the user travel cost following the shortest path in the street network between these two nodes. Route directness $d(\boldsymbol{r})$ is the average ratio between the travel cost along a route segment and that along the shortest path connecting the two end nodes of the route segment. The route directness $d(\boldsymbol{r})$ has the property of $d(\boldsymbol{r}) \leq 1$, with $d(\boldsymbol{r}) = 1$ being the optimum when all route segments coincide with the corresponding shortest paths between the ending points of these segments.

Solving Transit Network Optimization Problems with Meta-Heuristic Search

Various methods have been proposed to solve the problem defined above. Some of them include greedy search, branch-and-bound, simulated annealing (SA), genetic algorithms (GA), Tabu search, ant colony algorithms, etc. To be effective, a combination of these methods may be necessary.

Greedy search is a local search strategy, which follows the path that seems to be promising compared to a current solution without exploring alternative, and potentially better, paths.

In branch-and-bound methods, some integer sets and their descendant subsets are discarded if they have no chance of containing optimal results under given constraints. Alternatively, instead of searching through the entire branch-and-bound subsets, rollout algorithms may be implemented to allow the search to be performed along a sequence of subsets that are most probable to contain better solutions.

The SA method (Kirkpatrick *et al*. 1983) is a Monte Carlo simulation-based stochastic search algorithm. SA has a solid theoretical foundation and theoretically guarantees a global optimum. However, SA generally has exponential computational intractability with respect to the problem size, and needs interventions or enhancements to optimize or accelerate the random search process (Bertsekas 1998).

GA is a stochastic search technique based on the mechanism of natural selection and evolution. Unlike SA search methods and other traditional mathematical search methods that follow a single search thread at a time, GAs search simultaneously among a set of solutions, referred to as a solution population. While GAs enjoy wide applicability and popularity, especially for problems of complex nature, they suffer from combinatorial explosive solution costs due to huge solution search spaces for large-scale problems. Thus far in the literature of TRN research fields, most GA approaches seem to be limited to small optimization problems. However, instead of an independent solution search tool, a GA can be used to identify promising search locations whenever the current search process slows down or "freeze" near a local optimal point. Various GA search schemes can be found in (Goldberg 1989), and applications of GAs on TRN and TNS optimization in (Chakroborty and Dwivedi 2002, Bielli *et al*. 2002).

Tabu search schemes are described in (Bertsekas, 1998) and are designed to avoid repeated evaluation of the same solutions that have been explored previously. A description of greedy search, SA, tabu, and methods and local search spaces may be found in (Zhao and Gan, 2003).

The ant colony optimization algorithm (ACO), initial introduced by Marco Dorigo (1992), is a probabilistic technique based on swarm intelligence for solving computational problems of routing, inspired by how ants find paths from the colony to food. It has been applied to transit network design problems (Kuan *et al*. 2006, Yang *et al*. 2007)

The implementation of search strategies should allow for relatively complete solution search spaces, particularly because the prospect of solving TN optimization problems of realistic sizes with metaheuristic approaches has become increasingly promising with the rapid advancement of computing technologies. The following should also be considered when designing search strategies: (a) The method should be generally applicable to the design and optimization of a wide range of TN problems in practice, and should not favor a particular TN configuration; (b) Solutions obtained from this method should give reasonably good results in a reasonable amount of time, as permitted by the current computing power affordable to most transit agencies; (c) The results should improve and approach a global optimum as computer resources (mainly CPU processing speed) increase.

Numerical Examples

A number of solution strategies have been implemented by the author. The detailed descriptions of these solution strategies, including their implementations, may be found in (Zhao 2006, Zhao and Zeng 2006, Zhao and Zeng 2007). Here two numerical examples are presented. The results are obtained using a combined SA and GA method (SAGA) for optimizing both network layout and headways (Zhao and Zeng 2006).

Table 1. Comparison of Results from Different Methods (Zhao and Zeng 2006)

Problem Source	Mandl (1979)		Baaj & Mahmassani (1991)						Shih & Mahmassani (1994)			
Route layout case	1		1		2		3		1		2	
Search Method	Mandl	SAGA	B&M	SAGA	B&M	SAGA	B&M	SAGA	S&M	SAGA	S&M	SAGA
0-transfer trips (%)	69.94	95.12	78.61	94.03	79.96	95.95	80.99	92.87	82.59	94.03	87.73	93.00
1-transfer trips (%)	29.93	4.88	21.39	5.97	20.04	4.05	19.01	7.13	17.41	5.97	12.27	7.00
2-transfer trips (%)	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Total user travel cost (min)	219094	181708	205646	184267	209318	186545	217954	183823	203936	186822	204028	190990
Savings in total user travel cost (hr)	-	565.60	-	244.13	-	230.87	-	457.93	-	241.27	-	117.87
Total Fleet sizes	99	99	89	89	77	77	82	82	84	84	68	68
Number of routes	4	4	6	6	8	8	7	7	6	6	8	8
Average Boarding	1.302	1.048	1.214	1.060	1.200	1.045	1.190	1.000	1.174	1.060	1.123	1.070
Network directness	0.811	0.971	0.912	0.928	0.959	0.970	0.856	0.908	0.835	0.908	0.940	0.979
CPU (sec)	-	7.56	-	17.91	-	21.98	-	14.58	-	22.15	-	25.38

The first problem and its solutions are given by Mandl (1979), which are also used by Mahmassani and Baaj (1991) and Shih and Mahmassani (1994) as a benchmark problem in their studies. Mandl's problem consists of a street network of 15 nodes with a total demand of 15570 trips per day. The length of a street segment is defined in terms of in-vehicle travel time in minutes. The transfer penalty is five minutes, and the maximum route length constraint is 40 minutes. In Table 1, the first row identifies the source of the solutions to the benchmark problem. The second row indicates solutions to the benchmark problem with different numbers of routes. The methods used to obtain the results are indicated in the third row, where SAGA indicates the method of this study, Mandl indicates Mandl's method, B&M stands for Baaj and Mahmassani's method, while S&M for Shih and Mahmassani's method. For each solution, the unshaded column provides the statistics for the layouts produced in the original studies, and the shaded column gives the statistics for the results produced from the SAGA method. The fleet size, which affects the operator cost, is shown in row seven. It may be seen that the percentages of zero transfer trips are higher for all solutions produced by the SAGA method. Except for Mandl's original results, all solutions provide 100% trip coverage with zero or one (one-or-less) transfer involved in each trip. The savings in total user travel time in hours from the SAGA method are given in the eighth row. The average boarding statistics shown in row eleven is defined as the average number of vehicle boarding(s) needed by a transit user to accomplish his/her OD trip. The optimal value of average boarding is 1, meaning that all transit users accomplish their OD trips with one vehicle boarding, i.e. without transfers.

The second example involves optimizing a large-scale transit network based on the service area of the Miami-Dade Transit (MDT), which encompasses a region of about 780 square kilometers with a population of about 2.3 million. MDT is ranked the 16th largest transit agency in the U.S. In 2005, MDT operated 83 transit routes, including a rail rapid transit system of 36.2 route-km (Metrorail), a 7.6-km downtown automated circulation system (Metromover), and 81 bus routes with about 4500 transit stops. In the optimization process, Metrorail and Metromover alignments are fixed. The street network used in this experiment consists of 4,300 street segments and 2,804 street nodes. The total length of the transit system is about 2,092 route km, with some small loops omitted at the ends of some routes or in shopping centers. The longest bus routes are about 51.5 km. The OD matrix was generated from the 1999 validated Miami-Dade travel demand model, which provided the daily number of passenger trips between each pair of traffic analysis zone centroids. These trips were manually distributed to the surrounding street network nodes with considerations given to land use patterns and street network connectivity. The total demand is 161,944 daily transit trips, which are distributed, sparsely and unevenly, between about 120,000 demand (OD) pairs. The network fleet size is 600 transit vehicles, which is about the same as that operated in MDT. The total number of transit lines in the example remains the same as the existing system. Other constraints and data used in this example include:

- bus routes minimum and maximum headways are four and 20 minutes, respectively;
- Metrorail minimum and maximum headways are six and 10 minutes, respectively;

- Metromover minimum and maximum headways are one and 6 minutes, respectively;
- transfer penalty for bus routes is 15 minutes;
- transfer penalty for Metrorail and Metromover are 2.5 minutes;
- minimum and maximum bus route length (in-vehicle time) are 10 minutes and 80 minutes; and
- average transit vehicle operating speed is 33.8 km/h for buses and 49.9 km/h for Metrorail.

All the numerical results were obtained on a PC with a 2.8GHz CPU and 1GB RAM memory. Table 2 presents the results. For comparison, network parameters for the existing network are also included. The first row in the table identifies the objective functions used to generate the corresponding numerical results. For each objective function case, the corresponding column provides the statistics for the TN obtained from the optimization process.

Table 2. Comparison of SAGA Search Results with the Existing Transit Network (Zhao and Zeng 2006)

Network Parameters	Results generated from existing TN	Based on maximization of zero-transfer trips and minimization of total user cost	Based on maximization of zero- and one-transfer trips and minimization of total user cost
0-transfer trips (%)	14.38	33.49	33.11
1-or-less transfer trips (%)	55.17	83.27	87.35
2-or-less transfer trips (%)	65.23	87.29	91.32
Total covered trips (%)	65.69	87.32	91.39
Average user cost (minutes)	46.8	42.1	39.66
Total fleet sizes	600	600	600
Average Boarding with Penalty for Uncovered Trip	2.652	1.960	1.882
CPU Time (hours)	-	1.593	49.28

It may be seen that results have significant improvements over the existing network. The zero-transfer trips (trips not requiring transfers) increased from the existing 14.38% to 33.49%, an improvement of about 133% (column 3). The one-or-less transfer trip (trips requiring at most one transfer) coverage increased from 55.17% to 87.35%, an improvement of about 58%. Assuming most transit riders are only willing to transfer once per trip (Stern 1996), the one-or-less trip coverage shown in the sixth row would be the actual total service coverage of the corresponding TN. The remaining trip demands either require two or more transfers or are not satisfied. The average user vehicle boarding results shown in the table include the penalty for uncovered trips is penalized as four vehicle boardings. The best average per user vehicle boarding shown in Table 3 is 1.822, a 30% improvement over the existing value of 2.652. The results demonstrate that given a network and demand, the solution method is able to improve the network based on user cost and operator cost.

State-of-the-Practice and Challenges

In the real-world of transit planning, either long-range or service planning, system optimization is still rare. Transit networks have evolved over a long period of time of decades or even centuries. When an urban area grows and service expansion becomes necessary, new routes are usually added or old routes extended, with relatively minor changes to the existing system structure. It is not that transit planners do not wish for an optimized network. Even assuming that optimization techniques have advanced to the point that such solutions are feasible, these solutions, as they are churned out by the optimization machinery, are impractical for implementation. The simple answer to the question about the reluctance of the planners is user confusion. Take transit schedules as an example. Typically, schedules are adjusted once or twice a year with perhaps minor changes in the routes. For fear of user confusion, such changes must be advertised long before they come into effect. Even so, some users may still be uninformed and become frustrated with missed rides. If more than a few routes are to change, there will be even more confusion to a larger degree, as the users not only have to adapt to new schedules, but also new routes and transfer locations. When transit service is inferior to private automobiles, such inconveniences may result in a significant decrease in ridership. For this reason, some studies that focus on the design of bus route extensions (Matisziw et al. 2006), schedule coordination (Cevallos and Zhao 2006), feeder route design (Kuan *et al*. 2006, Shrivastav and Dhingra 2001), etc., are more likely to produce immediate results than those that attempt to optimize an entire transit network.

The difficulty with implementing the solutions of new network design is not to deny that there is a significant value in research on large-scale transit network optimization. It will be very desirable for planners to understand how far away their networks are from being optimal and which part of the network can be modified with minimum inconveniences to the users and maximum benefits to both the users and the operators. It is relatively easy to determine how far away a network is from being optimal because the objective functions may be evaluated for both an existing and a new network. It is a little more difficult to determine which route or routes are the best to change to result in the least inconveniences and most benefits. The reason is that this implies that only a limited number of routes are to be changed, but selecting a subset of routes to be optimized is itself an combinatorial problem. For example, for a system with 10 routes, there are $C_{10}^{1} + C_{10}^{2} + C_{10}^{3} + \ldots + C_{10}^{10} = 1{,}023$ possible ways to optimize one through 10 routes at a time. For a system with 50 routes, which is not uncommon, and one to five routes being considered for modification, the number of cases to be investigated will be 2,369,935. Although methods are already available to optimize any selected subset of routes, the required computation will be intensive.

Another challenge is transit ridership estimation. Currently, travel mode share models are still far from being accurate and reliable. Aside from the complexity in the decision process by a user to choose transit as a travel mode, there is also a lack of

sufficient data, especially in areas with low transit ridership, to support the development of accurate transit ridership models.

For transit network optimization, simply knowing the number of riders or transit trips in a given area is not enough. The origin and destination of the transit trips are also needed. Such information is commonly supplied by a transportation planning model, often a four-step model, which are known for limitations in the trip distribution methodology. One possible way to improve the accuracy of OD estimation is through the development of better models, such as joint trip generation-distribution-mode choice models or activity-based models. Another is to find better ways to collect OD data, which is possible with new fare technologies. While directly collected data are useful to estimate an OD matrix for an existing system, the origin and destination information still needs to be estimated for potential trips in either an area that is previously not served or where services are to be expanded.

In most existing studies, the demand has been assumed to be a given, i.e., the demand is unchanged during network optimization. However, it is well understood that changes in the network, which may be the route structure, service frequencies, or schedules, can affect the demand. Not only is the demand a function of the supply of transit service and the level of transit service relative to that of private automobiles, but also does the service area change with the geographic coverage of the network and the level of service. It is not technically difficult to integrate a transit network model with a transit demand model to allow interactions between the two models. One of the main challenges will be the high demand for computing resources. Furthermore, the convergence of such an integrated system also needs to be investigated.

Transit demand changes not only with the transit supply, but also with time during a day. Most transit systems have different schedules or even different route coverages during peak, off-peak, and sometimes evening hours. Therefore, optimization may be performed for each period of service. Although the route structure should not change significantly between peak and off-peak hours, other than possible truncation of some routes during off-peak or evening hours, the frequencies and schedules can both be optimized. Such period modeling requires that the transit network be coded and demand to be estimated for each service period. The development of time-of-day transit networks require good information on travel time, including all components of the travel time. The time-of-day modeling of transit demand and OD matrices makes the transit demand estimation more challenging.

In addition to hourly variation in network operating conditions and transit demand, there are also seasonal changes in both. The decision therefore needs to be made as whether to use the peak season demand or an annual average daily demand as the basis for optimization. To be able to meet the peak season demand will mean underutilization of the services during off-peak season. Conversely, designing a network for a demand less than the peak season demand will mean a strained system during the peak season, and possibly negative impact on ridership.

Finally, at a large scale of things, there is also the interaction between transit supply and land use. As land use and demographics change, the transit demand also changes. This implies that a transit network will never be an optimal one, although it may be closer to an optimum at one time point than at another. This also raises a question – assuming that future land use and transit demand can be predicted with reasonable accuracy, is it possible to optimize a future network and evolve the existing network into that future one? This seems to imply that do to so simulation must be employed, in which modeling of land use and demographics will alternate with the optimization of transit network.

Conclusions

In this paper, selected literature on transit network design is briefly reviewed. A formulation of the transit network optimization is given. Some solution strategies are described and one set of solutions obtained from an implementation of the metaheuristic search method that combines GA, SA, Tabu, and bi-section search for a relatively large transit network are presented. The state-of-the-practice of transit network optimization is also discussed. Many challenges remain, some computational, some technical, and others practical. Although much progress has been made in solving the problem computationally, the solutions will generate real benefits only when many of the practical and technical issues are addressed.

References

Agrawal, J. and T. V. Mathew (2004). "Transit Route Network Design Using Parallel Genetic Algorithm." *Journal of Computing in Civil Engineering,* ASCE, 18, 248-256.

Baaj, M. H and H. S. Mahmassani (1991). "An AI-based Approach for Transit Route System Planning and Design." *Journal of Advanced Transportation,* 25(2), 187-210.

Bertsekas, D. P. (1998). *Network Optimization: Continuous and Discrete Models*. Athena Scientific. Belmont, MA.

Bielli, M., M. Caramia, and P. Carotento (2002). "Genetic Algorithms in Bus Network Optimization." *Transportation Research Part C*, 10(1), 19-34.

Cevallos, F. and F. Zhao (2006). "Minimizing Transfer Times in a Public Transit Network with a Genetic Algorithm," *Transportation Research Record* 1971: 74-79.

Charkroborty, P. and T. Dwivedi (2002). "Optimal Route Network Design for Transit Systems Using Genetic Algorithms." *Engineering Optimization*, 34(1), 83-100.

Dorigo, M. (1992). *Optimization, Learning and Natural Algorithms* (in Italian), Ph.D. thesis, DEI, Politecnico di Milano, Italy. pp. 140.

Fan, W. and R. B. Machemehl (2004). *Optimal Transit Route Network Design Problem: Algorithms, Implementations, and Numerical Results*. Report No. SWUTC/04/167244-1, Center for Transportation Research, University of Texas at Austin (http://swutc.tamu.edu/Reports/ 167244-1.pdf).

Garey, M. R. and D. S. Johnson. (1979). *Computers and Intractability. A Guide to the Theory of NP-Completeness*. Freeman, N. Y.

Gao, Z. H. Sun, L. L. and Shan (2002). "A Continuous Equilibrium Network Design Model
and Algorithm for Transit Systems." *Transportation Research Part B*, 38, 235–250.

Goldburg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley.

Kirkpatrick, S., C. D. Gelatt, Jr., and M. P. Vecchi (1983). "Optimization by Simulated Annealing." *SCIENCE* (220)4598, 671-680.

Kuan, S. N., H. L. Ong, K. M. Ng (2006). "Solving the Feeder Bus Network Design Problem by Genetic Algorithms and Ant Colony Optimization." *Source Advances in Engineering Software Archive*, Elsevier Science Ltd., Oxford, UK, 37(6), 351 – 359.

Mandl, C. E. (1979). "Evaluation and Optimization of Urban Public Transportation Networks." *Third European Congress on Operational Research*, Amsterdam, Netherlands.

Matisziw, T. C., A. T. Murray, and C. Kim (2006). "Strategic Route Extension in Transit Networks." *European Journal of Operational Research,* 171, 661-673.

Newell (1979).

Shih, M.-C. and H. S. Mahmassani (1994). *A Design Methodology for Bus Transit Networks with Coordinated Operations*. SWUTC/94/60016-1, Center for Transportation, the University of Texas at Austin, Austin, Texas.

Shrivastav P, S. L. Dhingra (2001). "Development of Feeder Routes for Suburban Railway Stations Using Heuristic Approach." *Journal of Transport Engineering*, 127(4), 334-41.

Stern, R. (1996). *Passenger Transfer System Review*. Synthesis of Transit Practice 19, Transportation Research Board, National Research Council, Washington, DC.

Van Nes, R., R. Hamerslag, and L. H. Immers (1988). "The Design of Public Transport Network." *Transportation Research Record*, 1202, 74-83.

Wu, C. and A. Murray (2005). "Optimizing Public Transit Quality and Access: the Multiple-Route Maximal Covering/Shortest Path Problem." *Environment and Planning B: Planning and Design,* 32, 163-178.

Yang, Z., B. Yu, C. Cheng (2007). "A Parallel Ant Colony Algorithm for Bus Network Optimization." *Computer–Aided Civil and Infrastructure Engineering*, 22(1), 44-55(12).

Zhao, F. (2006). "Large-Scale Transit Network Optimization by Minimizing Transfers and User Cost." *Journal of Public Transportation*, 9(2), 107-129.

Zhao, F. and A. Gan, A. (2003). *Optimization of Transit Network to Minimize Transfers: Methodologies for Route Network Optimization*. Final Report. Dept. of Civil and Environmental Engineering, Florida International Univ., Miami, FL. http://www.dot.state.fl.us/research-center/Completed_Proj/Summary_PTO/FDOT_BD015_02rpt.pdf

Zhao, F. (2006). "Large-Scale Transit Network Optimization by Minimizing Transfers and User Cost," *Journal of Public Transportation*, 9(2), 107-129.

Zhao, F. and X. Zeng (2006). "Optimization of Layout and Headway for Large-Scale Transit Networks with a Combined Genetic Algorithm and Simulated

Annealing Method," *Journal of Engineering Optimization*, Taylor and Francis Publisher, 38(6), 2006, 701-722.

Zhao, F. and X. Zeng (2007). "Simultaneous Optimization of User and Operator Costs for Large-Scale Transit Network with a Simulated Annealing, Tabu, Greedy, and Bi-Section Search Method," *ASCE Journal of Transportation Engineering*, 133(4), 240-251.

PRICE NEGOTIATION MODEL IN E-COMMERCE

Jinkai WANG
School of Transportation Engineering, Tongji University Shanghai 200092, China
E-mail: wjkmail@gmail.com

Abstract: With the rapid development of electronic commerce, more and more people have begun to go shopping on the Internet. Price negotiation has become increasingly important for two reasons. From the consumer's view, price negotiation provides an opportunity to debate the price; from the vendor's view, the ability to negotiate the price allows for flexibility in pricing that a rigidly fixed price cannot offer. An obstacle that hinders the development of electronic commerce is that price negotiation on the Internet is not like shopping in store. In this paper, the author wants to design an agent-based price negotiation system for electronic commerce. It would allow the seller to negotiate with the buyers on the Internet. Firstly we put forward a framework of agent-based price negotiation, then we propose a price negotiation model based on equilibrium theory of game theory. Finally, the paper describes the behavior definition of agent and explains the rules for communication between agents.
Key words: *agent-based, price negotiation, game theory*

1. Introduction

The United Nations Organization for Economic Cooperation and Development (OECD) has given the following definition of electronic commerce: "An Electronic Commerce is the sale or purchase of goods or services, whether between businesses, households, individuals, governments, and other public or private organizations, conducted over computer mediated". For the transaction both sides, and the substance content of the electronic commerce (EC) could be divided into following groups:

- Business to Business (B2B)
- Business to Customer (B2C)
- Business to Government (B2G)
- Customer to Customer (C2C)

At present, more and more net users are accepting shopping on the Internet. According to the CNNIC's newest survey in January 2006, it shows that the total number of net users has already achieved 111 million in our country. Among those users the proportion of people who have bought something online has reached 40.7%, and the crowd who frequently uses the Internet to shop is close to 6.351 million people. Most electronic commerce practitioners feel that the future prospect is bright for the reason that the proportion purchasing commodities or services on the Internet during recent one year has reached 37.8%. The proportion of customers who are

planning to shop in this way will be close to 58%, that is to say the number will pass 60 million.

The old shopping custom has changed because of the fast development of EC, and this way is favored by more and more people because of its convenience. In the past people hoped to bargain with the seller in order to buy products at the lowest price. Therefore, the same question how to bargain in EC on the Internet has been put forward.

From the view of vendor, the necessity of negotiation lies in in the following:

1) Increase the vendor's income in the short term. It is essential to negotiate with clients in EC because price negotiation can make clients feel psychological satisfaction in one transaction.

2) Guarantee the vendor's income in the long term. Price negotiation is a method to attract clients to return after a successful transaction.

According to the consumer's motivation, they would like to bargain when they buy goods. It is suitable for e-commerce trade. From the view of clients, it is necessary for the following reasons:

1) Economical reason: When most people want to buy an item, they always hope to buy the article at the lowest price in order to satisfy his or her needs.

2) Psychological reason: During the business activity, buyers always think the business is a mutual process that requires the seller and buyer to negotiate and at the same time it can satisfy customers' psychological expectations.

3) The feeling of time value: Most people think that time value is not so sensitive as the article price, and people would like to spend a little more time to bargain in order to reduce the article price more.

4) At present our country it has the environment and tradition of bargaining. While going shopping in a real shop people have been used to bargaining. So they hope to keep such a habit while shopping in the e-commerce trade.

But it is very troublesome to negotiate a price on the Internet, especially when a seller faces a lot of customers at the same time. If he delays serving for one of them, it could be possible to make the customer very dissatisfied or even lose the customer forever. So many e-commerce suppliers have set "the fixed price" with no price negotiation. It is also an obstacle of e-commerce development at present.

A price negotiation system based on Agent can meet the requirements of the development of e-commerce.

The aim of this paper is to show how to design an agent-based price negotiation system for electronic commerce in order to negotiate prices with buyers replacing vendor. First, the paper puts forward a framework of an agent-based price negotiation system in Section 2. Section 3 proposes a price negotiation model based on the equilibrium theory of game theory. Section 4 describes the behavior definition of agents and explains communication between agents.

2. Agent Definitions and Characteristics

An Agent is anything that can be viewed as perceiving its environment through

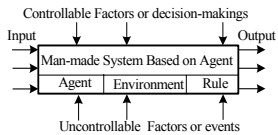


Figure 1. A man-made system based on agent

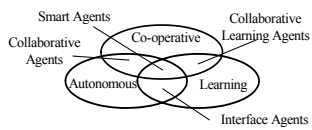


Figure 2. Nwana's classification for agent

sensors and acting upon that environment through effects. Figure 1 represents a man-made system based on agent.

Nwana's classification for an agent may be neatly shown as a Venn diagram (seen as Figure 2). This diagram provides a framework into which we can place almost all software agents currently in the literature[5].

Owing to the complexity of the real world, people have considered a lot of methods to get a better solution in order to approach the result of real world to solve the complicated questions. Artificial intelligence technology based on agent can solve many complicated problems. This technology has been approved by more and more people and has been applied in the fields of economic systems, natural phenomenon, artificial life etc. It is mainly different from other artificial software as the following characteristics show:

1. Autonomous: An agent is an independent entity capable of reasoning on its own. It is able to exercise a non-trivial degree of control over its own actions on behalf of its owner, without requiring explicit permission for every action.

2. Reactive: An agent can sense its dynamic environment (which may be the sensor inputs of the physical world, explicit user inputs, information gathered from another agent of from sources such as the Internet) and respond to those changes with suitable reactions.

3. Proactive: An agent is goal-oriented and takes initiatives to fulfill its goals.

4. Persistent: The notion of agency involves a sense of temporal continuity; an agent is usually continuously running.

5. Trustworthy: An agent will only do what its owner expects it to do.

6. Personalized: An agent can either learn or be explicitly taught what to do for each individual or group of users.

7. Social Behavior: An agent can interact with the user or other agents in order to best accomplish its goals.

3. Agent-Based Price Negotiation Model of E-Commerce

3.1 Transaction Process of E-commerce

Figure 3 describes the transaction process of e-commerce. In Figure 3, the client and vendor achieve agreement consistently through price negotiation, but the purchaser does not pay money to vendor directly, but to an intermediate institution. Only when buyer receives his goods, will the intermediate institution transfer the money to vendor.

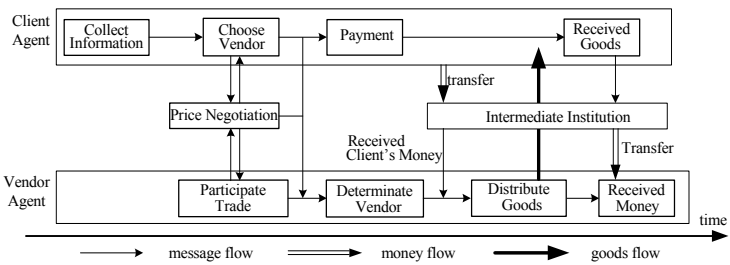


Figure 3. The transaction course of e-commerce

Negotiation is a process of reaching an agreement on the terms (such as price and quantity) of a transaction for two or more parties. However, negotiation is not just a matter of finding an acceptable deal, but is an attempt to maximize all parties' payoffs (Figure 4).

In the EC price negotiation system the conditions needed in a transaction are as follows:

1. The bargaining participant's lowest sell price is smaller or equal to bargaining demander's buy price.

2. The bargaining participant's maximal buy price is bigger or equal to bargaining demander's sell price.

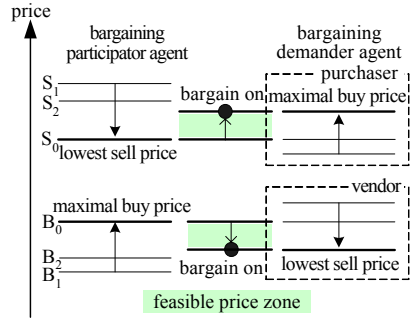


Figure 4. Transaction deal process

3.2 Agent-based Price Negotiation System

Using these characteristics of agents, we apply it to the price negotiation system, and put forward to an agent-based price negotiation system. The system architecture described in Figure 5 consists of a Cognitive Agent (CA), a Study Agent (SA), a Processing Agent (PA), a Reactive Agent (RA), and Knowledge Library, Data Library, and Model Library.

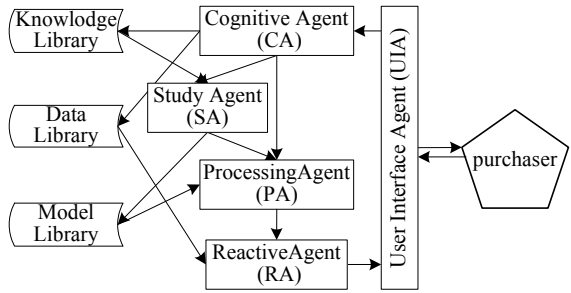


Figure 5. Agent-based price negotiation system architecture

1. User Interface Agent (UIA): UIA is used for obtaining the customer's information. It is different from the interface of traditional human-computer interaction for the reason that UIA does not need users to offer the data for computer like the traditional human-computer interface needed. It only needs users to interact with the agent according to the goal, and agent can treat different users individually. The user interface could also be able to finish some jobs that the users proposed under special conditions.

2. Cognitive Agent (CA): CA belongs to one kind of reactive agent by function. It can get the information on outside conditions through the interface agent to provide a study agent to study and a processing agent to process. CA would be able to store the information obtained from the outside into the database. And CA itself does not have the "knowledge-reasoning" mechanism. Its intelligence is shown as perceived ability and action ability, and its merit is in a quick response to an exterior stimulus..

The basic functions of CA are to recognize and identify the input condition from the outside. When several external neurons are in a certain condition CA would be able to appear in an "excited" status; On the other hand when several external neurons are in another condition CA would be able to appear in "prohibitive" status.

3. Study Agent (SA): SA belongs to a duty Agent. Its responsibility is to decompose, assign and manage tasks, and has the right of appointing other agents to finish a certain task together. It can compare the knowledge in the knowledge library and database with those obtained from the outside. From the attribute it could also be a deliberative agent. The knowledge library includes two kinds of knowledge: One kind is rules, and another kind is a knowledge block.

4. Processing Agent (PA): PA belongs to the duty agent and deliberative agents like SA. It can combine the information transferred by SA and CA using a processing model in the model library to process and judge.

5. Reactive Agent (RA): RA sends the knowledge from PA to client with a way in which client can understand. RA is made up of a perceiving device, an effecting device, a communication mechanism and an information library.

4. Implement

4.1 Price Negotiation Algorithm

The process price negotiation in electronic commerce can be described by multi-phase game theory.

Let $a_1^t(t=0,1,2,...T)$ be a price that a bargaining participant offered at time t (can also be interpreted as the number of times). Let $a_2^t(t=0,1,2,...T)$ be a price that a buyer answered at time t (can also be interpreted as the number of times). The sequence of actions $h_i^t=(a_i^0,...,a_i^{t-1}),(t=0,1,2,...T,i=1,2)$ is the vector of historical price offered by both sides of bargaining before time t.

$\sigma_i(a_i^t \mid h_i^t,\theta_i)$ is probability when both sides of bargaining take action a_i^t under the condition of h_i^t and bidding strategy θ_i.

In order to extend the concept of a sub-game-perfect to these games, we need to construct Bayesian Nash equilibrium. It not only adapts to the whole game, but also

adapts to the following game after the historical sequence of actions h^t. The conditional probability of participant i is $\mu_i(\theta_{-i} \mid \theta_i, h^t)$ when the strategy of his opponent is θ_{-i}. For all participants, we assume that time t, historical sequence of actions h^t, and strategy θ_i have their definition.

Assume that the decision influence factors of participant *i* are as follows:

(i) $\mu_i(\theta_{-i} \mid \theta_i, h^t) = \prod_{j \neq i} \mu_i(\theta_j \mid h^t)$

$(t = 1, 2, ..., T;\ i = 1, 2;\ j = 1, 2)$ (1)

(ii) $\mu_i\left(\theta_j \mid (h^t, a^t)\right) = \dfrac{\mu_i(\theta_j \mid h^t)\sigma_j(a_j^t \mid h^t, \theta_j)}{\sum_{\theta_j^\circ} \mu_i(\theta_j^\circ \mid h^t)\sigma_j(a_j^t \mid h^t, \theta_j^\circ)}$

$(t = 1, 2, ..., T;\ i = 1, 2;\ j = 1, 2)$ (2)

(iii) $\mu_i = (\theta_j \mid (h^t, a^t)) = \mu_i(\theta_j \mid (h^t, a^t))$,

$if\ a_j^t = a_j^t\ (t = 1, 2, ..., T;\ i{=}1,2; j{=}1,2)$ (3)

(iv) $\mu(\theta_{-i} \mid h^t)\mu(\theta_i \mid h^t) = \mu(\theta \mid h^t)$

$(t = 1, 2, ..., T;\ i{=}1,2)$ (4)

For every participant *i*, bidding strategy θ, other strategy σ_i', and historical sequence of actions h^t, there is:

$$u_i(\sigma \mid h^t, \theta_i, \mu(\cdot \mid h^t)) \geq u_i((\sigma_i', \sigma_{-i}) \mid h^t, \theta_i, \mu(\cdot \mid h^t))$$

$(t = 1, 2, ..., T;\ i{=}1,2)$ (5)

Then the expected income of game can be written:

$$u_i(\sigma \mid h^t, \theta_i, q) = u_i(h^t) \cdot u_i(\sigma \mid h^t) + u_i(\theta_i) \cdot u_i(\sigma \mid \theta_i) + q \cdot u_i(\sigma \mid q) \quad (t = 1, 2, ..., T;\ i{=}1,2) \quad (6)$$

where： σ——the probability of a^t bidding;

q——the probability distribution of a^t bidding.

It needs to further estimate the probability distribution of bidding strategy θ in the course transaction.

The transaction succeeds the condition is described by table 1.

Table 1 The transaction succeeds the condition

Algorithm 1

S_0, B_0, a_1^0, a_2^0, tg=1;

repeat

$$u_1(\sigma \mid h^{tg}, \theta_1, q) = u_1(h^{tg}) \cdot u_1(\sigma \mid h^{tg}) + u_1(\theta_1) \cdot u_1(\sigma \mid \theta_1) + q \cdot u_1(\sigma \mid q);$$

$$u_2(\sigma \mid h^{tg}, \theta_2, q) = u_2(h^{tg}) \cdot u_2(\sigma \mid h^{tg}) + u_2(\theta_2) \cdot u_2(\sigma \mid \theta_2) + q \cdot u_2(\sigma \mid q);$$

tg=tg+1;

until $\left(\left\|U_1(\sigma \mid h^{tg}, \theta_1, q) - U_2(\sigma \mid h^{tg}, \theta_1, q)\right\| < \varepsilon \text{ or } B_0 \geq S_0\right)$

4.2 The Form Definition of Agent

The form definition of agent can be described as follows:

```
public class Agent{
Agent  AgentName;
Public static void main()
{ privet data1, data2, …
Knowledge-base rule1, rule2, …
Process<process-name1>
  On<event1> Do <action1> at Priority
<Prior1>
Process<process-name2>
  On<event1> Do <action2> at Priority
<Prior2>
  …
Action<action1>
Action<action2>
}
}
```

4.3 Communication between the Agents

The communication between the agents means that agents send and receive message mutually. The message is composed of a message name and a set of parameters. Its syntax can be show by **<Message-Name>[(p0, p1, …, pn)]**.

In the communication, **<massage-name>** is message name. It describes the news type and the implication. Sequence **(p0, p1, …, pn)** is message parameters. The sentence if sending message is follows:

Send<Receiver-Agent><Message-Name>(p0, p1, …, pn) [Wait<Message-Name> (q0, q1, …, qm)].

where, **<Receiver-Agent>** is message receiver agent; **<Message-Name>**(p0, p1, …, pn)are message name and message parameters.

[Wait<Message-Name> (q0, q1, …, qm)]is optional parameter, it indicates the message to wait to answer.

5. CONCLUSION

Price automated negotiation has become increasingly important since the advent of electronic commerce. In this paper, we proposed an agent-based price negotiation system for electronic business. We discussed a framework of agent-based price negotiation system, and proposed a price negotiation model based on equilibrium theory of game theory. Furthermore, the paper interpreted the structure of agent and explained how the communication is done between agents. We used the game theory of Nash equilibrium. The paper only demonstrates the implement of model and accounts for some influential parameters. There is further work required to verify these parameters.

References

[1] Samuel P.M. Choi, Jiming, Liu, Sheung-Ping Chan, "A genetic agent-based negotiation system" *Computer Networks*, 37(2001): pp195-204.

[2] Yoichi Sawasaki, Yutaka Kimura, Kensuke Tanaka, "A TWO-PERSON ZERO-SUM WITH FRACTIONAL LOSS FUNCTION", *Journal of the Operations Research Society of Japan*, Vol. 43, No. 1, March 2000: pp209-218.

[3] Ralf Neubert, Otmar Gorlitz, Töbias Teich, "Automated negotiations of supply contracts for flexible production networks", *Int. J. Production Economics*, 89(2004): pp175-187.

[4] Gerald J. Tesauro, Jeffrey O. Kephart, "Foresight pricing algorithms in agent economies", *Decision Support Systems*, 28(2000): pp49-60.

[5] C. J. Anumba, Z. Ren, A. Thorpe, O. O. Ugwu, L. Newnham, "Negotiation within a multi-agent system for the collaborative design of light industrial buildings", *Advance in Engineering Software*, 34(2003): pp389-401

[6] Partrick Desharnais, "Agent Assisted Price Negotiation For Electronic Commerce", *Master Thesis*, Concordia University, Montreal, Quebec, Canada, April 2000.

[7] Yang Junqin, "Distribution System Planning for B2C Electronic Commerce——Considering the relationship of the system resource configuration and service level", *PhD Thesis* , School of Transportation Engineering, Tongji University, Shanghai, China, July 2005.

[8] Ammar Attoui, "Real-Time and Multi-Agent Systems", Springer-Verlag London Berlin Heidelberg, United Kingdom, 2000.

[9] Drew Fudenberg, Jean Tirole, Translator: Huang Tao etc. "Game Theory", Renmin University of China Press, Beijing, China, October 2002.

[10] He Yanxiang, Chen Xinmeng, "Multi-agent system design and application", Wuhan University Press, Wuhan,China, June 2001.

MODELING OF CARGO VEHICLES CHOICE IN VARIABLE PRICING OF FREEWAYS

Qiaojun Xiang[1]， Yingwei Ren[2] ， Jian Lu[3] and Shourong Wu[4]

[1]Transportation College, Southeast University, 2 Si-Pai Lou, Nanjing, Jiangsu 210096, China, 86-25-83792560 (o), 86-25-83795352 (fax), email: xqj@seu.edu.cn
[2]College of Civil Engineering and Architecture, Shandong University of Science and Technology, Qingdao, Shandong, 266510, China, 86-532-86057595(o), email: renyingweicy@sohu.com
[3]Transportation College, Southeast University, 2 Si-Pai Lou, Nanjing, Jiangsu 210096, China, 86-25-83790776 (o), 86-25-83795352 (fax), email: jianjohnlu@sina.com
[4]College of Civil Engineering and Architecture, Shandong University of Science and Technology, Qingdao, Shandong, 266510, China, 86-532-86057595(o), email: wusrong@sohu.com

Abstract: Currently, freeway traffic demand during rush hours is very high and freeway traffic demand during non rush hours is very low. With variable toll pricing capabilities for cargo vehicles, traffic congestion on freeways could be reduced. Since cargo vehicle drivers have flexibility in selecting their departure time, variable toll pricing for cargo vehicles may have the feasibility in balancing cargo vehicle departure time distribution and improving traffic efficiency. In the paper, field questionnaire surveys were performed to obtain cargo vehicle drivers' behavior data regarding their departure choices if different variable pricing schemes would be implemented. Based on the survey data, variable pricing models with logit model formats were developed.
Key Words: *Variable Pricing, Cargo Vehicles, Logit Model, Driver Behavior, Questionnaire Survey*

1 Introduction

Variable pricing is one of traffic management methods. The basic principle of variable pricing is to use different pricing schemes for different time segments. Practically, during rush hours, a higher toll price would be useful to encourage some traffic to change their departure time from rush hours to non-rush hours so that traffic distribution over time could be somewhat balanced.

The principle of static congestion pricing was proposed by Pigou (1920) and Knight (1924). After that, many research studies have proved that variable pricing has the benefit to reduce roadway traffic congestion in certain degree with limited effect to travelers' benefits. In 1975, Singapore began to apply variable pricing (also called congestion pricing) models to control traffic demand in the central district area in Singapore. Past practice shows that the applications of variable pricing in Singapore

have been successful. In last three decades, variable pricing models have been applied in Canada, France, Japan, and other Europe countries(Lam,T. C., Small, K.A., 2001). In American, Sullivan and Jia(2000) used logit models to evaluate the impacts of variable pricing on drivers' route choice behavior, transportation modal choice behavior, and departure time choice behavior. Kawamura analyzed and evaluated the benefits of variable pricing to commercial vehicles. Kazimi et(2000) also analyzed the travelers' behavior change with the implementation of variable pricing on I-15 Interstate highway. In 2002, Burris developed discrete choice models to analyze the willingness of drivers with different social-economic characteristics to select the toll facility with variable pricing schemes. Holguin-Veras (2006)concluded toll collection polices in USA for various type of vehicles and the impacts of variable pricing in urban areas on cargo vehicle drivers' travel behavior.

In recent years, as the economic development in China, transportation demand has significantly increased and there are some real problems existing on China high-speed freeways. Some typical problems are:

1). Statistically, more than 30% traffic traveling on national freeways are cargo vehicles. The speed difference between cargo vehicles and passenger cars are more than 20 kilometers per hour in most cases. Such a large speed difference could result in the reduction of roadway capacity and, more seriously, may result in severe traffic conflicts and crashes.

2). Since economic developments in different regions are not the same, traffic demand distributions in different regions are significantly different. The unbalanced traffic distribution may provide some potential for drivers to select the best routes if necessary traffic information is provided.

3). There are significant hourly-traffic variations over time. Figure 1(Jian Lu, Qiaojun Xiang, Yingwei Ren,2006) presents the general traffic distribution and cargo traffic distribution during a 24-hour time period on the Hu-Ning Freeway (between Shanghai and Jiang Su Province). It can be seen that the general traffic has significant peak traffic volumes during morning and evening rush hours. If variable pricing schemes are implemented on this freeway, some traffic during rush hours could shift to non-rush hours.

4). Cargo traffic occupies a large part of the whole traffic. Figure 2 (Jian Lu, Qiaojun Xiang, Yingwei Ren,2006) shows the composition of traffic on freeways in Jiangsu Province in 2004. Slow-moving large cargo traffic that moves in a speed between 50 to 80 kilometers per hour may cause some problems to traffic capacity and safety. If some of the rush-hour cargo traffic could shift to non-rush hours to travel, freeway traffic capacity could significantly increase and traffic performance could improve.

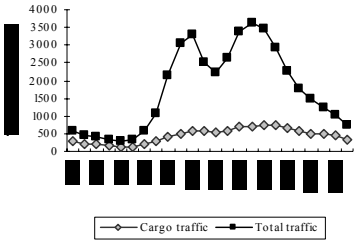


Figure 1 Time distribution of traffic volume on Jiang Su freeways

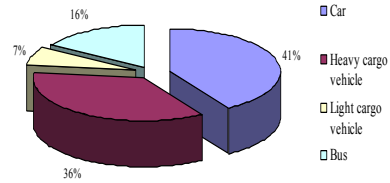


Figure 2 Composition of traffic on Jiang Su freeways

In recent years, China transportation agencies have emphasized more on traffic management to improve highway traffic efficiency and safety. Variable pricing is a good method of management.

2 Proposed Variable Pricing Schemes for Cargo Vehicles

The design for variable pricing schemes includes division of time segments, toll rate for each time segment, and definition of vehicle types. This research only focused on the variable pricing for toll collection of cargo vehicles. Thus, only cargo vehicle type was considered.

2.1 Division of time segments

As shown in Figure 1, traffic demands are not evenly distributed over the 24-hour time periods. Traffic demands during rush hours are much higher than that during non-rush hours. The division of time segments could be based on the hourly volume distribution of traffic over time. One day time is divided into two segments: Peak Traffic Segment and Non-Peak Segment.

Selection of the time segment division method for a particular freeway mainly depends on hourly traffic distribution characteristics of the particular freeway. The best division of time segments could be obtained through variable pricing model analysis if drivers' behavior models could be obtained.

2.2 Toll rates at different time segments

A basic rate is considered the rate before variable pricing is installed and the basic rate is a reference toll rate considered in the paper. If variable pricing is to be installed, the following options could be considered: (1) A higher toll rate is used during rush hours and the basic rate is kept for non-rush hours; (2) The basic rate is kept for rush hours and a lower rate is used during non-rush hours; (3) A higher rate is used during rush hours and a lower rate is used during non-rush hours; However, during average traffic demand time periods, the basic rate is used. The so called higher rate and lower rate can be determined based on variable pricing models which could be developed through discrete choice modeling with the data collected from field surveys. These models would generate the higher or lower toll rate that can be accepted by cargo vehicle drivers so that optimized benefits can be obtained.

2.3 Possible impacts of variable pricing on cargo vehicle drivers' behavior

Based on the current situation and the market characteristics of highway cargo transportation, the toll collection fee have been a large part of the cargo transportation costs. Changes in toll fee could be sensitive to cargo owners' decision in travel plans, such as departure time, route choice, transportation modal selection, destination, etc. However, the changes in toll fee would only affect cargo vehicle drivers in terms of departure time and route choice. In fact, in addition to toll fee, many other factors will affect cargo vehicle drivers' choice in terms of travel plans. These factors include: travel purpose, type of cargo, driver's social-economic characteristics, reimbursement method for toll fee, time urgency of cargo transportation, etc. Thus, to develop models for cargo vehicle drivers' choices in departure time and route choice, some of the factors mentioned here would be considered. Field questionnaire surveys to cargo vehicle drivers would be a good way to collect these data for modeling purposes.

The variable pricing models developed in the research were for the departure time choice only. In the region area that the research project was based on does not have adequate alternative routes for cargo vehicle drivers to select in order to avoid higher toll rates. Thus, the route choice behavior was not considered in the modeling process.

3 Modeling Specifications

After variable pricing is installed, cargo vehicle drivers will determine whether they will select the time segments with low toll fee to departure. The decision is based on the toll fee, travel purpose, type of cargo, driver's social-economic characteristics, reimbursement method for toll fee, time urgency of cargo transportation, etc, and a decision regarding departure time is to made so that cargo vehicle drivers' travel efficiency is maximized. The following describes specifications for the modeling of cargo drivers' behavior in terms of departure time selection if the variable pricing scheme is installed.

Since there are only two selections for cargo vehicle drivers: travel during the peak hours or to travel during the non-peak hours, binary discrete choice models could be attempted. In this research, the logit model format was used. The model format is as follows:

$$P_{1n} = \frac{\exp(V_{1n})}{[\exp(V_{1n}) + \exp(V_{2n})]} \quad (1)$$

$$P_{2n} = \frac{\exp(V_{2n})}{[\exp(V_{1n}) + \exp(V_{2n})]} \quad (2)$$

where:

n- number of surveyors,

P_{1n} - probability for the n^{th} driver in peak period to change his/her travel time,

P_{2n} - probability for the n^{th} driver in peak period not to change his/her travel time,

V_{1n} – utility function for the selection of changing travel time, and

V_{2n} - utility function for the selection of not changing travel time.

Mathematically, the relationship between P_{1n} and P_{2n} is: $P_{2n} = 1 - P_{1n}$. The utility functions can have the following formats:

$$V_{1n} = a_1 + a_2 P + \sum_{j=1}^{m} a_{(2+j)} X_{nj} \quad (3)$$

$$V_{2n} = b_2 P + \sum_{j=1}^{m} b_{(2+j)} X_{nj} \quad (4)$$

where:

m - number of driver's characteristics to be considered in the model;

a_i (i = 1, 2, …, m) – model parameters to be estimated for V_{1n},

b_i (i = 2, 3, …, m) - model parameters to be estimated for V_{2n},

P – toll rate, and

X_{nj} (j = 1, 2, …, m) – characteristics of the n^{th} driver and his/her vehicle.

In the paper, two types of variable pricing schemes were attempted: (1) The basic toll rate is used during rush hours and lower toll rate is used during non-rush hours, and (2) As compared to the basic toll rate, higher toll rate is used during rush hours and the basic toll rate is used during non-rush hours.

4 Field Question Surveys and Data Collection

Generally, there are two types of field questionnaire surveys: Revealed Preference (RP) Survey and Stated Preference (SP) Survey. The RP survey is to perform travel behavior questionnaire survey to the travelers who have made the choices, and the SP survey is to perform the same survey to the travelers who have not made the choices. Both RP and SP surveys seek to understand what choices that travelers will make.

In this paper, the SP survey method was used. Field questionnaire surveys were performed at freeway toll stations, service areas, and cargo vehicle parking fields. Cargo drivers were randomly selected and interviewed, and the survey results were entered into a survey form by the surveyors.

4.1 Design of the survey form

The quality and the usefulness of the survey form will have the direct impacts on survey results and model accuracy. To determine the survey form format and elements to be surveyed, focus group meetings were organized and expert surveys were carried out. With the considerations of current highway cargo transportation characteristics in China and the results from these focus group meeting and expert surveys, the following characteristics directly related to cargo vehicles and drivers were selected to be included in the field questionnaire survey: (1) type of cargo to be transported,(2) method to pay toll fee, (3) owner type of the cargo vehicle (the cargo vehicle belongs to the driver or the driver is hired by the cargo vehicle owner.), (4) weight of the cargo vehicle, (5) driver's income, (6) travel distance, (7) driver's age, and (8) the education degree of the driver. The reason for using the variable "type of cargo to be transported" is stated as follows: Since different cargo types have different values, different cargo types would have different travel time requirements, meaning the owners of different cargo types may have different departure flexibilities. If a cargo driver has sufficient departure flexibility, the driver could adjust his/her departure time to obtain economic benefits from cargo toll variable pricing. In the research, cargo type is classified as emergency cargo and general cargo. Emergency cargos have limited flexibility in selecting departure time. However, general cargos may have more flexibility in selecting departure time. Thus, drivers of trucks with general cargo types were the main body for field surveys. The survey was in the format of field interview, which would simplify the tasks for the completion of the survey.

4.2 Field survey and data collection

The main purpose of field surveys performed was to model the impacts of cargo vehicle drivers' characteristics on their departure time selections with given variable pricing schemes. Field questionnaire surveys were performed on the Jing-Hu Freeway (between Beijing and Shanghai). More specifically, the following locations were selected for the surveys: the service areas, toll collection stations, and cargo vehicle parting fields. The cargo vehicles towards Shanghai direction originated mostly from these provinces (cities) like Shandong, Tianjin, Hebei, Beijing, Liaoning, and Neimeng. Their destinations were Jiangsu, Zhejiang, or Shanghai. The cargo vehicles towards Beijing direction originated mostly from Jiangsu, Zhejiang, and Shanghai, and their destinations were Beijing, Shandong, Tianjin, Hebei, or Liaoning. In the survey, the travels covering two or more provinces were considered long-distance travels, and the travels within one province were considered short-distance travels.

Toll pricing in China is based on the three factors: vehicle type, cargo weight, and travel distance. In the field survey, cargo drivers were asked to make their departure time with several given characteristics. In order to understand their departure time choices, two toll variable pricing methods were proposed: (1) During rush hours, as compared current pricing base, toll price would increase by 5% to 50% with the incremental interval of 5% and during non-rush hours toll price would be kept unchanged; (2) During non-rush hours, toll price would decrease by 5% to 50% with the incremental interval of 5% and during rush hours toll price would be kept unchanged. It was found from the field survey results that for the first method more and more cargo drivers would select to shift their departure time from rush hours to non-rush hours as toll price would increase by 10% or more, and for the second method more and more cargo drivers would select to shift their departure time from rush hours to non-rush hours as toll price would decrease by 20% or more. This clearly indicates that variable pricing could be used to adjust cargo traffic demand on freeways.

During the field surveys, 712 cargo vehicle drivers were interviewed and 680 interviewed were considered valid. It was found from the surveys that more than 95% cargo vehicle travels were long-distance travels. Among these cargo vehicles surveyed, about 13% of the cargo vehicles had the load capacity of 10 tons or less, about 72% of the cargo vehicles had the capacity between 10 to 40 tons, and about 15% of the cargo vehicles had the capacity of 40 tons or more. In the surveys, cargo vehicle drivers were asked to rank the factors that could affect their travel plans. The most important factors considered by cargo vehicle drivers were travel cost and travel time. About 75% of interviewed drivers selected travel cost as the most important factor; about 15% interviewed drivers selected travel time as the most important factor, and about 10% interviewed drivers considered travel cost and travel time as equally important factors that affected their travel plan selections. Generally, the drivers who considered travel cost the most important factor would change their departure time if variable pricing is installed, meaning they would departure during the time segments with lower toll rates.

The factor of cargo types is also an important factor.The survey results indicated that most cargo types could be divided into two categories: general cargo and emergency cargo. Emergency cargo types include vegetables, fresh foods, emergency rescue materials, and other cargos with specified delivering time. According to survey

results, about 24% of cargo vehicles had emergency cargos and about 76% of cargo vehicles had general cargos only. It was found from the surveys that if the difference in fee rates for rush hours and non-rush hours was not significant, the probability for drivers of cargo vehicles with emergent cargos to change their departure time was not high. However, if the difference was significant, most drivers expressed that they would adjust their departure time to save some travel costs.

Table 1 lists percent of cargo drivers who would select to shift their departure time from rush hours to non-rush hours if cargo toll variable pricing would be applied. About 166 drivers of cargo vehicles with emergency cargos and 514 drivers of cargo vehicles with general cargos were surveyed and these drivers correctly completed the questionnaire survey.

Table 1 Percent of cargo drivers who would select to shift departure time from rush hours to Non-Rush hours if variable pricing would be applied

Cargo Types	Increased Toll Price During Rush Hours				Decreased Toll Price During Non-Rush Hours			
	10%	20%	30%	40%	10%	20%	30%	40%
Emergency Cargo	20%	45%	74%	75%	0%	0%	0%	25%
General Cargo	35%	74%	91%	96%	18%	81%	89%	91%

Table 2 Summary of selection distribution of cargo vehicle drivers with different characteristics

Characteristics		Percentage (%)
Responsible party for toll fee	Driver's cost	87
	Unit's cost	13
Cargo type	Emergency cargo	24
	General cargo	76
Travel distance	Short distance	4
	Long distance	96
Vehicle load	Less than 10 ton	13
	10 to 40 ton	72
	More than 40 ton	15

Based on the results shown in Table 1, it is found that as toll price increased during rush hours the percent of drivers who selected to shift departure time was much more than the percent of drivers who selected to shift departure as toll price decreased during non-rush hours. Therefore, cargo drivers are more sensitive to the increased toll price than to the decreased toll price. The main reason could be explained as follows: Cargo drivers/owners may need to figure out their transportation cost and profit. If travel cost increases (as toll price increases during rush hours) they may change their departure time. However, if travel cost decreases (as toll price decreases during non-rush hours), they may keep the same departure schedule without sacrificing their total

travel time because their transportation contracts may have certain items that could specify their arrival time.

Another factor that could affect cargo vehicle drivers' departure time is whether the toll fee can be reimbursed by other responsible unit. If a driver has to take care of the toll fee, the driver would be more willing to change his/her departure time to save some toll fee. In this category, "driver's cost" (meaning the driver is responsible for the toll fee) and "unit's cost" (meaning the toll fee will be reimbursed by a unit) are used to represent the characteristic. From field survey results, it was found that 87% of drivers were responsible for the toll fees (drivers' cost) and 13% of drivers could reimburse their toll fees (unit's cost). It is clear that the majority of drivers are responsible for toll fees. Thus, toll fee changes due to variable pricing schemes could have significant impacts on the selection of travel time. The factors discussed above and corresponding survey results are summarized in Table 2. At the beginning of the research, there was a plan on surveying cargo firms and cargo receivers. However, from the field survey on cargo drivers, it was found that majority drivers (about 87%) were responsible for all transportation costs. These drivers will get fixed payment for the cargo transportation. Therefore, if toll costs could be saved, drivers could get more income. That is way cargo toll variable pricing has so sensitive impacts on the departure plan of cargo vehicles. In fact, this is a very general situation in China for cargo drivers' payment. Also, from field survey results, it was found that majority cargo drivers have certain flexibility in their arrival time once they start their cargo transportation trips. The main reason for this is that cargo owners usually want to ensure their cargos can be safely delivered to the receivers. Therefore, cargo receivers usually put time restrictions on when the cargo could be delivered, not on when the cargo could be received. With such field survey results and the considerations addressed here, the inputs from cargo firms and receivers were not included in the modeling procedure in the paper.

4.3 Factors to be included in modeling

Based on the survey results and the above discussion, the following factors were included as the characteristics in the modeling processing: (1) cargo type –emergency cargo and general cargo, (2) responsible party for toll fee – driver's cost and unit's cost, (3) travel distance – short distance and long distance, and (4) vehicle load – 10 tons or less, 10 to 40 tons, and 40 tons or more.

Usually, total travel time could be a factor in affecting drivers' travel plan. However, as stated before, cargo drivers in China usually have sufficient flexibility in determining their total travel time and departure schedule, mainly for the safety reasons. Besides, the additional time spent for departure time change due to the impact of variable pricing is a very small part in the total travel time and could be adjusted or made up by their rest schedule. Also, cargo drivers in China pay more attention on the trip costs than on trip time because the flexibility in travel time. In fact, the variable "travel distance" may have certain correlation with travel time. With such considerations, the factor "total travel time" was not included in the model specifications.

5 Model Development

As stated previously, division of time segments is categorized as: rush hours and non-rush hours. There are two scenarios (model types): Model Type One - basic toll rate for rush hours and lower toll rate for non-rush hours, Model Type Two – higher toll rate for rush hours and basic toll rate for non-rush hours.

5.1Model specification

Cargo vehicle drivers would determine their departure time based on their characteristics and the toll rate. Generally, the drivers traveling during a time segment with a low toll rate would not change their departure time. Only the drivers traveling during a time segment with a higher toll rate would shift their departure time from the segment to another segment with a lower toll rate. Thus, the models will focus on the time segments with relatively higher toll rates including the basic rate if a lower toll rate is used for non-rush hours.

The logit models which estimate the probabilities for drivers to change their departure time and not to change their departure are shown in Equations 1 and 2. The utility functions expressed in Equations 3 and 4 can be simplified based on the previous discussion and expressed as follows:

$$V_{1n} = a_1 + a_2P_{Nr} + a_3C + a_4R + a_5D + a_6L_1 + a_7L_2 + a_8L_3 \quad (5)$$

$$V_{2n} = a_2P_r \quad (6)$$

where:

a_i (i = 1, 2, …, 8) – model parameters to be estimated,

P_{Nr} – toll rate during non-rush hours,

P_r – toll rate during rush hours,

C – cargo type (urgent cargo: C = 0 , general cargo: C = 1),

R – responsible party for toll fee (driver's cost: R = 1, unit's cost: R = 0),

D – travel distance (short distance: D = 0, long distance: D = 1),

L_1 – vehicle load (if 10 tons or less, L_1 = 1, otherwise, L_1 = 0),

L_2 – vehicle load (if 10 to 40 tons or less, L_2 = 1, otherwise, L_2 = 0), and

L_3 – vehicle load (if 40 tons or more, L_3 = 1, otherwise, L_3 = 0).

5.2 Model type one

A 24-hour time segment is divided into rush hours and non-rush hours. As stated before, for model type one, the basic toll rate is used during rush hours and lower toll rate is used during non-rush hours. The data collected from field questionnaire surveys were analyzed with the use of SAS analysis toll. The modeling procedure performed in the research was to estimate the parameters specified in Equations 5 and 6.

With ample recognition in the discrete choice modeling that the critical t-values for determining the inclusion of a variable should be about 1.00. According to the calibration results shown in Table 3, the t values of all parameters are greater than 1. Thus, it is reasonable to keep all the proposed variables specified in model specification. The following equations present the model calibration results:

$$V_{1n} = -2.64 - 14P_{Nr} + 1.14C - 0.13R - 0.22D - 0.9L_1 - 0.84L_2 - 0.94L_3 \quad (7)$$

$$V_{2n} = -14P_r \quad (8)$$

As shown in the above equations, the parameter sign for toll rate is negative, meaning as toll rate increases, the probability for drivers to change their departure time will decrease; The parameter sign for cargo type is positive, meaning drivers of

cargo vehicles with general cargos would have more probability to change their departure time if variable pricing schemes are installed.

Table 3 Preliminary modeling results for model type one

variables	Toll rate (x_{1n},x_{2n})	Responsible party for toll fee (R)	Cargo type (C)	Travel distance (D)	Vehicle load			constant (a_1)
					Less than10 ton (L_1)	10 to 40 ton (L_2)	More than 40 ton (L_3)	
parameter	−14	−0.13	1.14	−0.22	−0.90	−0.84	−0.94	−2.64
Standard deviation	0.35	0.11	0.09	0.21	0.10	0.08	0.10	0.19
T-statistic	−39.96	−1.20	13.24	−1.04	−9.26	−10.71	−9.89	−14.13
LogLikelihood	−2552							
Likelihood Ratio(R)	3381.1							

Generally speaking, the logit models shown in Equations 1, 2, 7, and 8 are reasonable. These models can be applied to real variable pricing practices.

5.3 Model type two

The basic scheme for model type two is that higher toll rates will be used during rush hours and the basic toll rate will be used during non-rush hours. The basic purpose of model type two is to analyze the probability for drivers to change their departure time as the toll rate during rush hours increases. Table 4 presents the modeling results which are based on the specification specified in Equations 5 and 6. Similar to model type one, the t values for the characteristics R is less than the t table value which is 1.0 The characteristics R is considered not significant in the model specification. Thus, the characteristics R can be removed from the model specifications, and the same model specifications as used in model type one can be used. Based on the simplified model specification, another modeling result are obtained and presented in Table 5. The utility functions corresponding to the model specification is given as follows:

$$V_{1n} = -2.74 - 16.89P_{Nr} + 1.17C - 0.26D - 0.91L_1 - 0.87L_2 - 0.97L_3 \quad (9)$$

$$V_{2n} = -16.89P_r \quad (10)$$

The discussion to parameter signs is similar to model type one. It will not be repeated here.

Table 4 Preliminary modeling results for model type two

Variables	Toll rate (x_{1n},x_{2n})	Responsible party for toll fee (R)	Cargo type (C)	Travel distance (D)	Vehicle load			constant
					Less than10 ton (L_1)	10 to 40 ton (L_2)	More than 40 ton (L_3)	(a_1)
Parameter	−16.9	−0.09	1.18	−0.28	−0.84	−0.79	−0.9	−2.47
Standard deviation	0.44	0.11	0.09	0.22	0.10	0.08	0.10	0.2
T-statistic	−38.32	−0.83	13.12	−1.25	−8.21	−9.61	−8.96	−2.53
LogLike lihood	−2279							
Likelihood Ratio(R)	2982.7							

Table 5 Final modeling results for model type two

variables	Toll rate (x_{1n},x_{2n})	Cargo type (C)	Travel distance (D)	Vehicle load			constant (a_1)
				Less than 10 ton (L_1)	10 to 40 ton (L_2)	More than 40 ton (L_3)	
parameter	−16.89	1.17	−0.26	−0.91	−0.87	−0.97	−2.74
Standard deviation	0.44	0.09	0.21	0.09	0.06	0.08	0.09
T-statistic	−38.32	13.07	13.11	−10.57	−15.78	−11.99	−30.11
LogLikelihood	−2280						
Likelihood Ratio(R)	2980.6						

6 Conclusions

Variable pricing models have been used in many applications. In China, cargo vehicle traffic has been a large part in the traffic stream. Due the relative poor operational characteristics, cargo vehicles are more frequently involved in freeway traffic crashes, particularly, severe crashes. A possible way to balance cargo vehicle demand distribution over time, variable pricing models could be used as a potential management method. The paper estimated variable pricing models for cargo vehicles. The variable pricing models have good potential applicability to balance cargo traffic demand over time. However, for the applications of cargo toll variable pricing, optimal variable pricing schemes should be implemented. There is an urgent need to market the applications of variable pricing so that more and more cargo drivers could understand the benefits of variable pricing and take advantage of the applications.

REFERENCES

Pigou(1920), *The Economics of Welfare*. London: Macmillan.

Knight(1924), F.H. “Some Fallacies in the Interpretation of Social Cost”, Quarterly Journal of Economics, 28A, pp.582-606.

Weiming Liu, Zheren Wang, Xitao Zheng(2000), *Theory and Method of Expressway Toll Collection System*, China Communications Press, PP31.

Lam,T. C., Small, K.A.(2001), "The Value of Time and Reliability: Measurement from a Value Pricing Experiment". Transportation Research Part E37, pp. 231-251.

Sullivan,E.(2000). *Continuation Study to Evaluate the Impacts of the SR91 Value-priced Express Lanes*. Cal Poly State University, San Luis Obispo, CA.

Jia Yan, Kenneth(2002), "A.Small. Choice Models of Route, Occupancy, and Time-of-Day with Value Priced Tolls", Transportation Research Record.

Kawamura, Kazuya(2000). "Perceived Value of Time for Truck Operators", Transportation Research Record 1725, pp.31-36.

Brownstone, D., A. Ghosh, T.F. Golob, C. Kazimi and D. Van Amelsfort(2003), "Drivers's Willingness to Pay to Reduce Travel Time: Evidence from the San Diego I-15 Congestion Pricing Project", Transportation Research A, 37A, pp.373-387.

Burris, M.W.and R.M. Pendyala(2002). "Discrete Choice Models of Traveler Participation in Differential Time of Day Pricing Programs", Transport Policy 9(3), pp.241-251.

Jose Holguin-Veras, Mecit Cetin, Shuwen Xia(2006), "A Comparative Analysis of US Toll Policy", Transportation Research Part A, pp.852-871.

Jose Holguin-Veras, Qian Wang, Ning Xu etc(2006), "The Impacts of Time of Day Pricing on the Behavior of Freight Carriers in a Congested Urban Area: Implications to Road Pricing", Transportation Research Part A, pp.744-766.

Traffic Volumes of Trunk Roads in Jiang Su Province in 2002, Jiang Su Province Communication Department Highway Bureau, 2003.8.

Traffic Volumes of Trunk Roads in Jiang Su Province in 2003, Jiang Su Province Communication Department Highway Bureau, 2004.8.

Traffic Volumes of Trunk Roads in Jiang Su Province in 2004, Jiang Su Province Communication Department Highway Bureau, 2005.6.

Jian Lu, Qiaojun Xiang, Yingwei Ren(2006), *Research on Impacts of Variable Pricing Program on Traffic Flows Distribution on Jiang Su Freeways*, A Research Report Submitted to Jiang Su Province Department of Transportation, Transportation College, Southeast University, Nanjing, China.

Hongzhi Guan(2004), *Disaggregate Model-A Tool of Traffic Behavior Analysis*, China Communications Press.

Manski, C.F. and D. McFadden(1981), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press.

Logical Definition and Application of Expressway Network Guide Sign System

Yuren CHEN[1], Yueru WANG[2] and Shaojun CHEN[3]

Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai, China, ST 200092
[1]PH (8621) 6598-2276, FAX (8621) 6598-2507, email: Chenyr@mail.tongji.edu.cn
[2]PH (8621) 6958-7730, FAX (8621) 6598-2507, email: wyr_1982@163.com
[3]PH (8621) 6958-4649, FAX (8621) 6598-2507, email: knightofchj@163.com

Abstract

During the design and administration of expressway traffic signs, people always treat one single sign, instead of thinking about the relationships of signs on an entire network composed of different expressways. In fact, traffic signs, especially the guide signs, act as a system, which seems to have great logic. Their locations, types and information on the panels are reasonable and correlative. The paper designs the topology model of the expressway network, with the consideration of exits, entrances, interchanges and nodes of expressways, makes a logical definition of the guide signs, and discusses the relationship between the guide signs' logical definition and expressway network topology. With the logical definition, there are a lot of applications, such as automatically choosing as well as altering the guide signs that are related to the added expressway or the dropped one, and altering the relative guide signs with the adjustments of the topology of the expressway network during the expressway rebuilding and traffic administration.

Key words: *guide signs, expressway network, topology, information management system, traffic administration*

1 Introduction

As the expressway system develops as a network, it is more and more difficult to add guidance for expressway users. Guide signs, which contain the most important parts of the traffic information, are becoming quite complicated in their information and locations. Guide signs help drivers to choose their routes, show them where to get into the expressway and where to get off, and also tell them how far are their destinations. So the information and locations of guide signs have a close relationship with the topology of the expressway system. In the meanwhile, it is necessary to set guide signs in groups to give enough information to the drivers and help them to make decisions during their journeys. It is of great use to study the logic of the information and locations of guide signs to make correct guidance for the drivers, as well as making the work easier and more effective for the expressway administrators.

2 The Expressway Network Topology Structure

As the location and information setting on guide signs are decided by the expressway network topology, it is necessary to build the topology structure of the expressway network before studying the logical definition of guide signs. The route is the basic element of an expressway, and then there will be primary reference points, such as exits, entrances, the starting point, and the end point. In the end, there should

be shifting information, which mainly consists of information of interchange ramps and ramps of exits and entries.

2.1 Expressway network and nodes

There should be two nodes (the starting point and the end point) for the radial expressways and more for the annular ones. The basic topology (see Figure1.) could be built by lines referencing the routes and points referencing the nodes. Then, there should be KM-number system put on the lines. After this, there should be the relationship between the Descartes coordinates and the linear coordinates within the precision. For each node, there should be a pointer in the database to indicate which expressway it belongs to.

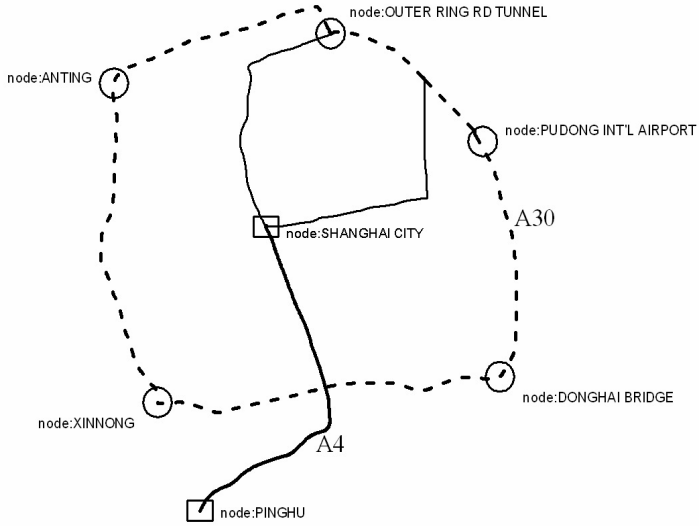


Figure 1. Basic Expressway Topology

2.2 Topology of exit, entry and interchange ramp systems

In the basic expressway topology, each expressway is treated separately. There is neither the information of showing how to get into or off the expressway nor the information of shifting between expressways. So the topology information of entries, exits and interchange ramps should be added to the topology model (see Figures 2 and 3), and Tables 1, 2 and 3 show which kinds of information should be involved.

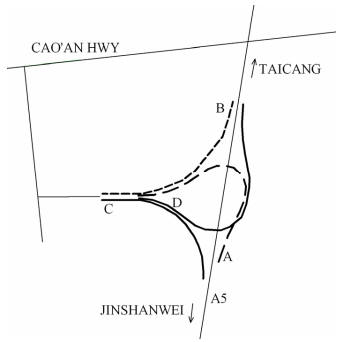


Figure 2. Topology of Exit and Entry Systems

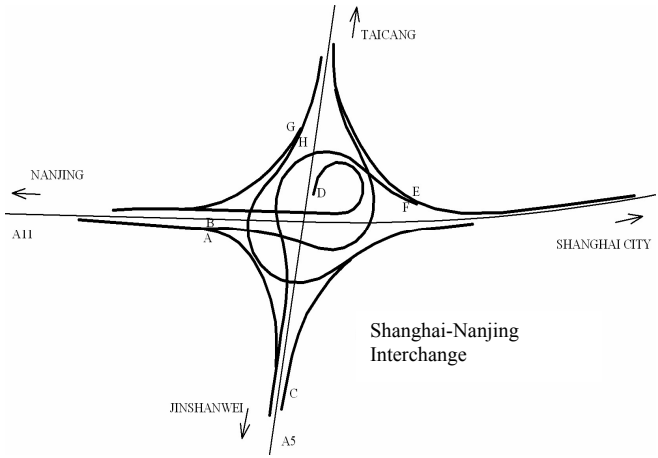


Figure 3. Topology of Interchange Ramps

Table 1. Topology Information of Exit System

Exit_NO	Exit_Name	Expwy_NO	KM-number	Ramp_ID	Expwy_Node	Ramp_Type
5	CAO'AN HWY	A5	23000	A	TAICANG	Left
5	CAO'AN HWY	A5	22750	B	JINSHANWEI	Right

Table 2. Topology Information of Entry System

Entry_NO	Entry_Name	Expwy_NO	KM-number	Ramp_ID	Expwy_Node	Ramp_Type
15	CAO'AN HWY	A5	23000	C	TAICANG	Left
15	CAO'AN HWY	A5	22375	D	JINSHANWEI	Right

Take Exit ramp A in Figure 2 for example. Table 1 shows that the exit number is 5; the exit name is CAO'AN highway; the exit belongs to the A5 expressway; the KM-number on A5 of the starting point of the ramp is K23+000; the ramp of the exit is named A; the ramp starts from A5 with the direction of to-Taicang; and the ramp type is left-turn.

Table 3. Topology Information of interchange Ramps

ID	Name	Exit_ NO	RP_ ID	S_ EPX	E_ EPX	S_KM	E_KM	S_Node	E_Node	RP_ Type
12	Shanghai -Nanjing	2A	A	A11	A5	11400	11625	SHANG-HAI City	JIN-SHAN-WEI	Y-Right
12	Shanghai -Nanjing	2A	B	A11	A5	11400	10650	SHANG-HAI City	TAI-CANG	Y-Left
12	Shanghai -Nanjing	6	C	A5	A11	12000	3097	TAI-CANG	SHANG-CITY	A (of AB)
12	Shanghai -Nanjing	6	D	A5	A11	11125	11400	TAI-CANG	NAN-JING	B(of AB)
12	Shanghai -Nanjing	2A	E	A11	A5	3097	10575	NAN-JING	TAI-CANG	Y-Right
12	Shanghai -Nanjing	2A	F	A11	A5	3097	11625	NAN-JING	JIN-SHAN-WEI	Y-Left
12	Shanghai -Nanjing	6	G	A5	A11	10575	11400	JIN-SHAN-WEI	NAN-JING	Y-Right
12	Shanghai -Nanjing	6	H	A5	A11	10575	3097	JIN-SHAN-WEI	SHANG-CITY	Y-Left

Take ramp A in Figure 3 for example. Table 3 shows that the ramp belongs to the interchange named Shanghai-Nanjing with the ID of 12 in the database and the exit number is 2A; the ramp name is C; it takes the function of making shift-driving from A11 with the to-Shanghai-city direction to A5 with the to-Jinshanwei direction; the ramp starts from A11 K11+400 and ends at A5 K11+625; the ramp type is turn-right of the Y-type.

3 Logical Definition of Guide Signs on the Expressway Network

When the topology is built, it is time to make the logical definition of guide signs follow the relative design and plan rules. In the topology model of the expressway network, the reference points (consisting of exits, entries, interchanges and the nodes), divide an expressway into several sections. Within each section, there are rules to set a group of guide signs. Figure 4 shows how to set guide signs on a section (on the A5 expressway, with the to-Taicang direction, from the reference point of Cao'an highway to the reference point of Bao'an highway).

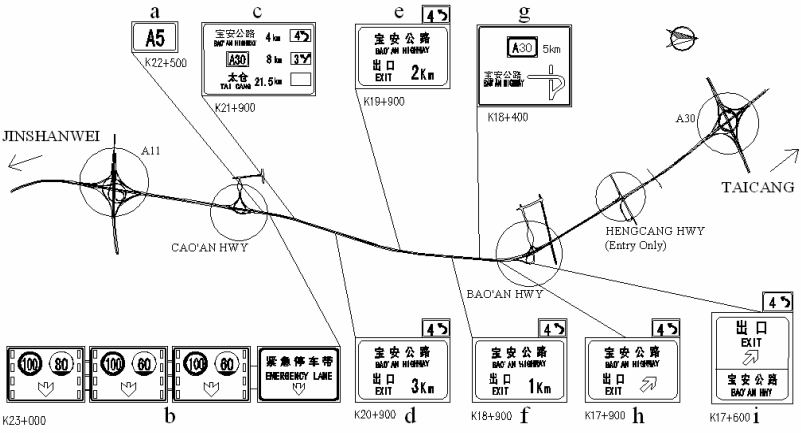


Figure 4. Guide Signs on an Expressway Section

There are rules for the setting of these guide signs according to the section topology. As shown in Figure 4, at the end of the entry ramp, where one gets into the expressway, there should be an Expressway-Number sign (Sign a in figure 4) announcing the number of the expressway, which is A5 in this instance. Then there should be a group of Lane-Speed-limit signs (b) to advise expressway users about the permitted speed range or function of each lane. Also there should be a Destination-Distance sign (c) to give users the information of the next exit, the exit past the next, and the node to arrive in the current direction on the expressway. The information of each destination consists of name, distance, exit-number and exit ramp type (for the next exit and the exit past the next). The information has close relationship with the expressway topology information, for the destination is either a reference point or a node; the distance is decided by the KM-number on the road, and the exit number is decided by the exit number system of the entire expressway. The exit-ramp type just represents the exit topology. When it comes to the zone near the exit, there should be Exit-Advance signs (d,e,f), and they should be placed at 3km, 2km and 1km in advance of the beginning of the decelerating lane carrying the distance information. At 0.5km before the beginning of the decelerating lane, there should be a Diagram-Advance sign (g), and a diagram giving the shape of the interchange, or exit ramps should be placed on the sign panel with the name of the exit and next exit or the number of the expressway to shift to, as well as the distance to the next exit or expressway. At the beginning of the decelerating lane, there should be an Exit-Direction sign (h) with the same face-design as the Exit-Advance signs, except that there should be an arrow instead of the distance information. An Exit-Gore sign (i) should be located where the ramp branches from the main road. For the Exit-Advance sign, the Exit-Direction sign and the Exit-Gore, there should be a small panel carrying the information of the exit number and the exit type on the right top of the main panel.

4 Application of the logical definition

In the situation of dropping the exit ramp of the reference point of Bao'an highway on A5 expressway with the direction of to-Taicang (see Figure 4), in this direction, the section of Cao'an highway to Bao'an highway and the section of Bao'an highway to Hengcang highway would combine into one, and the guide signs placed on the primary two sections should be modified. As shown in Figure 4, Sign a and Sign b would get no change, while the others would be updated or dropped and some new signs should be placed. On the panel of Sign b, the information on the first line would be of the primary information of the second line(A30), and the second line would carry the information of the exit past A30(for Hengcang highway has only entries), and the bottom line of the node information would remain unchanged. Signs d, e, f, g, h and i would be dropped for they carry the topology information of Bao'an highway exit. There would also be some modification on other sections for their signs to carry information of the dropped exit.

5 Conclusions

As the expressway network becomes more and more complicated, it is getting harder and harder to set guide signs to guide road users clearly, and the administration of the signs is becoming a difficult job. With the logical definition of guide signs based on the topology model of the expressway network, an information management system of expressway guide signs could be built which will help to make efficient administration of guide signs.

References

Federal Highway Administration. (2003) *Manual on Uniform Traffic Control Devices for Streets and Highways (Part 2 Sighs)*

Shanghai Transportation Police Department. (2002) *General Rules for Traffic Signs and Indications on Expressway of Shanghai*

Research on Pedestrian-Vehicle Collision Model and Simulation Calculation

FANG, Shouen[1], WANG, Junhua[1]

[1]Key Lab Road and Traffic Engineering, Ministry of Education, Tongji University, 1239 Siping Rd, Shanghai, China 200135; PH (86-21) 65982276; FAX (86-21) 65982507; email: fangsek@mail.tongji.edu.cn, benwjh@163.net

Abstract: In recent years vehicle-pedestrian crash is significant, and pedestrian safety has been a serious problem all over the country. The conflict between the vehicle and the pedestrian is the key to the pedestrian safety research and a road layout safety plan. Based on the traffic conflict technology, this paper sets up the vehicle-pedestrian conflict model and vehicle-pedestrian crash model in the condition of one pedestrian crossing a single lane. In addition, the Monte-Carlo system simulation is used to calculate the probability of vehicle-pedestrian collision. With the tool of spatial analytic geometry, the Volume-Speed curve is built as the condition of setting pedestrian safety device.

Key words: *Pedestrian Vehicle System, Pedestrian Safety, Conflict Technology, Pedestrian-Vehicle Collision Model, Monte-Carlo Simulation*

1 General

In the recent years, the vehicle-pedestrian accident rate has been high in China. During 2003-2005, 53407 pedestrian casualties happened due to their violation behavior. Pedestrian accidents make up 17% of all the road traffic accidents [3]. Pedestrian safety has been a significant problem in China, and it is essential to provide a safe walking environment.

The conflict between pedestrian and vehicle is the key to the pedestrian safety research. Historical data shows that careless pedestrian rushing into the vehicle lane is one of the main causations for the pedestrian vehicle accident [6]. From the viewpoint of pedestrian safety, it is very dangerous for a single pedestrian to cross a free traffic flow [4]. There is neither an effective method to evaluate pedestrian's crossing behavior nor a standard to set pedestrian protection devices such as zebra crossings, traffic control signals etc. in China. Such problems could be solved by studying the probability of pedestrian vehicle conflict rate.

Road traffic accident especially the pedestrian-vehicle accident is such a random issue that it is almost impossible to get the traffic flow character when the accident happened. Thus it is more feasible to set up a simulation model to study the situation.

Ashton induced the relationship between speed distribution and pedestrian-vehicle collision in 1982 and the probability of being collided to death was calculated to illustrate the contribution of vehicle speed to the casualty [8]. Pasanen developed Ashton's model and Gray used the model to evaluate the effectiveness of Local Traffic Calming Device [4]. Xiaoliang Ma utilized the pedestrian vehicle conflict simulation to evaluate the Intelligent Speed Adaption System [8]. Ashton and Pasanen's

conflict and collision possibility model pays more attention to the speed variation. However, the traffic volume is not involved. Secondly, the road traffic environment and pedestrian's behavior in foreign countries differs widely from that in China so that the model could not be fully applied. Moreover a pedestrian's critical gap which represents his safety perception is not included in the above models, which is essential.

Based on the traffic conflict technique, the pedestrian vehicle conflict and collision model when a single pedestrian crosses a free-flow traffic lane is built and Monte—Carlo simulation is used to calculate the probability of pedestrian conflict and collision in different traffic condition. According to the calculation result, the justification for setting a pedestrian control device is discussed in the end.

2 Pedestrian-Vehicle collision model

2.1 Conflict mechanism

The procedure of pedestrian vehicle conflict when a pedestrian crossing the street is shown in Figure. 1.The vehicle and pedestrian's characters are described by t_p and t_v which represent the time they take to reach the conflict point. V_v is the initial speed of vehicle, d_v the distance from the conflict point, V_p the pedestrian's walking speed, and d_p the pedestrian's distance from the conflict point. Therefore $t_v = d_V / V_v$ and $t_p = d_p / V_p$, and the time gap needed by the pedestrian to cross the street $t_{cross} = W / V_p$

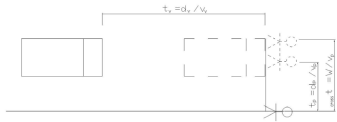


Figure 1. Pedestrian's crossing procedure

In the case $t_v > t_{cross}$,, the vehicle is far enough from the pedestrian that no conflict occurs. In the case $t_v < t_p$,, the vehicle reaches the conflict point before the pedestrian and the pedestrian will stop in front of the vehicle to avoid collision. In both two situations, the vehicle will not slow down. While $t_p < t_v < t_{cross}$,the vehicle must decelerate or it will collide the pedestrian who is crossing the street. Therefore the conflict between the pedestrian and vehicle comes into being. Considering the pedestrian's safety perception (critical cross gap) t_c=4.5s[7] , the probability of the conflict is shown as in Eq.1.

$$P_{conflict} = P(t_p < t_v < t_{cross}, t_v > t_c) \qquad (1)$$

2.2 The probability model of collision

Conflict is the direct cause that induces the accident and collision is the most

serious conflict[9]. It is supposed that t_{dv} is the time that the decelerating vehicle approaches the conflict point and t_r is the pedestrian's reaction time. Thus 3 possible situations exist when a conflict occurs:

Case1 $t_p > t_r + t_{dv}$, The driver finds the pedestrian dashing into the crosswalk and steps on the brake, and the vehicle stops completely in front of the conflict point.

Case2 $t_r + t_{dv} > t_{cross}$ The driver finds the pedestrian dashing into the crosswalk and steps on the brake, and the pedestrian reaches the end of the crosswalk before the vehicle comes.

Case3 $t_{cross} > t_r + t_{dv} \geq t_p$, The driver finds the pedestrian dashing into the crosswalk and step on the brake, and the vehicle could not entirely stop in front of the conflict point and will bump into the pedestrian. .

Therefore in the conflict condition, the probability model of pedestrian vehicle accident is shown as the following function

$$P = P(t_{cross} > t_r + t_{dv} > t_p, t_p < t_v < t_{cross}, t_v > t_c) \qquad (2)$$

The relationship between conflict and collision could be described as Eq.3.

$$P(A_{collision}|B_{conflict}) = P(A_{collision}, B_{conflict}) / P(B_{conflict})$$

$$= P(A_{collision}) / P(B_{conflict})$$

Therefore

$$P(A_{collision}) = P(B_{conflict}) P(A_{collision}|B_{conflict}) \qquad (3)$$

2.3 Parameter establishing

The parameters of the model consult the reference [1], [5], [8], [9].

The distance from the pedestrian's starting point to the conflict point d_p obeys uniform distribution d_p（d_p~U(2,6)）. The minimum distance 2m is considered by the width of the bicycle lane. The pedestrian's walking speed conforms to normal distribution (v_p~N(1.5,1),$v_{p\,min}$=0.5). While the pedestrian is confronted by the conflict his running speed conforms to (v_p~N(3.02,0.51) which is calculated through the national 4×10m shuttle racing standard

The vehicle's arrival time conforms to negative exponential distribution.

The driver's reaction time conforms to logarithm normal distribution（t_r~LnN(1.2,1)）.

The vehicle's minimum brake time t_{dv} refers to middle vehicle brake effect in sun time.

2.4 Hypothetic condition of the model

The model comes into existence under the following 4 conditions.

(1) The model object is to calculate the collision probability when a single pedestrian crosses a free traffic flow.

(2) The model neglects the situation that pedestrian goes back to the footpath when he feels threatened.

(3) Although the bicycle lane width is considered, the model supposes that there is no

bicycle when the pedestrian is crossing the street.

(4) It is assumed that the vehicle decelerates and keeps original trace when it is exposed to the conflict. But the vehicle sometimes swerves to evade the pedestrian in the real world.

3 Simulation calculation

Because most variables in the model are random and conform to a certain contribution model, Integral calculation on Eq.2 is complex and Monte-Carlo Simulation is used. The simulation procedure is shown in Figure. 2. The input data in the model are the operating speed from 30km/h to 70km/h and traffic volume 200PCU/h to700PCU/h. According to the distribution characters of the different parameters in the model, The tool of MATLAB is used to generate the random. When combining different vehicle rates of speed and traffic flow and simulating N time, the probability pedestrian-vehicle conflict could be obtained as shown in Figure. 3 and the pedestrian-vehicle collision probability calculation result is shown in Table1. When n=30000, the result tends to convergent.

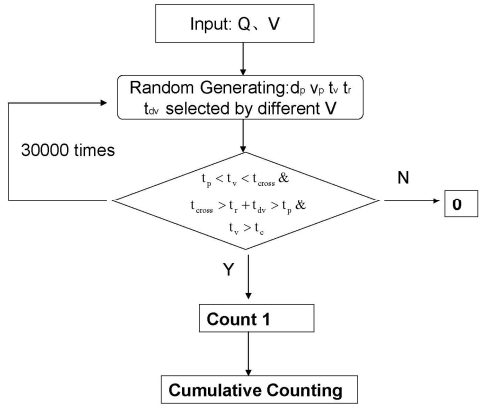


Figure 2. The simulation procedure

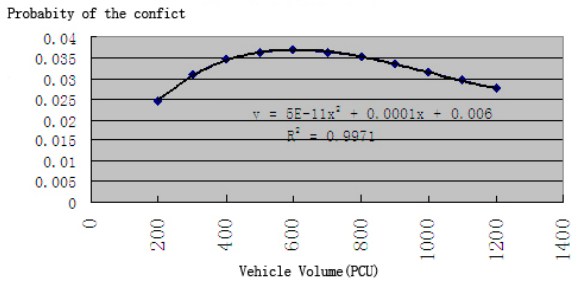


Figure.3. The relationship between traffic flow and conflict probability

Table.1. Collision probability

Traffic Vol. (PCU) / Speed （km/h）	200	300	400	500	600	700
30	0.0327%	0.0708%	0.1015%	0.1522%	0.2172%	0.2299%
40	0.0379%	0.0786%	0.1098%	0.1651%	0.2217%	0.2542%
50	0.0468%	0.0938%	0.1409%	0.1976%	0.2686%	0.3038%
60	0.0497%	0.0990%	0.1590%	0.2259%	0.2972%	0.3342%
70	0.0549%	0.1102%	0.1701%	0.2429%	0.3162%	0.3611%

4 Result discussion

(1) According to the model, the probability of the pedestrian vehicle conflict is determined by traffic volume and the severity of the conflict is determined by the speed. Additionally quadratic function could describe the relationship between traffic flow and conflict probability. When the traffic volume is in the vicinity of 600pcu/h, the conflict probability is the highest. Because when the volume exceeds 600pcu/h, fewer safe gaps will exist and the exposure to the conflict risk is lower.

(2) Table1 could be described as a pedestrian vehicle collision risk sinuate surface model and it could be imitated with successive dualistic function. To be simplified, AcDbPolyFaceMesh of ObjectArx in AutoCAD environment is utilized to construct a dispersive triangle surface model with the point in Table 1 and all these figures could be depicted with a certain point in the dispersive triangle surfaces shown in Figure. 4.

(3) The collision risk value is between 0.0327% and 0.6311% in Table1 with different traffic volume and rates of speed. 15% risk value 0.0008196 is applied in traffic control warrant. Projecting the points of Prob＝0.0008196, a Volume-Speed Curve in Figure 5 could be obtained.

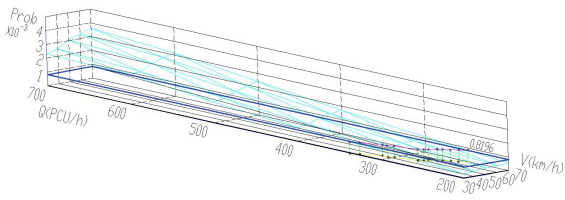


Figure. 4. Dispersive triangle surfaces risk model

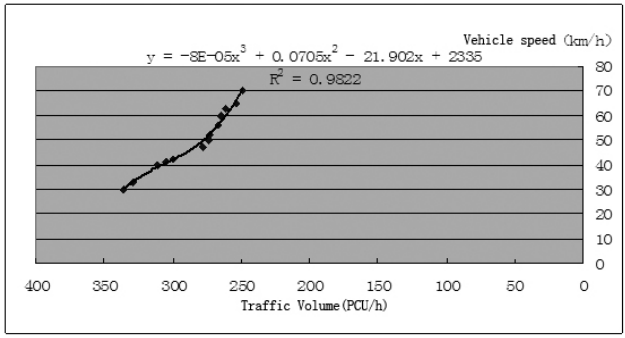


Figure.5. Projected curve

(4) The curve in Figure.5 is the justification to set pedestrian Safety Device, and it is proper to set the device when the traffic volume and speed parameters are located above the curve.

(5) This paper sets up the basic model of pedestrian vehicle collision and the risk of a single pedestrian crossing a single lane is calculated, based on this the risk of a pedestrian crossing multiple lanes could be further studied.

Reference

Feng Shulin(2004). P*edestrian's Crossing Speed In The Intersection.* Journal of Haerbing Industry Uni. Jan 2004 Vol36 No.1

GARY A. DAVIS (2000). *Method for Estimating Effect of TrafficVolume and Speed on PedestrianSafety for Residential Streets*，*TRANSPORTATION RESEARCH RECORD 1636* Paper No. 98-1330

Luo Yuan(2005). *Traffic Accident in China 2004*, Juristic Daily

Pasanen,E,Salmivaara,H. (1993). *Driving speeds and Pedestrian safety in the city of Helsinki*.Traffic Eng. Control 308-310

Standard of National Physical Education People's Physical Education Pub. 2000

Stutts,J.,Hunter,W.,Pein,W.(1996). *Pedestrian crash type:1990's update*. Trans. Res. Rec.1583,68-74,\

Wang Junhua, Fang Shouen *Research on the Pedestrian's Critical Gap*, Submitted to PIARC 2007

iaoliang Ma□, Ingmar Andr´easson(2005). *Predicting the effect of various ISA penetration grades on pedestrian safety by simulation*

Zhang Su (2000). *Chinese Traffic Conflict Technique*, People's Traffic Pub.

Zhou Shangwu(2001). *Traffic Engineering,* Tongji Uni Pub.

SAFETY EVALUATION APPROACH OF HIGHWAY INTERSETION

Li YUAN[1], Jian LU[2], Guoqiang ZHANG[3]and Qiaojun XIANG[4]

[1] Transportation College, Southeast University, Nanjing, 210096, P. R. China; PH (86)025-8390776; FAX (86) 025-83795352; email: yuanlibox@sina.com
[2] Transportation College, Southeast University, Nanjing, 210096, P. R. China; PH (86)025-8390776; FAX (86) 025-83795352; email: jianjohnlu@sina.com
[3] Transportation College, Southeast University, Nanjing, 210096, P. R. China; PH (86)025-8390776; FAX (86) 025-83795352; email: guoqiang_zhang@163.com
[4] Transportation College, Southeast University, Nanjing, 210096, P. R. China; PH (86)025-8390776; FAX (86) 025-83795352; email: xqj@seu.edu.cn

Abstract: Highway safety is one of the most important issues in transportation. Intersections are the locations with higher traffic crashes as compared with other highway locations. This paper introduces an approach to evaluate highway intersection safety performance. This approach is fully based on the existing conditions of the intersection, including geometrics, sight distance, pavement surface conditions, traffic control devices, traffic signal timing and phasing, etc. The non-accident based approach is based on field survey to the conditions mentioned before. The approach will also result in a safety index to indicate the safety performance degree of the intersection. Meanwhile, corresponding countermeasures are ranked and recommended based on the cost-benefit analysis. Results from the real application indicate that the approach has good applicability and can be used by field safety engineers in real application.

Key words: *highway safety, intersection, safety diagnosis, safety improvement*

1 Introduction

A highway intersection is the place where traffic and non-motorized users interact. Due to the conflicts from all users, more traffic crashes could happen at the intersection as compared with roadway segments. Thus, traffic safety at intersections is one of the most important issues to be resolved by traffic engineers. Based on statistics, in the whole world, about 45% traffic crashes and 25% fatal traffic crashes happen at intersections (FHWA,2000) (TRB,2003). In China, traffic safety problems at highway intersections are also critical, and past statistics indicate that about 30% traffic crashes happen at intersections (Zhang,1994). In order to improve highway intersection traffic safety performance, many research studies have been done. The most popular methods to analyze traffic safety performance are based on traffic crash analysis methods (Danief,1983). Non-crash-based analysis methods have been used by

many researchers. One of the methods is traffic conflict analysis (Lu,2004). Based on the definition from Perkins and Harris (Hauer,1986), a traffic conflict is defined as the interactions between vehicles and such interactions would result in some actions, such as braking and movement direction change (Sayed,1999) (Lu,2001). In the research presented in this paper, another approach called diagnostic approach was attempted.

Different from the crash-based and conflict-based methods, this approach is a non-crash and non-conflict based analysis method. With such characteristics, this approach could give a relatively fast evaluation to intersection safety performance. In this approach, a safety index is introduced to give a diagnostic summary of the safety performance of an intersection.

2 Methodology Description

As stated previously, the diagnostic approach for intersection safety evaluation is based on the field survey. Then, potential safety problems are ranked according to the criteria such as the significance and severity. After safety problems with the highest ranks are identified, possible countermeasures are proposed and the cost-benefit analysis is performed to estimate the benefit-cost ratio for each countermeasure. Finally, corresponding countermeasures with the highest benefit-cost ratios are recommended to improve the safety of the intersection. The diagnostic approach can be described by a flow process as shown in Figure 1. The following subsections give a brief description to each step shown in Figure 1 (Yuan, 2005).

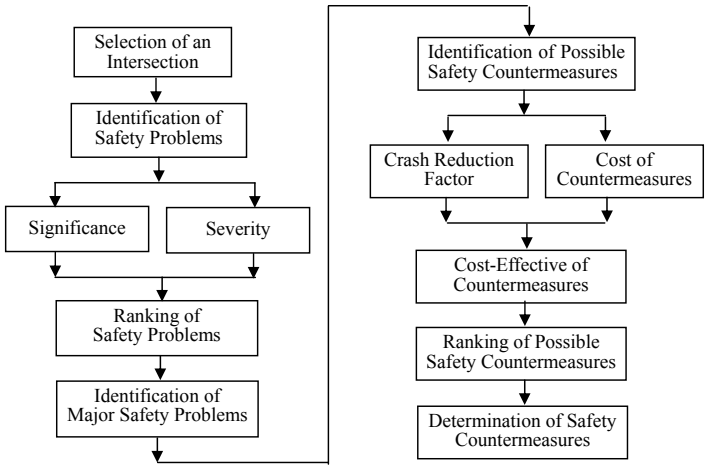


Figure1. Flow process of diagnostic approach

2.1 Survey Form

For intersection safety diagnostic process, field surveys are necessary. Field engineers use the survey forms in the fields to perform diagnostic process by filling the survey forms (also called safety diagnostic forms) with information collected from each intersection. Development of the survey form is based on past crash analysis results. The basic information (elements) in the form covers intersection existing conditions that have direct relationships or impacts to traffic safety at intersections. Typical elements include, but not limited to, geometrics, sight distance, pavement surface conditions, traffic control devices, traffic signal timing and phasing, lateral clearance, lighting, and approaching speed. Field engineers will check these elements at each intersection and they will give scores to each element based on their personal judgments. One score is under the category of "significance" and the other one is under the category of "severity". Field engineers need to fill scores for each element under the categories of "significance" and "severity". This form is relatively easy to be used in fields by an "average" field engineer.

2.2 Identification of Safety Problems

This step includes two parts: (1) Field engineers need to use the survey form to survey each intersection. The survey items cover all related conditions. The surveyors need to use their field experience to judge the potential safety problems associated with the existing conditions at each intersection; Filling in the diagnostic table mainly depend on the researcher's own practical experience, Significance and the severity evaluation criterions, which are established according on present criterions, guides practical experience and the expert's suggestion, the purpose is to provide a uniform reference of the safety problems to the researchers. (2) This part needs field engineers to collect some data (such as traffic data) and analyze the data. The main purpose of data analysis is to give certain adjustment to the first part. With the qualitative judgment (from the first part) and the quantitative judgment (from the second part), the survey could produce better identification of safety problems at each intersection.

2.3 Ranking of Safety Problems

After completion of field survey and identification of intersection potential safety problems, the potential safety problems should be ranked. Some safety problems may be significant, but would not result in severe traffic crashes, and some safety problems may not be significant, but could result in severe traffic crashes. Thus, the indices "Significance" and "Severity" are used in ranking the potential safety problems. With the consideration of significance and severity, a safety index is introduced to rank a particular safety problem, such as non-sufficient sight distance for a given intersection. The safety index can be calculated by the following equation (Equation 1):

$$R_i = W_1 C_i + W_2 S_i \qquad (i = 1, 2, ..., n) \tag{1}$$

where:

R_i – safety index for the ith particular safety problem,

W_1 – weight for significance index,
C_i – significance index for the ith particular safety problem,
W_2 – weight for severity index,
S_i – severity index for ith particular safety problem, and
n – number of the potential safety problems.

In Equation 1, the safety index R_i represents the ranking of each particular safety problem, and if the ith particular safety problem has a larger R_i value, this problem should have a higher rank for safety improvements.

To use Equation 1, the weights (parameters W_1 and W_2) should be estimated. In this research, the AHP (Analytic Hierarchy Process) method was used to estimate the two parameters. AHP method is to combine expert's experience and theory analysis, and uses inter-comparing method. It will reduce the uncertain factors during the comparing process which is a usual one to estimate the weights.

Researchers and engineers with sufficient experience in highway design, construction, safety, maintenance, and traffic operations were visited and interviewed in Shanghai, Beijing, Jiangsu Province, and other places. Their judgments regarding the importance of "significance" and "severity" of each intersection element to traffic safety were entered into the AHP algorithm. Since all the statistical significance levels (I) were more than or equal to 0.6 ($I \geqslant 0.6$), the results from the AHP algorithm were acceptable. With the data collected from the interviews and AHP algorithm, the final parameters are: $W_1 = 0.38$ and $W_2 = 0.62$, or Equation 1 has the following form:

$$R_i = 0.38C_i + 0.62S_i \qquad (i = 1, 2, ..., n) \tag{2}$$

In order to rank intersections for safety improvements, the overall safety index Y for a particular intersection is used and described by Equation 3:

$$Y = \sum_{i=1}^{n} R_i \tag{3}$$

The overall safety index is the summary of the safety index for each particular potential safety problem. A larger Y value means the corresponding intersection has more significant and/or severe safety problems.

2.4 Identification of Possible Safety Countermeasures

The selection of possible safety countermeasures should follow several principles such as (1) reduction of crash rate and crash severity, (2) reduction of the conflicts between non-motorized users and motorized vehicles, and (3) reduction of conflicts between vehicles.

The selection of possible roadway safety countermeasures is relatively simple. Generally, one safety problem could have several countermeasures. On the other hand, one safety countermeasure could solve several safety problems. Many documents present potential roadway safety problems and corresponding possible safety countermeasures (ITE,1999).

2.5 Ranking of Possible Safety Countermeasures

In real applications, some countermeasures may cost less, but their safety effectiveness is also low; and some countermeasures are effective, but their costs are also high. Therefore, there is a need to optimize the process to select countermeasures. The general way to rank possible safety countermeasures is based on cost-benefit analysis and the benefit-cost ratio is often used as an indicator. In this research that is summarized in the paper, the following equation (Equation 4) is used as a ranking indicator:

$$R = \frac{BENEFIT}{COST} C \tag{4}$$

$$BENEFIT = \sum_{i=1}^{3} CRF_i C_i \tag{5}$$

where:

R – benefit-cost ratio,

C – average cost for each traffic crash,

i – 1, 2, 3, representing fatality, injury, and property damage only (PDO), respectively,

CRF_i – crash reduction factors for fatality, injury, and PDO, respectively (i = 1, 2, 3), and

$COST$ – average cost for a particular countermeasure.

R values can be used to rank possible safety countermeasures. The countermeasure with a larger R value should have higher priority (higher rank) for implementation. In real applications, due to the budget limitations, not all safety countermeasures will be implemented. Only the countermeasures with higher priority will be implemented. *CRF* values are obtained from long-term crash data accumulation and statistics analysis. The cost for a particular countermeasure includes implementation cost, maintenance cost, and operational cost. Cost data are obtained through long-term application accumulation and cost data statistics analysis.

2.6 Determination of Safety Countermeasures

Technically, the ranking of possible countermeasures can be used to determine safety countermeasures that will be implemented. In other words, Equation 4 or 5 is used to calculate the benefit-cost ratio for each possible safety countermeasure. If the implementation budget allows to the installation of countermeasures, the top countermeasures with the highest benefit-cost ratios will be selected for real implementation.

3 Application of the Diagnostic Approach

In the research, the diagnostic approach was applied to evaluate the safety performance of the intersection of Province Highway S327 and Province Highway S103. This application was to rank the potential safety problems caused by existing conditions and rank the possible safety countermeasures.

The intersection of S327 and S103 is in a hill area and is a STOP controlled intersection. Both highways S327 and S103 are major arterials and their main function is mobility. According to past crash data, the intersection has a high annual crash rate. Local transportation agencies have planned to improve the safety performance of this intersection. Therefore, there is a need to rank the safety problems at this intersection.

This intersection is not a right-angle one. The intersecting angle is about 40 degrees, which limits the sight distance for some approaches and makes it difficult for vehicles to make turns, especially for heavy vehicles, mainly due to the small turning radius. The intersection has a moderate traffic demand with the traffic volume on S103 being higher than the one on S327. Based on field observations, traffic conflict at the intersection is high, meaning traffic safety is not good.

Researchers from Southeast University in Najing performed a field survey in fall of 2005 to apply the diagnostic approach to rank safety problems at the intersection. Based on the safety index values, the safety problems can be ranked.

After the safety problems were identified and ranked in the research, possible safety countermeasures were proposed. Practically, a safety problem may have several corresponding countermeasures and one countermeasure may be used to improve several safety problems. After careful assessments, a list of possible countermeasures were proposed. In the same table, crash reduction factors (CRF) for each proposed countermeasure are also listed. With the use of Equation 5, BENEFIT values were obtained. Meanwhile, the cost data were estimated from past experience. Thus, the benefit-cost ratio value for each countermeasure can be calculated according to Equation 4.

As stated previously, the ranking of the countermeasures is based on the benefit-cost ratio. The top countermeasures with higher ranking will be implemented if the budget for safety improvements is limited.

The above describes a real application case with the use of the diagnostic approach developed in the research. It is found the approach is relatively easy to be used in fields. As shown in the real application, the approach has good applicability. To effectively use the approach, it is very important for field engineers to complete field surveys to intersections. Field engineers should have good field experience and safety judgments.

4 Conclusions

Crash-based traffic safety analysis needs a long-time accumulation of traffic crash data, and sometimes, it is difficult to obtain reliable traffic crash data. On the other hand, Conflict-based traffic safety analysis has the time-consuming nature, and sometimes, the determination of a conflict between vehicles depends on the personal judgment with certain subjective nature. To give a fast evaluation of the safety of an intersection, the existing conditions at the intersection could be used as the basis for intersection safety performance evaluation. The non-crash and non-conflict based

safety analysis approach has some advantages over the crash-based and conflict-based safety analysis approaches. However, to adequately apply the diagnostic approach, field survey forms should be well designed to include all possible safety problems and field engineers (surveyors) should be well trained.

References

Danief, S. (1983). "Turner and Edward R. Mansfield. Variability in Rural Accident Reporting", *In Transportation Research Record: Journal of the Transportation Research Board, No. 910,* TRB, National Research Council, Washington, D. C. 8-13.

Federal Highway Administration (2000). "*National Agenda for Intersection Safety*", U.S. Department of Transportation, Washington, D. C.

Hauer, E. (1986). "Research into the Validity of the Traffic Conflicts Technique", *Accident Analysis & Prevention*, 471-481.

Institute of Transportation Engineers (1999). "*Traffic Engineering Handbook",* Institute of Transportation Engineers, Washington, D. C.

Lu, J. J., Xiang, Q. J., Dissanayake, S., and Pernia, J. (2004). "The Ways to Improve Highway Safety", *Journal of Advances in Transportation Studies*, Section B 2, 3-14.

Lu, J. J., Yu, Z. (2001). *"Impacts of Intersection Signalization on Traffic Safety"*, Research Report to Florida Department of Transportation, Department of Civil and Environmental Engineering, University of South Florida.

Sayed, T., and Zein, S. (1999). "Traffic Conflict Standards for Intersections", *Transportation Planning and Technology,* 309-323.

Yuan, L., Lu, J. J., and Xiang Q. J. (2005). *"Research on Diagnostic Method of Highway Intersection Safety",* International Road Safety Seminar, Beijing, China, 348-340.

Transportation Research Board (2003). *"Guide for Addressing Unsignalized Intersection Collision, NCHRP 500, Vol.5"*, Washington, D. C.

Zhang, C. P. (1994). "Research on Traffic Safety Evaluation at Urban Road Intersections", *Journal of Tongji University*, 22, 47-52.

Multi-objective Transit Route Network Design Using Tabu search

Deming Lei[1 2] Xinping Yan[1]

[1]Center for Intelligent Transportation Systems, Wuhan University of Technology, Wuhan 430063, Email: deminglei11@163.com

[2]School of Automation, Wuhan University of Technology, Wuhan 430070

Abstract: This paper proposes a tabu search-based method to a multi-objective transit route network design problem. The whole design is performed in two phases. A route generation algorithm produces a good initial solution in the first phase. Tabu search continuously generates new solutions with route-merge, route-break and add-link operator. The external archive is used to store the non-dominated solutions that are maintained during the search process. The proposed algorithm is applied to a benchmark network for the following two objectives: operator cost and user cost. The sensitivity analyses are performed and computational results are presented.

Key words: *transit route network, multi-objective optimization, tabu search*

1 Introduction

Urban transit route network design (TRND) involves selecting the routes and the associated frequencies. The optimal design of transit routes and service frequencies can decrease the overall cost of providing transit service, which is generally comprised of user costs and operator costs. The chosen route network and suitable headway for each route can affect later planning steps, such as timetable development, bus and crew scheduling.

In the past decade, several models have been proposed in order to find a more efficient route network. Baaj and Mahmassani (1991) presented an artificial intelligence–based approach. Pattnaik et al.(1998) introduced a model using a generic genetic algorithm to select the optimal route network from the candidate routes set. Lee and Vuchic (2000) demonstrated an iterative approach that was based on the dynamic relation between variable demand and the optimal network. Tom et al.(2003) proposed a frequency coded genetic algorithm to the selection of a set of routes and the determination of the associated frequencies. Fan et al.(2006) developed a simulated annealing-based heuristic to solve the optimal bus transit route network design problem at the distribution node level. Ngamchai et al.(2003) demonstrated the efficacy of problem-specific genetic operators in handling bus transit route design and service frequency settings with the objective of minimizing the total system costs. Agrawal et al.(2004) proposed two parallel genetic algorithms for large scale transit route network design problem. Fan et al.(2006) introduced a genetic algorithm-based heuristic to research the underlying characteristics of the bus transit route network design problem with variable transit demand.

The above-mentioned algorithms use artificial intelligence, genetic algorithms or simulated annealing and most of them are used to select the optimal routes sub-set from the candidate routes set. However, the number of the candidate routes greatly increases, and the computational burden grows

exponentially with the size of the transit network. Some optimal routes may not be included in the candidate route set and the search of heuristic becomes invalid. This makes the direct application of meta-heuristic to the route generation.

Multi-objective transit route design is also seldom investigated. Generally, the two parts of the overall cost of providing transit service, operator cost and user cost conflict with each other. Pareto optimality-based algorithm is an effective method for the problem with conflicting objectives and can provide decision makers with a group of Pareto optimal solutions. This paper presents a tabu search-based solution approach to multi-objective transit route network design problem in such a way to provide decision maker with a group of Pareto optimal solutions.

2 Basic Concepts

The general multi-objective optimization problem is of the form

$$\begin{aligned} &max \quad \boldsymbol{f}\left(\vec{X}\right)=\left[f_1\left(\vec{X}\right),f_2\left(\vec{X}\right),\mathrm{L}\ f_M\left(\vec{X}\right)\right] \\ &Subject\ to \quad g_i\left(\vec{X}\right)\le 0 \quad i=1,2,\mathrm{L}\ T \end{aligned} \tag{1}$$

Where $\vec{X}=\left(x_1,x_2\mathrm{L}\ x_n\right)^T$, $\vec{X}\in\Theta\in R^n$, f_i is objective function and g_i is constraint.

The following basic concepts are used in multi-objective optimization case.

Definition 1: Let a solution $\vec{X}^0\in\Theta$

1. $\vec{X}^0$ is said to dominate a solution $\vec{X}^1$ ($\vec{X}^0$ f $\vec{X}^1$) iff $f_i(\vec{X}^0)\ldots f_i(\vec{X}^1), i=1,2,\mathrm{L}\ M$

2. $\vec{X}^0$ is said to be Pareto optimal iff $\neg\exists\vec{X}^1:\vec{X}^1$ f $\vec{X}^0$

3. Non-dominated, a decision vector which is not dominated by any decision vectors in a given set is non-dominated regarding this set.

3 **Problem Formulation**

Transit route network design is an important planning step in the urban transit planning procedure (Ceder and Wilson 1986). The important components of TRND are identified as estimating demand, identification of objective function, constraint, passenger behavior, solution techniques and the computational time. The main aim in TRND is to determine a route configuration consisting of a set of transit routes and associated frequencies that archives the maximum or minimum value of the desired objective, subject to the constraints. The constraints consist of the minimum and maximum available frequency, maximum load factor, allowable fleet size et al. The objective of the single objective TRND often is to minimize the total cost measure covering user cost as well as operator cost.

In this paper, multi-objective TRND is formulated as a two-objective problem. The mathematic formulation is shown as follows.

$$\begin{aligned} &\textbf{min} \quad OC = \frac{2C_v}{V}\sum_{k=1}^{R}\frac{d_k}{h_k} \\ &\textbf{min} \quad UC = \frac{\gamma_v}{V}\sum_{i=1}^{m}\sum_{j=1}^{m}q_{ij}D_{ij} + \frac{\gamma_w}{2}\sum_{k=1}^{R}\sum_{i=1}^{m}\sum_{j=1}^{m}q_{ij}\alpha_{ijk}h_k \\ &\textit{subject to} \quad h_k \le h_{max} \quad \forall k \in R \\ &\qquad\qquad\quad l_k \le L \quad \forall k \in R \end{aligned} \tag{2}$$

Where $k =$ route index; $l_k =$ load factor on route k. $d_k =$ length of route k; $m =$ the total number of nodes; $h_k =$ operating headway of route k; $R =$ total number of routes in the network; $V =$ speed of buses on the network; and $C_v =$ per-hour operating cost of buses; $D_{ij} =$ length of the shortest path between origin i and destination j. In this paper, the shortest path between origin i and destination j is the one with the minimum transfer times. If two paths have the same minimum transfer times, the path with the shortest length is the shortest one. Origin-destination demand is represented as a symmetric matrix $Q = \{q_{ij}\}$, q_{ij} is the demand between origin i and destination j; $\gamma_v =$ user's value of in-vehicle time; $\gamma_w =$ user's value of waiting time; $h_{max} =$ maximum headway required for any route; $L =$ load factor.

$$\alpha_{ijk} = \begin{cases} 1 & \textit{if OD flow}(i,j)\ \textit{begins its journey on route k} \\ & \textit{or if OD flow}(i,j)\ \textit{transfers to route k in an} \\ & \textit{uncoordinate fashion} \\ 0 & \textit{otherwise} \end{cases} \tag{3}$$

The first objective is the operating cost of the transit route network and the second is the user cost. The first term in the user cost formula is the user in-vehicle cost and the second is the user waiting cost. The first constraint is the headway feasibility constraint, and the second is the load factor constraint, which guarantees that the maximum flow on the critical link of any route cannot exceed the bus capacity on that route.

4 Tabu Search

Tabu search (TS) is a stochastic neighborhood search algorithm that was first suggested and applied by Glover (1989). The basic tabu search algorithm operates in the following way. It starts from a randomly selected or a known feasible solution *S*. From this initial solution, a set of neighborhood solutions, *S**, is generated using a number of previously determined movement strategies and the best neighbor replaces the current solution, even though it may be worse than *S*. The algorithm iterates, repeating the procedure with the new solution, until certain given stopping conditions are reached. However, the algorithm may recycle. To avoid this situation, certain attributes of the last replaced

solutions are stored in a list, which is called the tabu list. The neighbors of the current solution that satisfy conditions given by the tabu list are systematically eliminated unless they meet an aspiration criterion, so at each iteration the algorithm is forced to select a point not recently selected.

4.1 Tabu search-based Algorithm

The proposed algorithm consists of two main phases: an initial route network is generated in the first phase and the initial solution is improved in the second phase. The tabu search produces new neighbor solutions through three movement strategies: route-merge operator, route-break operator and add-link operator. The movement strategy is always performed on the current solution and is used to change the configuration of the current route network. In this paper, the tabu list is implemented as a circular queue with the first in first out queue principle, and three tabu lists are available and respectively constructed for three operators. The proposed algorithm consists of the following steps.

1 Generate an initial solution as the current solution with a route generation algorithm.

2 Generate a random number s following the uniform distribution on interval $[0,1]$, if $s < p_{rm}$, perform route-merge and add-link operator; otherwise, perform route-break operator. When an operator is performed, the external archive is maintained, if the new solution can be included into the external archive, replace the current configuration with the new one and update the corresponding tabu list.

3 If the termination condition is met, terminate search; otherwise, go to 2.

Where p_{rm} is the probability of route-merge operator.

4.2 Initial Solution

A route generation algorithm is applied to construct an initial route configuration with the constraints that all nodes have to be served. When constructing the initial route network, the demand is ignored. We define the connected nodes of a node by a single link as the adjacent nodes of the node.

The route generation algorithm can described as follows.

1 All adjacent nodes of each node are determined and stored in the adjacent set. Each node is also included into its adjacent set.

3 A route is built for node i in the following way: the first node randomly chosen from the adjacent set Θ (except node i) becomes the first node of the route; then all the nodes in the set Θ are chosen if they are connected to the end of the route by a single link; 1) if more than one node is chosen, a node is selected from all chosen nodes except the node i and turns into the new end; 2) if no node is chosen or only node i itself is chosen and at least two nodes exist in the set, a adjacent node of the end acts as the new end of the route if the adjacent node is connected a node in the set Θ of node i, or node i becomes the new end if no adjacent node is connected any nodes of the set Θ; 3) otherwise, the unique node in the set Θ is added to the end of the route. Once a node in the adjacent set is on the route, the node is removed from the set. Repeat the above procedure until the adjacent set becomes empty or no node can chosen as the new end of the route.

3 When a route is built for each node, the whole construction procedure is terminated.

The adjacent node of the current end of the route becomes the new end provided that this adjacent node isn't contained in the adjacent set of node i and the route, and is connected at least a node of the

adjacent set by a single link. As shown in the above procedure, node i must be included on the constructed route for it because this node connects with all nodes in its adjacent set. Moreover, node i is always chosen last because this method can make as many nodes as possible to be included on the route.

4.3 Operators

The route-merge operator tries to combine two routes with common nodes and have them perform in the following way.

1 All routes out of the merging-list are determined and each of them is assigned a fitness equal to $c_{max} - q_k^{cr}$ .where c_{max} is a big enough positive number. A route l is chosen by using the roulette-wheel method.

2 The following conditions are used to construct all possible combined routes: 1). If all nodes of a route k are included in the route l, delete the route k and make the route l be the combined route. 2). If a node i of a route k is the beginning or the end of the route l and all nodes from the beginning of the route k to the node i are also on the route l, then merge the route k and l into a combined route in the following way: the first part of the combined route is built by including all nodes of route k from the end to the beginning; the second part is formed by adding all nodes of the route l from beginning to end if node i is the beginning or constructed by adding all nodes of the route l from end to beginning if node i is the end. The common nodes are not included on the second part of the new route. 3). If a node of the route l is the beginning or the end of a route k, the above procedure is used to build possible combined routes by interchanging route k and l each other.

3 A combined route is chosen if the demand between two randomly chosen nodes can be served by the new network by substituting the combined route for two routes being used to form the new route. If the objective-function value of the new network is smaller than the current value, the new network replaces the current one, add the new route into the merging-list and the breaking-list and update these lists.

The route-merge operator defined in Ngamchai et al. (2006) only connects routes that have a terminus in common. The proposed route-merge operator will try to connect two routes that a node of one route is the beginning or the end of other route. It is obvious that the candidate routes for the latter are significantly more than those routes for the former.

The route-break operator will try to split a route into two routes. First, a route k , which consists of at least 4 nodes and is out of the breaking-list, is randomly chosen. If the total number of routes in the current network is lower than a predetermined value, break route k into two different routes; otherwise, route-break is aborted. If the objective function of the new network obtained by route-break is lower than the current objective value, replace the current network with the new one, add the new routes into the merging-list and breaking-list and delete the tail of these lists if they become full.

The route-break operator defined by Ngamchai et al. (2006) may split a route into more than two routes. The proposed operator always randomly separates a route into two different routes.

In the proposed algorithm, when the route-merge operator or the route-break operator are performed for some routes, the new routes are included into both merging-list and breaking-list. The

main purpose of this strategy is to avoid a situation in which a new combined route obtained by route-merge in this iteration is divided into two routes through route-break in next iteration or vice versa.

Routes for merging and a route for splitting are randomly determined provided that these routes are out of two tabu lists. Many routes can be chosen and merged with the other route; as a result, a number of the combined routes are produced. The splitting of a route is not based on difference between the least and the second least link flows and completed at random. This feature also guarantees that the route-break operator also can produce various different transit route configurations. Thus, the proposed movement strategies are effective and feasible.

The add-link operator will try to add some links to the route with small length. First, all routes out of the adding-list are determined and each of them is assigned a fitness equal to $m-\left|\Omega_k\right|$.where $\left|\Omega_k\right|$ is the number of links on route k. A route l is chosen by using the roulette-wheel method. The beginning or end of the route l is considered and a new beginning or end is decided. Finally, if the new network has a better performance than the old one, replace the old one with it and include it in the adding-list.

4.4 External Archive maintenance

The external archive is a basic concept of multi-objective optimization. It is a set of the non-dominated solutions and is used to store the non-dominated solutions produced by the multi-objective optimization algorithms. In most cases, a maximum size is specified for the external archive. In this paper, the archive is maintained in the following way:

For each new solution, if it dominates some archive members, replace these dominated members with it; if it is non-dominated regarding the whole archive and the archive size is lower than the predetermined value, include it into the archive.

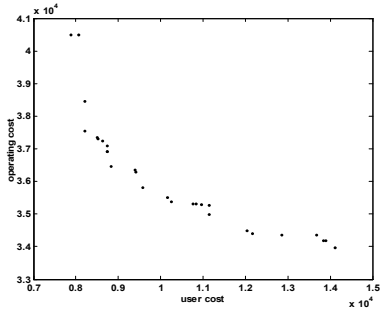


Fig. 1 the distribution of the final non-dominated solutions

5 Numerical results

The proposed algorithm is validated on a sample network. The example network is defined by 25 nodes and 39 links and the corresponding link data, node data and demand date are available in Ngamchai et al. (2006). The following parameters are used in the performance validation: bus operating

cost =\$55/h; load factor =1.2; bus capacity =50 passengers/bus; bus travel speed = 40km/h; user in-vehicle time value = \$8.5/h; user waiting time value =\$17/h.

The sensitivity analyses of the proposed algorithm are performed when the probability of route-merge operator p_{rm} varies from 0.1 to 0.9 and 10 different runs are conducted for each case. The maximum archive size is 20. The maximum objective function evaluation times of 1200 are the termination condition. Table 1 shows the final non-dominated solutions regarding all archive members produced by the proposed algorithm in 90 runs. These solutions in table 1 are the best ones which the proposed model can obtain. The data in each unit grid are $\boldsymbol{OC}$, $\boldsymbol{UC}$, the number of routes and fleet size. The passengers will transfer 2 times at most for each route configuration found by the proposed algorithm. Fig. 1 shows the distribution of the final non-dominated solutions on the objective space, which the horizontal coordinate is the user cost and the vertical coordinate is the operating cost. Fig. 3 computational results on Generational distance

As can be seen, the proposed algorithm obtains the total 29 non-dominated solutions with different combinations of two objectives. These solutions don't concentrate on a narrow region of objective space and distribute on a wide region with good uniformity. The features of the final solutions are beneficial to the final decision of the decision makers.

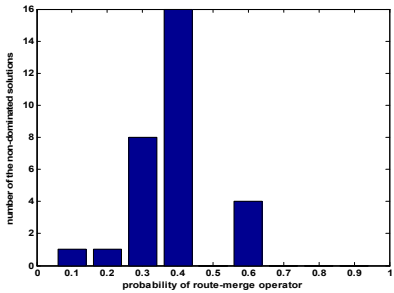


Fig. 2 The number of the final non-dominated solution in each case

Figure 2 shows how many of the final non-dominated solutions are produced in each case. In this paper, the generational distance GD is used to measure how far the archive members obtained in each case are close to the solutions in table 1.

$$\boldsymbol{GD} = \frac{\sqrt{\sum_{i=1}^{n} \boldsymbol{d}_i}}{\boldsymbol{n}} \tag{4}$$

Where $\boldsymbol{n}$ is the number of archive members and $\boldsymbol{d}_i$ is the distance between each archive member and the nearest final non-dominated solutions.

Fig.3 shows the generational distance values of the proposed algorithm with different probability of route-merge operator. As can be seen, 15 of 29 final non-dominated solutions and the least

generational distance value are achieved at 0.4. Eight and four non-dominated solutions are respectively obtained when p_{rm} is respectively equal to 0.3 and 0.6. All archive members are dominated by the final non-dominated solutions and these members are far away from the final non-dominated solutions when p_{rm} is assigned values such as 0.5, 0.8, 0.9. Therefore, 0.4 is chosen as the optimal probability of the route-merge operator.

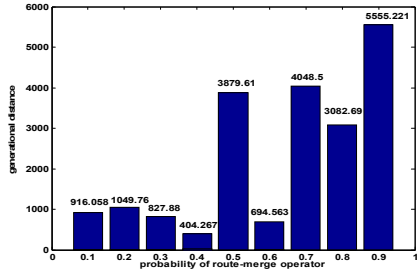


Fig.3 computational results on Generational distance

Table 1 The final non-dominated solutions obtained in 90 runs

7880.46,40488 9,144	9405.93,36342.9 10,172	12041.9,34487.3 9,219	8519.31,37300 12,155	8218.39,37540.8 13,150
8061.99,40480 9,147	9423.49,36292.9 11,172	12165.9,34398.7 10,222	8737.85,37091.2 12,159	8741.48,36921.2 10,159
8059.45,40482.2 10,147	10247.9,35375.5 10,187	13672.8,34359.4 9,249	10167,35507.4 10,185	8508.12,37353.7 13,155
11141.7,35256.7 9, 203	11146.7,34976.5 9,203	14115.4,33971.5 8,257	10841.4,35300.2 11,198	8203.77,38433.9 12,150
8835.13,36452.3 10,161	10960.6,35289 9,200	13902.4,34174.4 7,253	12852.9,34365.4 10,234	8743.6,36904.4 11,159
9580.09,35798.7 9,175	10758.7,35310 10,196	13849.8,34176.3 7,252	8623.93,37228.9 12,157	

6 Conclusions

A tabu search with three operators is used to solve the multi-objective transit route network design problem in this paper. The proposed algorithm is composed of two phases. A good initial solution is first produced by a route generation algorithm and then is improved through route-merge, route-break and add-link operator. An external archive is used to store the non-dominated solutions produced by the tabu search. Once an operator is performed, the archive is maintained by using the new solution. If the new solution can become the archive member, it substitutes for the current solution. Three tabu lists are

available and respectively constructed for three operators. The sensitivity analysis are presented and show that a tabu search is a simple and effective method to multi-objective transit route network design.

Acknowledgements
This work is sponsored by National Basic Research Program of China (2005CB724205）

References

Baaj, M. H., and Mahmassani, H. S. (1991). ‘‘An AI-based approach for transit route system planning and design.’’ *J. Adv. Transp.,* 25 (2), 187–210.

Pattnaik, S. B., Mohan, S., and Tom, V. M. (1998). ‘‘Urban bus transit route network design using genetic algorithm.’’ *J. Transp. Eng.,*124 (4), 368–375.

Lee, Y. J., and Vuchic, V. R. (2000). Transit network design with variable demand, *Proc., 79th Annual Meeting of the Transportation Research Board*, Transportation Research Board, Washington, D.C.

Tom, V.M., and Mohan, S. (2003), “Transit route network design using frequency coded genetic algorithm.” *J. Transp. Eng.,*129 (2), 186–195

Fan, W., and Machemehl, R.B. (2006), “Using a simulated annealing algorithm to solve the transit route network design problem.” *J. Transp. Eng.,*132 (2), 122-132

Ngamchai, S., and Lovell, D.J. (2003), “Optimal time transfer in bus transit route network design using a genetic algorithm.” *J. Transp. Eng.,*129 (5), 510-521

Agrawal, J., and Mathew, T.V. (2004), “Transit route network design using parallel genetic algorithm.” *J. Comput. Civ. Eng.,*18 (3), 248-256

Fan, W., and Machemehl, R.B. (2006), “Optimal transit route network design problem with variable demand: genetic algorithm approach.” *J. Transp. Eng.,*132 (1), 40-51

Ceder, A., and Wilson, N. H. M. (1986), ‘‘Bus network design.’’ *Transp. Res., Part B: Methodol,* 20 (1), 331–344.
Glover, F. (1989), “Tabu Search, Part I,” *ORSA Journal on Computing*, 1, 190-206.

Dynamic Model for Evaluation of Advanced Traveler Information System

Chao YANG[1], James Luk[2]

[1] School of Transportation Engineering, Tongji University, Shanghai, 200092 China
PH: 86-21-65985033; fax: 86-21-65982274; e-mail: tongjiyc@mail.tongji.edu.cn
[2] Australia Road Research Board Group, Australia. e-mail: james.luk@arrb.com.au

Abstract: This paper describes a dynamic model for evaluating the benefit of an Advanced Traveler Information System (ATIS). The model consists of five modules, including a route choice module and a traffic simulation module. The route choice module considers different categories of drivers' route choice behaviors with different levels of access to traffic information. The traffic simulation module simulates dynamic network conditions in response to traffic information. The dynamic model allows the investigation of a network-wide performance in a road network under different traveler information systems. Some results are provided to demonstrate the use of the model. They suggest that providing traffic information can reduce the total network delay. However, there is an optimum level of market penetration beyond which real-time information can no longer reduce delay of the network and the individual.

Key words: *Advance Traveler Information System, Evaluation, Dynamic Model, Traffic Simulation*

1 Introduction

There are many forms of an Advanced Traveler Information System (ATIS), such as traffic advisory radio, traffic information via the Internet, traffic information from roadside infrastructure, in-vehicle route guidance systems, and variable message signs (VMS). The Highway Advisory Radio in the US and the Radio Data System - Traffic Message Channel (RDS-TMS) in Europe are some examples of this kind of traffic advisory radio. Many transport authorities set up websites to provide travel speeds of critical links in the network and other traffic information, such as incidents, road work and vehicle breakdowns. Examples of these web-sites include the Traffic Smart site in Singapore (http://www.onemotoring.com.sg/) and the real-time traffic conditions status report in Shenzhen (see website: http://www.sutpc.com). Traffic information from roadside infrastructure (infrared beacons) is used in Europe and Japan. Beacons are installed at specific sites and communicate with cars passing by. This system uses cars as probes to report journey times in the network. There are on-board route guidance systems - drivers can obtain real-time information via in-car display units which receive data from information service providers. These systems recommend optimal routes to drivers based on real-time information. An example is the service tested by Karl and Trayford in Melbourne. VMS systems are normally installed on

various segments of a freeway to provide travel times or relative congestion levels. The information provided by VMS can be in text form, graphics or a combination of both (e.g. Drive Time in Melbourne).

Little is known about the network-wide effect of ATIS in terms of improving the conditions in a traffic system. In the last decade, there has been active research in time-dependent or dynamic traffic assignment due to advancements in ITS. The research can be classified into two broad types: mathematical models and simulation-based approaches. Mathematical models are naturally based on analytical formulae. In order to solve a problem, assumptions and simplifications of the problem have to be made. For example, Lam and Chan adopted the traditional static user-optimal assignment and extended it into a time-dependent framework. This analytical model, however, is suitable only for freeway networks because intersections are not modeled in it.

On the other hand, the simulation approaches such as CONTRAM, INTEGRATION and DYNASMART have been used in a few applications. In these simulation-based approaches, individual vehicles are tracked as they move through a network while their velocities are determined by macroscopic speed-flow relationships.

Originally developed by the Transport and Road Research Laboratory (TRRL) in the late 1970's, CONTRAM is a dynamic assignment model that predicts traffic routes, link flows, queues, and delays at intersections as traffic movements evolve over time. It has the capability to represent time-varying network conditions. It uses an iterative procedure to establish the equilibrium distribution of traffic in a network. In each iteration, the assignment program assigns packets to their minimum journey time routes in the network. In this model, all drivers are assumed to have the same route choice behavior except fixed-route vehicles. However, CONTRAM does not have the ability to model incidents and advanced traveler information systems.

Van Aerde and Yagar proposed a comprehensive model called INTEGRATION, which can evaluate networks that incorporate expressways, arterial roads and other surface roads. At present, INTEGRATION can not model the driver behavior of VMS and pre-trip information systems.

DYNASMART uses a simulation-assignment model in the context of the ATIS and Advanced Traffic Management System (ATMS) context. DYNASMART builds its own traffic simulation model, especially the intersection model.

Recently, Tong and Wong proposed a method for determining the dynamic equilibrium in a road network with three different classes of drivers. Each class receives a different level of information. By varying the percentages of drivers in each class, the smallest total travel time in the network can be found. Their approach, however, does not include an intersection model to account for intersection delays.

This paper provides one approach that accepts a time-dependent origin-destination trip matrix. The interaction between the drivers and the traffic

information provided is modeled in a dynamic framework. The aims of this paper are to

(1) Provide background information on ATIS and the related modeling research;
(2) Outline the modeling framework proposed in this study;
(3) Describe some results of simulating the impact of ATIS on the road network of Singapore.

2 Modeling framework

In the modeling of ATIS, the framework, the traffic simulator and the route choice model are most important. In urban areas, a detailed intersection simulation model is necessary for accurate results. This study makes use of the key concepts reported in the literature. In particular, it employs the transport-planning package, TRIPS, to provide intersection modeling.

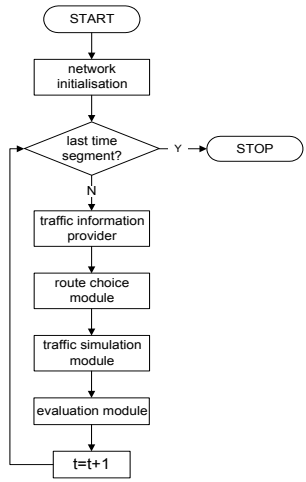


Figure 1. Framework of the dynamic simulation of ATIS

A dynamic model for evaluating the network performance of ATIS is shown in Figure 1. The framework consists of five modules:

(1)Network initialization - This module prepares the initial network condition for the algorithm. A time-dependent origin-destination (OD) matrix is an input to the module. In order to compare the network performance under different ATIS, the OD matrix is considered to be fixed. The consideration of elastic demand, which could be affected by traffic information, especially pre-trip information, is for future investigation;

(2)Traffic information provider - The travel time of each road link of the current

time segment (i.e. traffic information) is calculated in this module;

(3)Route choice - According to the traffic information and four categories of choice behaviour described below, this module determines the routes for different categories of drivers at each time segment;

(4)Traffic simulation module - Link flows of a network are changed in this module according to drivers' route choices. The network condition for the subsequent time segment is then calculated based on the updated flows of the network;

(5)Evaluation module - The performance indices, such as network delay and link-specific delay and flow rate, are calculated in this module.

The study period is divided into time segments; each time segment represents 5 min. The framework therefore simulates the period from the beginning of loading the traffic of the first time segment to the time segment when all traffic has been loaded onto the network.

After each pass of the algorithm, the time frame moves forward to the next time segment. The procedure is carried out for every time segment until all the time-dependent OD matrices are loaded onto the network. In the context of modeling ATIS, it is reasonable to assume that vehicles currently assigned would not be influenced by vehicles to be assigned in the future, because the currently assigned vehicles would probably be out of the network by that time. The traffic information provided to drivers is only based on the current or instantaneous network conditions and not on future or predictive conditions.

Currently, the initial network for the first time segment is set to have no traffic. Other initial conditions can be easily employed in the framework.

Of the five modules in the framework, the route choice module and the traffic simulation module are the most important. These two modules are further described below.

A. Route choice module

In this module, drivers are classified into the following four categories according to the level of access to traffic information:

(1)Drivers with no driver information - Drivers of this category are assumed to have no traffic information. Drivers choose the shortest route according to the travel distance when they enter the network. After that, they would keep on this route until they arrive at their destination. This category forms the benchmark for comparison with other information categories.

(2)Drivers with pre-trip traffic information - Drivers with pre-trip traffic information can get the prevailing network conditions, i.e. the instantaneous travel times of links in the network at the time they begin to travel. After they choose the shortest route according to the instantaneous route travel times, they would keep on this route until they arrive at their destinations.

(3)Drivers with real-time information - To determine the behavior of drivers with full information access, the concept of Reactive Dynamic User Optimal

(RDUO) is used. The drivers are assumed to choose their route at time segment t by the reactive or instantaneous route travel time. This travel time is the sum of the travel times on the subsequent links comprising the route at time segment t. The real-time information relayed to drivers is the link travel times that are used to compute the instantaneous route travel time. Drivers receive the real-time information while driving and are supposed to have the ability to deal with the information they get and make a correct choice of the minimum time route.

(4) Drivers driving past VMS signs - VMS's are different from other kinds of ATIS. They provide traffic information on specific links at limited sites only. Drivers passing by these sites can get the information and use this information to choose the shortest time route. Research on site-specific information systems is continuing and this paper does not include results for this category of drivers.

Mahmassani and Jayakrishnan suggested that a driver's route choice behavior exhibits a 'boundedly-rational character'. Under this behavioral assumption, a driver would divert from the current route only if the improvement in the remaining travel time after re-routing exceeds a threshold level. The behavior could be represented by a binary indicator variable, δ, as follows:

$$\delta = \begin{cases} 1 & if\ T_c - T_o > \alpha T_c \\ 0 & otherwise \end{cases} \qquad (1)$$

Where T_c is the travel time of the current route; T_o is the travel time of the optimal alternative route; α is a threshold level representing the willingness to divert.

When δ equals to 1, a driver switches from the current path to the best alternative, and when δ equals 0 the current path is maintained.

The threshold level represents the driver's willingness to divert. Low values of α correspond to the driver's increasing sensitivity to current conditions, resulting in considerable diversion activities. At α = 0, drivers will switch routes if the alternative route offers a shorter travel time no matter how small the difference is. At α = 1, drivers will be least sensitive to the shorter travel time on an alternative route. As drivers in category (a) and (b) do not switch en route, only drivers in category (c) and (d) are modeled with this behavior in the current research. A value of α = 0.1 is adopted in this study to represent the situation when category (c) drivers are willing to switch routes if the percentage time is different, (Tc – To)/Tc, is 10 %.

At each time segment, the route choice module calculates the routes for different categories of drivers based on the drivers' behavior and the information from the traffic information provider module. For example, for a category (c) driver, at each time segment the driver receives the travel times of all links in the network. He or she compares the travel time of the shortest-time route and the travel time of the current

route from the current link to his or her destination. The driver then makes a decision to divert or not according to the 'boundedly-rational character'.

The proposed framework also quantifies the level of market penetration or subscription to ATIS. A 10% penetration means that 10% of drivers subscribe to a certain type of information system, either the pre-trip (category (b)) or the real-time (category (c)) system. The other 90% continue to remain as category (a) drivers - choosing routes according to the shortest distance routes.

In this framework, the traffic information received by all drivers is identical and corresponds to the simulated network conditions. Information quality and transmission delays are not considered.

B. Traffic simulation module

(1)Representation of traffic

The traffic flows generated from various origins of a network at different time segments are grouped into packets. Each packet consists of a number of vehicles. It is assigned independently to its route through the network based on the route calculated in the route choice module. For a packet generated at a time segment, it is added to links along the route from its origin to the farthest link that it can reach in this time segment according to the travel times of the links. For a packet already in the network, the module moves it through the network based on its route travel time. During each time segment, the value of flow is the average flow in that time segment.

The link travel time is assumed to consist of two components: the cruise time along the link and the delay at the downstream intersection. The commercial transportation software, TRIPS, is employed to model link travel time. The intersection model of TRIPS adopts the approach used in the TRRL (now the Transport Research Laboratory, TRL) intersection programs ARCADY, PICADY and OSCADY.

(2)Speed-flow function

The link cruise time is calculated from a link flow and the link speed-flow function. For different link types, speed-flow functions with different shapes are used. Up to ten points can be input to define the speed-flow function in TRIPS, followed by a continuous function after the point with the largest degree of saturation is specified.

(3)Intersection delay

For undersaturated conditions, only delays would be calculated. For oversaturated cases, queues caused by oversaturation are considered in the model. Intersection delays are calculated using the time-dependent queuing theory, which allows delays at oversaturated intersections to be calculated. This makes it more suitable to simulate network conditions with a time-dependent OD matrix. The initial queue length of the current time segment is the final queue length of the previous time segment. At this stage, a point queue concept (vertical queue without length) is used. In this model, 'block-back' of a link is considered. Blocking-back occurs when the queue on a link

reaches a sufficient length to prevent vehicles entering the link from an upstream intersection. TRIPS assumes that the degree of blocking is a function of the mean queue-length. Each link has a Link Storage Capacity (LSC) based on the length of a link. With LSC and the mean queue-length, the model determines the percentage blocking of the current time segment. At the next time segment, the capacity of turning from an upstream link to a blocked link would be modified according to the blocking percentage of the downstream blocked link.

As already mentioned, the flow of each time segment is the average flow. The link travel time is also the average travel time in a time segment. The module calculates the average delay of all vehicles in each time interval of 5 minutes. If the duration of a time interval is less than the cycle time of the intersection, this average delay would be sensitive to the start time of the time segment. In this study, the time interval is set to 5 min to avoid this problem.

3 The simulation study

The network is described in terms of origins, destinations, links and intersections. Figure 2 shows the network used in this study. The numbers in the network are origins and destinations. The network is based on the full expressway network and most parts of the arterial road network in Singapore. It consists of 15×14 OD pairs, 287 nodes and 974 links.

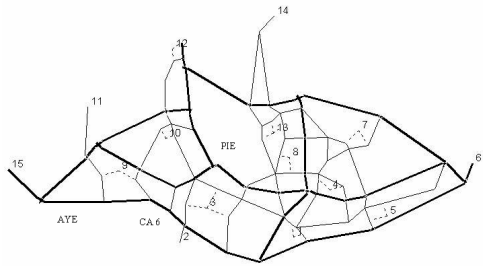


Figure 2. Experimental network of Singapore

The OD matrix is estimated using TRIPS based on an existing partial OD matrix and traffic counts of some links at the peak hour. Each cell of the matrix represents the average flow rate for each OD pair. To model the time-dependent OD demand, a flow profile is employed (see Figure 3). If the average flow rate of the origin at a peak hour is 1,000 pcu/h, the flow rate for time segment 2, as shown in the Figure 3, is 93% of the average flow rate or 930 pcu/h. Demands from one origin to all destinations share the same profile curve. One profile curve is input for each origin.

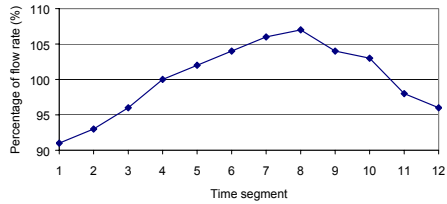


Figure 3. Time-dependent traffic demand profile from an origin

In this study, the estimated OD matrix represents high (or 100%) demand. Another two levels of demand are also analyzed: low (20%) and medium (60%) demand to represent different levels of congestion.

An analysis of the impacts of different levels of market penetration is conducted. The different levels to be analyzed are between 10% and 50% in steps of 10%. For 10% market penetration level, it is assumed 10% of the drivers in a network subscribed to the driver information system (pre-trip or real-time information) and the rest (90%) are drivers without information who choose shortest-distance routes.

For comparison, the scenario of all drivers without information is set to be the benchmark. Combining six market penetration levels with three demand conditions, 18 different scenarios for each type of ATIS are constructed. The performance index for comparison is the total delay rate of the network.

4 Simulation results

Results are obtained from five levels of market penetration (10, 20, 30, 40 and 50 per cent) and the benchmark (0% of market penetration) for three cases of demand was investigated. Figure 4 shows the results of total network delay rate for different levels of market penetration under the low (20%) demand. The total network delay rate (TD) is determined from:

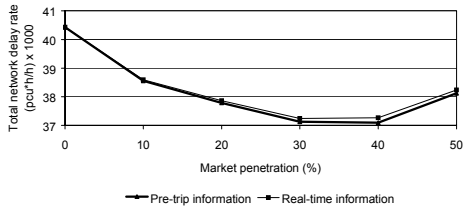


Figure 4. Total network delay rate at low demand (20%)

$$TD=\sum_{k=1}^{K}\sum_{a\in A} q_a(k)d_a(k) \qquad (2)$$

$$d_a(k) = c_a(k) - c_a^0 + t_a(k) \qquad (3)$$

where $q_a(k)$ is the flow rate of link a at time segment k (pcu/h); k is the time segment index, from 1 to K (=12); A is the set of links in the network; $d_a(k)$ is the delay of link a at time segment k (h); $c_a(k)$ is the travel time of link a at time segment k (h); c_a^0 is the free flow travel time of link a (h); $t_a(k)$ is the delay of link a at the downstream intersection at time segment k (h).

The two curves in Figure 4 represent the performance of the pre-trip information system and the real-time information system respectively. Category (a) drivers without information (i.e. at 0% penetration level) provide the benchmark for comparison. It is further assumed that drivers are 100% compliant to driver information. As already mentioned, category (c) drivers are further constrained by the 'boundedly-rational character' with a threshold level (α) equal to 0.1.

Figure 4 shows that the network delay decreases while the level of market penetration increases from 0% to 40% for both pre-trip and real-time information. The performance of real-time information system could not be better than pre-trip information system, because there is no congestion in the network at low demand. Minor route diversions due to real-time information have also created unnecessary travel for category (c) drivers leading to slightly higher network delay.

Figure 5 shows the results of the total network delay rate for different levels of market penetration under medium demand. The network performance of scenarios with real-time information is worse than the scenario without traffic information when the level of market penetration exceeds 40%. At 30% market penetration, both ATIS perform at their best, with the real-time information system outperforming the pre-trip information system. Drivers with real-time information benefit from switching en-route as enough capacity exists in the network at this demand condition.

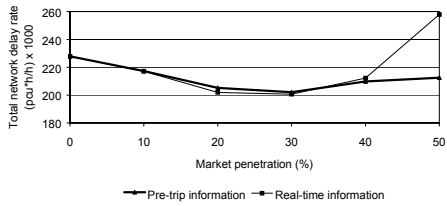


Figure5. Total network delay rate at medium demand (60%)

Figure 6 illustrates the results of the total network delay rate for different levels of market penetration under high demand. The optimal market penetration for the best performance of both systems is 20% at high demand. When the market penetration increases to 40%, neither system could perform better than the benchmark scenario of no information. Drivers receiving real-time information have more chances to switch en route than drivers with pre-trip information. This could lead to the phenomenon of overreaction. Because drivers are assumed to make decisions at each time segment (5 min) according to prevailing network conditions instead of predictive traffic information, the new route that drivers select does not necessarily produce less delay. These results suggest that providing real-time information to a large portion of drivers in the network may not give the better outcome for the system and the individual driver.

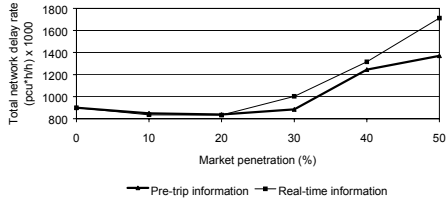


Figure6. Total network delay rate at high demand (100%)

From these results, the following points can be made:

(1)A driver information system is most effective at low demand. It can provide a potential benefit of up to 7.5% in reducing the total network delay at 30% and 40% of market penetration (using either type of ATIS);

(2)There is an optimal level of market penetration for each demand, i.e. 40% for low demand, 30% for medium demand and 20% for high demand. The optimal level of market penetration decreases while the demand increases;

(3)For all demand conditions, the performance of a real-time information system is better than or equal to that of the pre-trip information system at the optimal levels of market penetration, but worse at levels of market penetration higher than 50%.

5 Conclusions

This paper has provided a dynamic model of the impact of ATIS. Using an abridged network of Singapore as an example, it was shown that the model could be used as an evaluation tool to capture the dynamic interaction of the network performance with the driver information provided. The framework provided here is not an analytical approach, but a direct simulation of real-world ATIS. The results

highlight the complexity of the effects of information on the network performance as well as the drivers' route choice behavior.

The results obtained confirm that ATIS have a potential role in managing demand to match the capacity available. They could potentially alleviate acute problems caused by roadwork and accidents, and contribute to satisfactory performance of networks operating close to capacity over extended periods of high and variable demands. Future research will be on the modeling of accidents using the proposed dynamic framework.

References

A.C., Karl and R.S., Trayford. (2000). "Delivery of real-time and predictive travel time information: experiences from a Melbourne trial," *Proc. 7th World Congress on ITS, November 6-9, Turin.*

C.O., Tong, and S.C., Wong. (2000). "Evaluation of advanced traveller information systems using a multi-class dynamic traffic assignment model," *Proc. 6th Int. Conf. on Applications of Advanced Technologies in Transportation Engineering, June 28-30, Singapore*.

H.S., Mahmassani, and R., Jayakrishnan. (1991). "System performance and user response under real-time information in a congested traffic corridor," *Transpn Res*. vol. 25A (5), pp. 293-307.

H.S., Mahmassani, T.Y., Hu, S., Peeta and A., Ziliaskopoulos. (1994). "Development and testing of dynamic traffic assignment and simulation procedures for ATIS/ATMS applications", *Centre for Transportation Research, The University of Texas at Austin. Technical Report DTFH6 1-90-R-00074-FG*, Austin, TX, USA.

J. Li, O., Fujiwara and S., Kawakami. (2000). "A reactive dynamic user equilibrium model in network with queues," *Transpn Res.* Vol. 34B (8), pp. 605-624.

MVA Ltd. (1999) *TRIPS/32 – the professional's transportation planning package*. Woking Surrey, UK.

M., Van Aerde and S., Yagar. (1988). "Dynamic integrated freeway/traffic signal networks: a routing-based modelling approach," *Transpn Res*. 22A (6), pp. 445-453.

N.B., Taylor. (1990). "CONTRAM 5: An enhanced traffic assignment model," *Transport and Road Research Laboratory. TRRL Research Report RR249*, Crowthorne, UK.

W.H.K., Lam and K. S. Chan. (1996). "A stochastic traffic assignment model for road network with travel time information via Variable Message Signs," *Proceedings of the 1996 IEEE Intelligent Vehicles Symposium, September 19-20, Tokyo, Japan, pp. 99-104.*

Feasibilities and Challenges of Probe Technologies for Real-time Traffic Data Collection

Kai LIU[1], Toshiyuki YAMAMOTO [2] and Taka MORIKAWA [3]

[1] Researcher, Ph.D., Graduate School of Civil Engineering, Kobe University, 1-1, Rokkodai, Nada, Kobe 657-8501, JAPAN; Tel/Fax: +81-78-803-6346; E-mail: liukai@landscape.kobe-u.ac.jp

[2] Associate Professor, Department of Civil Engineering, Nagoya University, 1, Furo-cho, Chikusa, Nagoya, 464-8603 JAPAN; Tel: +81-52-789-4636; Fax: +81-52-789-5728; E-mail: yamamoto@civil.nagoya-u.ac.jp

[3] Professor, Department of Environmental Engineering and Architecture, Nagoya University, 1, Furo-cho, Chikusa, Nagoya, 464-8603 JAPAN; Tel: +81-52-789-3564; Fax: +81-52-789-5728; E-mail: morikawa@nagoya-u.jp

Abstract: Probe vehicles are playing an increasingly important role in the field of ITS, and showing high value for various applications over the last decade. Recent advances of probe vehicle technologies in Japan are explored by reviewing the existing studies in both academic fields and ITS practice projects. The methods of data collection and data processing are summarized, which is followed by a brief review of recent attentions and the possible solutions. In order to enhance the feasibility of probe vehicles for real-time traffic surveillance through the entire road network, this paper tries to identify some unresolved issues that may have impacts on the reliability of obtained real-time traffic data and the efficiency of data collection. Four addressed issues include: error propagation of map matching, the bias in link travel time estimation, lower accuracy of link travel time measurement through infrequent observations, and the efficiency in area coverage.

Keywords: *probe vehicles, Japan, attentions and issues*

1 Introduction

Cities around the world are now recognizing the need to pursue an effective method of monitoring traffic conditions to support real-time travel information systems and route guidance systems. Several data collection techniques can be used to

measure road segment travel times, among which probe vehicle equipped with GPS ranks highly because of its flexibility and cost-effectiveness within a large study area (Turner *et al.* 1998).

As compared with other infrastructure dependent technologies, such as loop detectors, probe vehicles have become an increasingly favorable tool to collect real-time traffic information due to its relative lower initial cost that can even be afforded by private commercial companies other than by government investments. Most importantly, these vehicles have the advantage of directly measuring roadway travel times. On the other hand, there are still some untapped issues of cost-effectiveness with respect to large scale probe deployment which might lead to a negative effect on the continuous data collection.

The main objective of this paper is to make a general review of existing studies and projects with respect to probe technologies, to understand the merits and limitations of such technologies in different environments, and to find what should be done and how to improve the cost-effectiveness for widely implementation in developing cities.

This paper first reviews the existing studies on the implementation projects of probe systems in Japan. Then, attentions focus on the introduction of the advances in the data collection and process, as well as the recent attentions and solutions in improving the cost-efficiency of probe data collection. This paper last discusses four issues that were neglected or have not attached enough attentions.

2 Probe Experiments in Japan

The initial probe experiment in Japan was recognized as at the end of 1980's when the AVI (Automated Vehicle Identification) system were taken into practical application and travel time information at limited road networks were provided. Then after 1996 the widely implemented VICS service that covered all highways and main arterials in Japan pushed probe technologies ahead. Since 1999, probe techniques through moving transmitters have been developed for road administration by the Japanese Ministry of Land, Infrastructure and Transport (MLIT) and for the traffic information surveillance by Japanese Automotive Research Institute (JARI) (Wada and Fujii, 2002). Those probe transmitters, sometimes called passive probes, include taxis, trucks, buses, private cars, and so on.

In addition, lots of researches have already tested the possibilities of mobile phones as a probe tool for travel behavior survey, including the early activity survey by PHS in Osaka in 1998 and 1999, the probe person experiment during World Cup 2002, the consumers' shop-around behavior experiment in Fukuoka in 2002, and the activity survey through GPS mobile phones in Matsuyama city (in 2003 and 2004),

Tokushima (in 2004), and Kochi (in 2004). Although traffic information is sometimes obtained from such kinds of data collection, those experiments did not aim for traffic surveillance and hence are not the attentions of this paper.

Japan has given a great impetus to the probe experiments and commercial development for at least one decade. The world witnessed the active probe implementation in Japan and recognized Japan as a leader in probe experiments on both taxis and private cars (Bishop, 2005). The Internet ITS Consortium is the major actor of such a great activity, which consists of government institutes, research institutes and universities, vehicle manufactures, cellular phone operating agencies, and so on. In 2000 and 2001, Japan MLIT and Internet ITS Consortium organized the probe experiment in Tokyo, Yokohama, and then a more large-scale probe field trial was scheduled in Nagoya.

P-DRGS (probe-based dynamic route guidance system) project in Nagoya, Japan that employed the IPv6 (Internet Protocol version 6) technology is a remarkable probe experiment in Japan with the largest scale and the longest periods up to now. Nagoya probe experiment consisted of 1,570 taxis with GPS receivers automatically reporting their locations thanks to the cooperation with 32 companies from Nagoya Taxi Association. Data collection was conducted in three stages: the first stage from January, 2002 to March, 2002; the second stage from October, 2002 to March, 2003; and the last stage from October, 2003 to June, 2004. To date, P-DRGS consortium has successfully developed a web-based route guidance system that can be visited through computers to provide multimodal traffic information and route guidance (Morikawa et al. 2007), which include the shortest travel time route (for the browser's own and for all travelers in the networks), the least CO_2 emission route (for the browser's own and from the aspect of all travelers in the networks), and so on. Here a data fusion, including a probe-based historical database, real time traffic information from VICS, real time weather information, and so on, is carried out for improving the system reliability. Meanwhile, a lot of application researches on the utilization of the collected probe vehicles data have been undertaken.

The probe taxis in Nagoya can be grouped into three types according to data transmission types and the in-vehicle map matching capability of the probe vehicles. The critical difference is that the GPS data submitted by type II probe vehicles are already map-matched owing to an enhanced GPS receiver that was integrated with an in-vehicle navigation device and can efficiently reduce the rate of missing GPS records, which is common in city downtown area due to the block of tall buildings. Moreover, seven patterns of data transmission schemes according to a pre-set polling interval are deployed during the second stage to send real-time information back to the information center. Table 1 lists the detailed description of all these three types and the corresponding seven patterns.

Table1. Probe types and polling intervals in Nagoya probe experiment *

Type	Polling intervals	Probe car numbers	Data transmission devices and in-vehicle navigation devices
Type I	300m	810	Packet communication device; Without in-vehicle navigation device.
Type III	300m+events#	35	DSRC (Dedicated Short Range Communication); Without in-vehicle navigation device.
Type II	10s	35	Packet communication device; and in-vehicle navigation device
	5s	10	
	100m	310	
	50m	100	
	100m+events#	200	

* During the second stage of probe data collection.

The events here include short stops (SS) and short trips (ST), please refer to Wada and Fujii (2002) and Horiguchi (2004).

As compared with the general collection method that saved probe data in a hard disk placed in a vehicle when vehicle was moving on the network and then sent to the information center periodically (one week or one day), probe data in Nagoya experiment are directly transmitted to the information center at real-time to be converted into traffic information. A mobile electronic communications network is employed to transmit these data with a communication charge depending on the amount of packets. For the purpose of real time data transmission, frequent exchanges are needed. Table 2 lists the charges for data at four kinds of polling intervals. As a result, almost all probe vehicles (except 10 vehicles at 5s intervals and 100 vehicles at 10s intervals) in Nagoya probe experiment cost at least 5,000 Japanese Yen per month for real-time data transmission.

Table 2. Transmission costs at various polling intervals

Polling intervals	Packets / Month	Costs (Japanese Yen)	Plan*
5 seconds	223,718	10,000	Flat Plan S
10 seconds	112,270	7,454	Middle Plan S
50 meters	61,156	5,000	Middle Plan S
100 meters	37,727	5,000	Middle Plan S

* Plan in detail：

Flat Plan S: 10,000 Japanese Yen until 300,000 packets, and 0.3 Japanese Yen

per packet for the excess parts.

Middle Plan S: 5,000 Japanese Yen until 100,000 packets, and 0.2 Japanese Yen per packet for the excess parts.

Light Plan S: Basic charge of 900 Japanese Yen per month and 0.2 Japanese Yen per packet.

Super light Plan S: Basic charge of 500 Japanese Yen per month and 0.35 Japanese Yen per packet.

The large-scale investment of Nagoya P-DRGS (Probe-vehicle-based Dynamic Route Guidance System) project during the last 4 years provides a unique opportunity for the evaluation of the system cost-effectiveness level, and make the field trial research available, such like the comparison between various polling frequencies (Liu et al. 2006a, Yamamoto et al. 2006a) and polling schemes (Liu et al. 2007a), the comparison in urban and suburban settings (Yamamoto et al. 2006b), the implication of car data navigation on data collection (Miwa, 2005), and the efficiency of information dissemination (in process). The experience of operating such a large probe system is valuable and helpful to other cities, especially those in developing countries, to make more efficient project planning.

3 Traffic Monitoring through Probe Vehicles

The development of traffic data collection is promoted by IT technology, which helps researchers and transportation managers nudge probe technologies towards a practical solution. This chapter reviews the state of practice of probe field tests in academic field and makes a general review on the probe techniques, along with recent attentions.

3.1 Data Collection

For the real-time data collection from probe vehicles, a GPS receiver and a communication set are required. The GPS receiver records the time and spatial coordinates, traveling speeds, directions, as well as other useful information, and then such kinds of data are transferred to the control center through a communication network at generally a regular frequency, on time-based or on space-based. There are lots of communication technologies that have been used for such kinds of communication, such as DSRC (Dedicated Short Range Communication), GPRS (General Packet Radio Services), and so on. It should be noted that the GPS data missing is a common issue in city centers around with dense high buildings, while in suburbs such issue can be released to a great extent (Yamamoto *et al.* 2006b).

Since the message capacity at the same time is not infinite in most existing

communication networks, it is very difficult to collect GPS data second by second for large amount of vehicles required for covering the entire network. Moreover, frequent transmission would cost too much. It is necessary to reduce the polling frequency to an affordable level, at which the accuracy of obtained traffic information per unit of cost is the highest. As compared with the space-based polling scheme, the time-based polling scheme is better in avoiding the communication congestion (Liu *et al.* 2007a). Although with great possibilities, taxi dispatching systems are not yet appropriate for direct implementation in real-time traffic monitoring due to great uncertainties (Liu *et al.* 2007b).

Recently, some taxi companies in Nagoya improve their dispatch systems by employing a new communication technology developed by Panasonic, named as mcAccess-e (digital multi channel access system, abbreviated as MCA e, website of mcAccess-e). This technology allows more transmission capacity at the same time and therefore has a lower data missing rate. Another advantage of this technology is that transmission frequencies can be set freely according to the location within a city.

In the case of taxi probes, it should be mentioned that data can be categorized into two situations according to passenger occupancy states, namely, occupied probes and vacant probes, where these two categories present different traffic characteristics. Generally only the data collected from occupied probes are used for traffic surveillance because it is suggested that occupied probes provide lower variance and are more realistic in reflecting traffic condition (Miwa, 2005).

3.2 Data Processing

Location and time information from the probe vehicles must be map-matched on the network links for transforming into valuable traffic information. Sophisticated map-matching algorithms are required to overcome uncertainties in the positioning system and inaccuracies in the digital road map, as well as some special conditions. For example, the requirements of differentiating elevated urban expressways from surface roads while coping with long intervals between data transmissions impelled Miwa *et al.* (2004) to utilize a weighted link cost, Dijkstra algorithm, and a screening method to maintain a list of all potentially feasible routes for one trip, and then re-evaluate the likelihood of each potential route as a candidate for the correct route.

The real-time link travel speed is the most valuable information for traffic management and route guidance, and even for the longer-term performance monitoring of the road network. While the GPS instruments are capable of monitoring instantaneous speeds and such observation were suggested to be preferable to the link length-based speed (Du and Aultman-Hall, 2006), the truth is that such perfect information are not available in almost all the large-scale probe vehicle systems. The

great costs for transmitting such information at high frequency, say, second by second, are not afforded generally, therefore only infrequent GPS speeds are available. Moreover the GPS speeds from regular reports are not always representative of traffic conditions due to some uncertainties in the networks, such like stopping at signalized intersections. The alternative method of calculating travel speeds using the measured link travel times and the length of links seems more feasible for infrequent probe data.

Travel time estimations were traditionally based on aggregate data sets that examine either historic travel times over a number of days or at travel times in previous time intervals. Thus, it is expected that by "averaging" data from many probe vehicles, the overall situation is well characterized (Bishop, 2005). However, studies by Zietsman and Rilett (2000) indicated that such methods are not always satisfactory for real-time travel time estimation. Meanwhile it is suggested that the accuracy of link travel time estimation is influenced by many factors including signal control at intersections (Hellinga and Fu, 2002), observations from fragment probe trajectories (Uesugi *et al.* 2003), polling intervals (Liu *et al.* 2006b), and so on.

Since large numbers of taxi probes always congregate around the city downtown area where demands for service are high, therefore there still are some road links with limited observations, or absolutely no observations. As a result, the travel time for links with blank or few observations should be complemented based on other available information, say, the average travel time in the adjacent time periods along with the changing trends, or the average travel speed in the upstream/downstream links in the same time periods.

4 Recent Attentions and Solutions

Probe was designed from the beginning for more efficient collection of traffic condition data (David and Elva, 2000), therefore the exploration on the cost-efficiency of a probe system never stops. Recent attentions consist mainly of two issues: the minimum required probe vehicles that can satisfy both enough network coverage and enough probe observations for a certain link, and the issue of communication cost per vehicle that is directly affected by the polling frequencies.

The discussion on the minimum required probe numbers or penetrate rates gave lots of solutions. For instance, Chen and Chien (2000) developed a heuristic algorithm of determining the minimum number of probe vehicles required and then discussed the impact of traffic volume on the probe numbers; Horiguchi (2002) analyzed the relationship between the total number of taxis and the link coverage rate, and proposed a theoretical framework to estimate the update frequency of link travel information obtained by probe taxis. It is clear that the earlier studies always treated the entire roadway network as a uniform entity. Green *et al.* (2004) showed that fluctuations in

flow conditions over time and across different locations require varying sample sizes and addressed the issue of dynamic sample size requirements in probe-based traffic monitoring. Furthermore, Yakkala *et al.* (2006) developed a zonal sampling strategy that can ensure probe penetration across the entire network. Generally speaking, the minimum required probes for covering the entire networks during the whole days are very large, and even some developed cities would have to face the dilemma of the expectation of higher reliability based on large amount of probes and induced great operation costs that burdened the developers.

Cost-effective traffic data collection from commercial probes at lower polling frequencies attracts increasing attractions recently. The wide implementation of probe technology with higher cost-effectiveness will no doubt be a great progress in traffic surveys and traffic monitoring. The cost-effective deployment strategy weighs increasing transmission frequencies against expanding probe vehicle size. The minimum required probe size for various polling frequencies can be estimated by considering the average variability of link travel speeds at roadway segment level and network level. The optimal polling frequencies were examined from the individual link level (Liu *et al.* 2006a) and from the network level (Liu *et al.* 2006b). In order to achieve such an objective, researchers should consider the combination of three aspects (Liu *et al.* 2006b): 1) the accuracy or the reliability that a probe system can supply, 2) the accuracy levels of market expectation, and 3) the communication system. Some commercial dispatching systems, such as a taxi dispatching system, have also collected the spatial coordinates and time information during their trip. Generally, reported information is infrequent and there still exist some limitations that need to be overcome before wide application.

5 Issues in Probe Vehicle-Based Traffic Surveillance

The overwhelming superiority of probe technology relies on a wide coverage and a high penetration rate of probe vehicles. Although conceivable, the expectation of ubiquitous coverage of the entire road network is still far away from actual implementation nowadays. Some technical issues arise with probe implementation, of which four issues are discussed here. Better consideration on these issues would help to improve the cost-efficiency and accelerate probe tools towards successful commercial application.

5.1 Error Propagation

It is necessary and very important to know the performance level of the data collection and the accuracy of the resulted travel time estimation. The error variance is usually used to reflect such levels, where the observed travel routes are generally

considered true regardless of neither the complication of network nor the polling frequency. However, map matching errors can not be absolutely avoided through the data at lower polling frequencies even if the network is very simple and the algorism is well-worked.

The incorrect probe reports in practice might lead to inaccurate travel time estimation. Therefore, the propagation of map matching errors should be considered to reflect the real reliability level based on which travel time prediction can be obtained. The problem here is that the map matching accuracy for data at a certain polling frequency is network specific, thus it is not a constant value.

5.2 Reducing Bias in Travel Time Estimation

Beside the influence of individual days, polling frequencies, and signal delays, there are still some factors that have effect on the accuracy of travel time estimation for individual links. While these factors specified below are seldom considered well, they do exist and play important roles.

The first one is the aggregation time, which is usually defined as an integral minutes, say, 5 minutes. Note that such a time is not always integer times of the signal cycles. As a result, the estimation for adjacent time periods from random observations lose the characteristics if the percent of valid green times are different between these time periods. The implication of aggregation time on the accuracy of link travel time estimation should be considered carefully for excluding such external noises.

Secondly, taxis travel routes may be largely different form those of general travelers. As skilled and long time drivers, they might be capable of avoiding the congestion area according to their own experiences unless the customers insist their favorite routes. In such a case, some congestion links might obtain few or absolutely no observations. The same thing happens to the time dimension, for instance, the occupied taxi trips in the evening are much more than those in the daytime. It is easy to imagine that the common used complementation algorithms that regarding to the uplinks or previous time periods can not overcome such issue. Moreover, taxi trips have much shorter average length per trip compared with general private vehicles. As a result, the traffic condition of the entire network can not be reflected perfectly.

5.3 Link Travel Time Measurements from Infrequent Data

It has been widely accepted that reducing the polling frequencies to an appropriate level would improve the cost-efficiency level of probe data collection. However, the accuracy of travel time observation from such infrequent data decreased rapidly under the assumption of constant travel speed between consequent observations.

It is suggested that more considerations on the driving behaviors would help make more accurate link travel time measurements form infrequent observations. Although reasonable, this issue has not yet attracted enough attentions. One of the possible solutions is to regenerate the data set with higher polling frequencies, where the concept of interpolation is worthy of consideration.

5.4 Efficiency of Probe Coverage

It is easily understood that the coverage area and the link numbers with required minimum observed samples are inversely proportional under a curtain number of total probe vehicles. On the other hand, neither the increased number of newly coverage links nor the increased observations will be proportional to the increased probe vehicles.

Nowadays the available service from probe-based ATIS are not satisfactory and thus do not yet achieve enough market penetration, that is to say, probe vehicles are still far away from valid implementation. Therefore it is better to control the distribution of probe vehicles in the entire network to improve the arrangement level for approaching to the aim of ubiquitous coverage.

It is suggested that selecting more taxi companies that located around the entire network might improve the probe distribution efficiency, under the consideration that taxi probes may like to congregate around public transit stations near to their company locations, while the actual distribution in Nagoya experiment shows that such a strategy can not sufficiently solve this issue, although it does work. Further studies are needed to propose some more efficient probe distribution strategies.

6 Conclusions

The cost-efficiency is recognized as the priority of probe technologies. As a kind of passive traffic monitoring tool, probe vehicles can provide real-time travel time information and congestion information, observe route choice behavior, assess the performance of signal control and support to traffic management, while special surveys are traditionally required for each application.

The development history over the last two decades in Japan suggests that this technology is maturing and towards to the commercialization stage. It is also recognized that a cost-efficient probe deployment scheme is of certainty to be the impetus to probe evolution. Recent attentions on the cost-efficiency mainly focus on two aspects: the minimum required probe vehicles and the optimized polling frequencies.

While probe technology was designed primarily for improving the efficiency of traffic survey and enabling real-time traffic monitoring, the expected superiority relies

extremely on a higher penetration rate, not too long a way but need more exploration. It is suggested that more studies are required to improve the feasibility and effectiveness of probe technologies for application in the developing cities. Four issues are discussed in this paper and recommended as useful in improving the accuracy and efficiency: to explore the error propagation of map matching and its implication in travel time estimation; to avoid the bias of travel time estimations from aggregation periods, as well as from the spatial-time characteristics of taxis behaviors; to regenerate higher polling frequency probe data from the lower one; and to improve the probe coverage efficiency through a "control" of time-space distribution.

References

Bishop, R. (2005). *Intelligent vehicle technology and trends*, Artech House, Norwood.

Chen, M. and Chien, S.I.J. (2000). "Determining the number of probe vehicles for freeway travel time estimation using microscopic simulation." *79th Transportation Research Board Anural Meeting*, Washington, D.C.

David, L. and Elva, C. (2000). "The economics of traveler information from probes." *79th Anural Meeting of Transportation Research Board,* Washington, D.C.

Du, J. and Aultman-Hall, L. (2006). "Using spatial analysis to estimate link travel times on local roads." *85th Annual meeting of Transportation Research Board*, Washington, D.C.

Green, M.W., Fontaine, M.D. and Smith, B.L. (2004). "Investigation of dynamic probe sample requirements for traffic condition monitoring." *83rd Annual Meeting of the Transportation Research Board*, Washington, D.C.

Hellinga, B. R. and Fu L. (2002). "Reducing bias in probe-based arterial link travel time estimates." *Transportation Research Part C: Emerging Technologies*, Vol.10, No. 4, pp. 257-273.

Horiguchi, R. (2002). "Strategic disposition of taxi probe to accomplish effective data collection for travel information provision – A theoretical framework and its practice." *Proceedings of the 1st Japan ITS Symposium*, pp. 677-683.

Horiguchi, R. and Wada, K. (2004). "Effective probe data transmittal with detection of congestion pattern." *Proceedings of 11th World Congress on Intelligent Transport Systems*, CD-ROM, Nagoya.

Liu, K., Yamamoto, T. and Morikawa, T. (2006a). "An analysis of the cost efficiency of probe vehicle data at different transmission frequencies." *International Journal of ITS Research*, Vol. 4, No.1, pp. 21-28.

Liu, K., Yamamoto, T. and Morikawa, T. (2006b). "Study on the cost-effectiveness of a probe vehicle system at lower transmission frequencies." *Proceedings of the 5th Japan ITS Symposium,* pp. 335-340.

Liu, K., Yamamoto, T. and Morikawa, T. (2007a). "Comparison of time/space polling schemes for a probe vehicle system." *Proceedings of the 14th World Conference on Intelligent Transport Systems*, Beijing (forthcoming).

Liu, K., Yamamoto, T. and Morikawa, T. (2007b). Feasibilities of taxi dispatch system as probes for collecting travel time data, Proceedings of the 3rd International Symposium on Transportation Network Reliability.

Liu, K., Li, Q, Yamamoto, T. and Morikawa, T. (2007c). Cost-effectiveness in a probe vehicle system, Proceedings of the 6th Eastern Asia Society for Transportation Studies, CD-ROM, Dalian. (forthcoming)

Miwa, T., Sakai, T. and Morikawa, T. (2004). Route identification and travel time estimation using probe-car data, International Journal of ITS Research, Vol. 2, No.1, pp.21-28.

Miwa, T. (2005). Travel behavior analysis in road networks based on probe vehicle data, Doctor Dissertation (In Japanese), Nagoya University, Japan.

Morikawa, T., Yamamoto, T., Miwa, T. and Wang, L.X. (2007). Development and Performance Evaluation of Dynamic Route Guidance System "PRONAVI", Journal of Transportation Engineering (in Japanese), Vol. 42, No. 3, pp. 65-75..

Turner, S., Eisele, W., Benz, R., and Holdener, D. (1998). Travel time data collection handbook. Report FHWA-PL-98-035, U.S. Department of Transportation, Federal Highway Administration.

Uesugi, Y., Iryo, T., Oneyama, H., Horigichi, R. and Kuwahara, M. (2003). Estimation of expectation and variance for a section travel time based on fragmentary probe trajectories, Journal of Infrastructure Planning Review, Vol.20, pp.923-929, 2003.

Wada, K. and Fijii,H (2002). Research, development and field testing of the. probe car information system(iii). Proceedings of 9th World Congress on Intelligent Transport System, CD-ROM.

Website of mcAccess-e, http://www.mcaccess-e.com/ (accessed Dec. 5, 2006).

Website of P-DRGS, http://www.p-drgs.com/english/index.html (accessed June. 5, 2006).

Yamamoto, T., Liu, K. and Morikawa, T. (2006a). Variability of travel time estimates using probe vehicle data, Proceedings of the Fourth International Conference of Traffic and Transportation studies (ICTTS), Edited by Mao, B., Tian, Z., Gao, Z. and Huang, H., Science Press, Vol.4, pp. 278-287.

Yamamoto, T., Liu, K. and Morikawa, T. (2006b). Preliminary analysis on taxi dispatching data as potential probe data, Journal of Infrastructure Planning Review, (In Japanese), Vol.23, No.4, pp.863-870.

Yakkala, A.P., Fontaine, M.D. and Smith, B.L. (2006). Investigation of zonal sampling strategies in probe vehicle-based traffic monitoring systems, 85th Anural

Meeting of Transportation Research Board, Washington, D.C., CD-ROM.

Zietsman, J. and Rilett, L.R. (2000). A comparison of aggregate and disaggregate based travel time estimation for sustainability and ATIS systems application, 79th Anural Meeting of Transportation Research Board, Washington, D.C., CD-ROM.

Driving Fatigue Identification Method Based on Physiological Signals

Zhe Mao[1], Xin-ping Yan[2], Chao-zhong Wu[3]

[1]Engineering Research Center of Transportation Safety Control and Equipments, Ministry of Education, Wuhan University of Technology. P.O. Box 125, Wuhan, 430063; PH (86) 027-50853212; email: yanar@whut.edu.cn
[2]Engineering Research Center of Transportation Safety Control and Equipments, Ministry of Education, Wuhan University of Technology. P.O. Box 125, Wuhan, 430063; PH (86) 027-50853212; email: xpyan@whut.edu.cn.
[3]Engineering Research Center of Transportation Safety Control and Equipments, Ministry of Education, Wuhan University of Technology. P.O. Box 125, Wuhan, 430063; PH (86) 027-50853212; email: wucz@whut.edu.cn

Abstract: Driving fatigue is a serious issue for traffic accident risk, for fatigue affects the drivers' ability to continue driving safely. On-line monitoring of physiological signals while driving provides the possibility of detecting fatigue in real time and is absorbing much attention in the road safety field. This study aims to identify driving fatigue through a driver's physiological signals. Physiological signals, including heart rate, skin conduction, electro- myogram, skin temperature and respiration, were collected on eight simulating experiments with 1Hz sample rate. Then wavelet transform was used to find feature vectors and a two-step fuzzy cluster analysis was developed to classify the different levels of fatigue, including alertness, a transitional phase and fatigue. Compared with the experimental data, 93.75% of the normal and abnormal states and 63.64% of the transitional and fatigue levels had been correctly identified. The results showed that such a method could provide an effective way to detect a driver's alert state level. To improve the reliability of this algorithm, more physiological signals are needed and an impersonal metrics must be used to judge the actual state of the driver.

Key words: *Driving fatigue; Fuzzy Clustering; Pattern recognition; Physiological factor*

1 Introduction

Driving fatigue is one of the most important causes of traffic accidents because driving fatigue can adversely affect the drivers' ability to assess their level of alertness in order to continue driving safely. Seventy percent of American drivers report driving fatigued. The National Highway Traffic Safety Administration (NHTSA)(M. R. Rosekind. et al. 2004) estimates that there are 100 000 crashes that are caused by drowsy drivers and result in more than 1500 fatalities and 71 000 injuries each year in the U.S. In China, driving fatigue results in about 20% of the traffic accidents and 40% of the fatal crashes each year. Thus, it is essential for accident prevention to develop a system that actively monitors the driver's status level and alerts the driver for drowsy

driving conditions. One of the key steps toward developing such an on-line monitoring system is to develop a fatigue identification method. Around this idea, many studies have been reported in the literature for developing an active safety system for reducing the number of automobile accidents due to driving fatigue. These studies can be classified into the following categories:

1) Monitoring driver behavior: Most of these studies attempt to observe the driver's behavior from an image detected by a CCD camera. The typical behaviors related to the driver's status level include head nodding (Philip et al. 1998), eye-blinking(Appl. Sci. Lab. 1999, Bretzner, L.& Krantz, M. 2003), and facial expression(yawns)(B. Fasel, and J. Luettin 2003), steering, and etc.

2) Monitoring the driver's physiological signals by biologic sensors: The driver's status level can be depicted by physical signals, such as skin conductance, EEG (electroencephalogram), ECG (electrocardiogram), EOG(electroculogram) (Roman Bittner et al. 2000), EMG (electromyogram) signals, etc. The EEG signal is one of the most predictive and reliable signals that indicate drivers' alertness level; when a driver is fatigued, significant changes in slow wave activity will emerge (Lal. S.K.L. & Craig A 2005). Many studies attempt to find the determinate relationship between the driver's status level and these signals.

3) Monitoring the posture of the moving vehicle: The posture of the moving vehicle can reflect the driver's status level indirectly. Some researchers judge the driver's vigilance level by observing if the vehicle keeps in lane(S.-S. Ieng et al. 2003).

Although numerous physiological indicators are available to describe an individual's level of alertness, there still are some technical hurdles in on-line monitoring; the important defect is that deriving driver's fatigue only from one aspect of physiological features is not completely reliable. Therefore, fatigue recognition is most likely to be accurate when it combines multiple kinds of signals. Considering contextual variables and visual cues, Ji (Q. Ji et al. 2004) has constructed a Bayesian Net (BN) used to infer and predict human being fatigue, enhancing the reliability of fatigue recognition. However, one of the disadvantages of this work is that only textual and visual features are considered. G. Yang(G. Yang et al. 2005) has proposed a prototype based on D-S evidence theory using the fusion of contextual, visual and non-visual features in order to obtain reliable driver's fatigue recognition. The model can be directly applied to the intelligent transportation system (ITS) for alarming or predicting by recognizing driver's fatigue. But the model is only a prototype and the "evidence explosion" would take place as the evidence increases greatly. Maybe a new available method should be found.

To find a new method to identify driving fatigue accurately and robustly, in this study, five physiological signals, including heart rate, skin conduction, electromyogram, skin temperature and respiration, were collected during simulating experiments. Then a identification algorithm was developed using wavelet and fuzzy

cluster analysis.

2 Simulator Experiments and Data Collection

2.1 Subjects

Eight non-professional drivers(4 males and 4 females), with average of 22±2 years, were recruited from the university students to complete the driving simulator task.

2.2 Driving Protocol

The fatigue study was conducted in a temperature–controlled laboratory as the subjects performed a driving simulator task.

The driving task consisted of 10 minutes of active driving to familiarizes the subject, followed by not less than three continuous hours of driving(speed≈40km/h). Four hours continuous driving task ensured that the subjects would show physical signs of fatigue.

2.3 Data Collection

Six types physiological signals were obtained during the driving task: Heart Rate(HR), Skin Conduction(SC), electromyogram(EMG), skin temperature, respiration rate and amplitude, which were recorded at 1Hz. These sensors were connected to a eight-channel ProComp Infiniti encoder unit, which was connected to a personal computer by USB cable.

Physical signs of fatigue were identified using a video image of the driver's face, linked in real time with the physiological signals. All six physiological signals were categorized into three levels, they are alert, transitional phase and fatigue, represented by different colors, which were green, orange and red. As indicated by the color, green is a safe level(alert), red is a dangerous level(fatigue), orange denote medium level(transitional phrase).

In order to analyze subjects' physiological feature of different state levels conveniently, each physiological signal was divided into many segments of 5 minutes in length. From all the physiological signals, two 5-min intervals of data of every subject's three state level, total 48(8×3×2=48) group of 5-min intervals of data, were selected as samples of this study.

3 Methodology

3.1 Data pre-processing

To show the signals fluctuation in time domain, every physiological data should be normalized due to the individual discrepancy of each subject. Normalization steps was describe as following:

1) Calculating the average value of each sample,

2) Subtracting the average from every sample data to get a residual error;

3) Dividing every sample data by the max absolute value of the residual error.

3.2 Wavelet analysis and feature extraction

Analyzing all the physiological data in time domain, only the standard

deviation(SD) of HR(σ_h) and EMG(σ_m) related to the subject's state level(Fig.1. and Fig.2.), and σ_h was sensitive to the transitional phase level, while σ_m was sensitive to the alert level well.

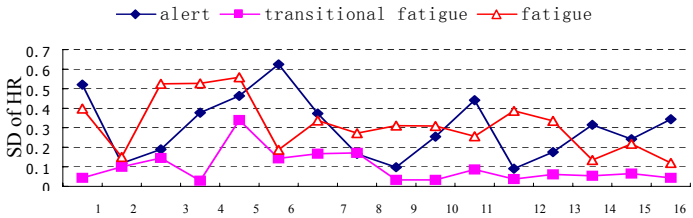


Fig.1. Fluctuation curve of σ_h in three different state levels

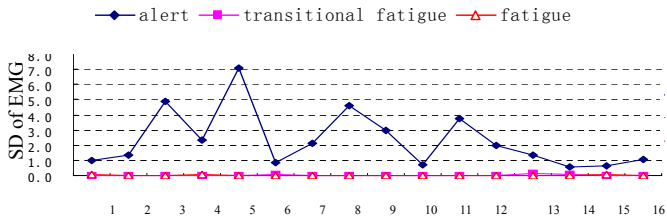


Fig.2. Fluctuation curve of σ_m in three different state levels

Fig.1. and Fig.2. showed that only σ_h and σ_m cannot identified the subject's state level correctly and more features should be mined from the data.

The wavelet transform, which is called "mathematical microscope", have fine resolution in both time and frequency domain and can focus on any detail of the signal by decomposing the high frequency signals gradually. Using the factor of wavelet transform, it is easy to decompose and reconstruct the signals at different frequency to get more features of the physiological signals availably. So, in the study wavelet transform was used to analysis the subjects' physiological signals in time and frequency domain.

There were many wavelet function could be selected and Daubechies4(db4) wavelet was chosen in this study because of it was sensitive to irregular signals and had fast maturity algorithm. Here, five-scales decomposing and reconstruction were used to analyze and contrast all the 48 segments of data. Combined with the FFT transform, energy, SD and range of different frequency signals were calculated and contrasted. Based on the result of contrast, six eigenvalues could be extraction to identify the subjects' state level and these 6 eigenvalues were HR's energy of the 5th lower frequency in wavelet decomposing(H_{ca5_e})(Fig.3.), HR's energy of the 1st higher

frequency (H_{cd1_e}) (Fig.4.), HR's range of the 2nd higher frequency (H_{rd2_range}) (Fig.5.), respiration capacity's SD of the 1st reconstruction signal, ($\sigma_{v(rd1)}$)(Fig.6.), EMG's energy of the 5th lower frequency (M_{ca5_e}) (Fig.7.), and EMG's energy of the 1st higher frequency(M_{cd1_e}) (Fig.8.).

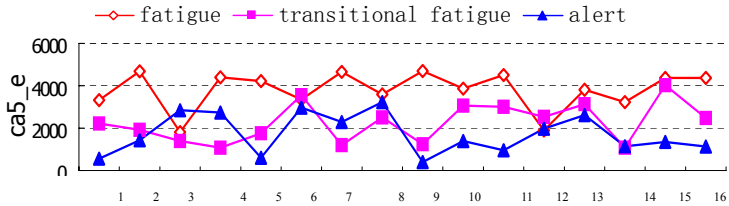


Fig.3. Energy contrast of H_{ca5_e}

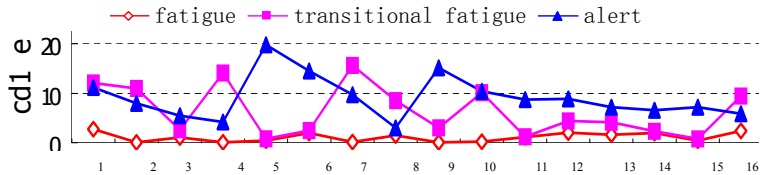


Fig.4. Energy contrast of H_{cd1_e}

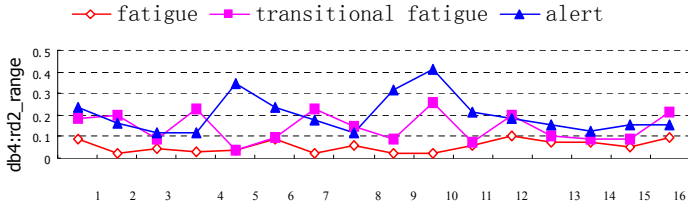


Fig.5. Range contrast of H_{rd2_range}

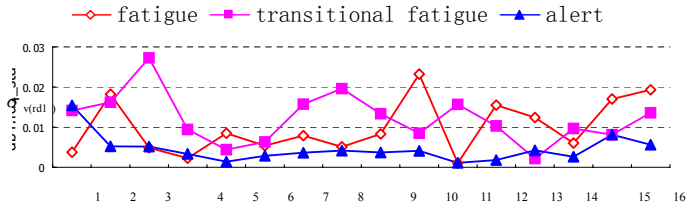


Fig.6. SD contrast of $\sigma_{v(rd1)}$

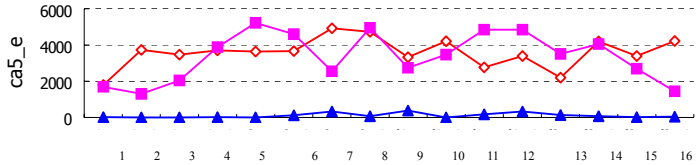


Fig.7. Energy contrast of M_{ca5_e}

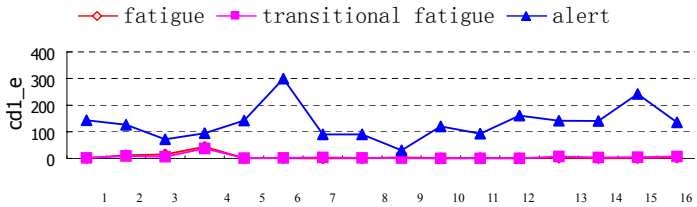


Fig.8. Energy contrast of M_{cd1_e}

Consequently, there are eight eigenvalues could reflects on the subjects' state level in a certain degree, besides two eigenvalues in time domain, σ_h and σ_m. But neither of these eight eigenvalues was enough to identify the subjects' state level alone. In order to get a robust method to identify subjects' state, the eight eigenvalues should be fused.

3.3 Identification subjects' state level

According to the analysis above, EMG's σ_m、M_{ca5_e}、M_{cd1_e} and respiration's $\sigma_{v(rd1)}$ were mainly sensitive to subjects' alert level while HR's σ_h、H_{ca5_e}、H_{cd1_e}、H_{rd2_range} were sensitive to subject' s fatigue and transitional phase level. So recognition of the subject's state level could be done by two steps. First, identify the alert level. Then, identify the fatigue and transitional fatigue. These steps could be describe as following:

1) Divided the whole state collection into two collections A and B. A was denoted the alert level collection and B indicate the non-alert level collection.

2) Divided the non-alert level collection B into B1 and B2. B1 was the transitional fatigue collection and B2 was the fatigue collection.

Both steps were carried out using the method of C-meaning fuzzy cluster. The clustering procedure was described as following:

1) Using formula (1) and (2), calculated each eigenvalue's cluster center of all the tree state level. The values of parameter c and b were 2.

$$m_j = \frac{\sum_{i=1}^{n}[\mu_j(x_i)]^b x_i}{\sum_{i=1}^{n}[\mu_j(x_i)]^b}, \quad j=1,2,...,c \tag{1}$$

$$\mu_j(x_i) = \frac{(1/\|x_i - m_j\|^2)^{1/(b-1)}}{\sum_{k=1}^{c}(1/\|x_i - m_k\|^2)^{1/(b-1)}}, \quad i=1,2,\ldots,n,\ j=1,2,\ldots,c \tag{2}$$

Here, $\mu_j(x_i)$ presented the degree of membership function of sample x_i to collection j.

2) Using formula (3) and (4), calculated the degrees of membership of different state value to both state collections.

$$A_1(x_i) = \frac{1/\|x_i - m_1\|^2}{(1/\|x_i - m_1\|^2) + (1/\|x_i - m_2\|^2)} \tag{3}$$

$$A_2(x_i) = 1 - A_1(x_i) \tag{4}$$

Here, m_1, m_2 were denoted cluster centers of A_1 and A_2, respectively.

3) According to the protocol of the maximum degrees of membership, identified the subject's state.

4 Results

4.1 Alert level identification result

The identification results of σ_m、M_{ca5_e}、M_{cd1_e}、$\sigma_{v(rd1)}$ 4 were listed in Table1. This result showed that the identification results of σ_m、$\sigma_{v(rd1)}$ were not satisfied, but the result of M_{ca5_e} was perfect which identified all the 48 samples' state level correctly. Therefore, M_{ca5_e} was chosen to identify the alert level only considering of the amount of computation.

Table 1. Contrast of σ_m、Mca5_e、Mcd1_e、$\sigma_{v(rd1)}$ 4 identification results

Eigenvalue	Alert	Non-Alert	Identification rate
σ_m	37.5%	100.00%	79.17%
M_{ca5_e}	100.00%	100.00%	100.00%
M_{cd1_e}	87.50%	100.00%	95.83%
$\sigma_{v(rd1)}$	93.75%	56.25%	68.75%

4.2 Transitional phrase and fatigue level identification result

The identification results of σ_h、H_{ca5_e}、H_{cd1_e}、H_{rd2_range} were listed in Table 2. This result showed that σ_h is sensitive to transitional phase level while H_{ca5_e}、H_{cd1_e}、H_{rd2_range} sensitive to fatigue level and the corresponding identification rate was higher than the other.

Table 2. Contrast of σ_h、Hca5_e、Hcd1_e、Hrd2_range identification results

Eigenvalue	Alert	Non-Alert	Identification rate
σ_h	68.75%	93.75%	81.25%
H_{ca5_e}	87.50%	68.75%	78.13%
H_{cd1_e}	100.00%	43.75%	71.88%
H_{rd2_range}	100.00%	50.00%	75.00%
Average	89.06%	64.06%	76.56%

Normalized column 1 in Table 2, the weights of σ_h、H_{ca5_e}、H_{cd1_e}、H_{rd2_range} identified fatigue level could obtained, which were 0.1930, 0.2456, 0.2807 and 0.2807 respectively. Analogously, normalized column2 in Table 2, the weights of this four eigenvalues identified transitional phase level is 0.3659, 0.2683,0.1707 and 0.1951. According to these weights, the membership functions of fatigue(B_1) and transitional phrase(B_2) collection, which were built according to weighted average model of fuzzy comprehensive assessment, were described like formula(5) and (6):

$$B_1(x_i)=0.1930B_1\sigma_h(x_i)+0.2456B_{1ca5e}(x_i)+0.2807B_{1cd1e}(x_i)+0.2807B_{1rd2range}(x_i) \quad (5)$$

$$B_2(x_i)=0.3659B_2\sigma_h(x_i)+0.2683B_{2ca5e}(x_i)+0.1707B_{2cd1e}(x_i)+0.1951B_{2rd2range}(x_i) \quad (6)$$

Combined above formulas and protocol of maximum degree of membership, the subjects' state level could be identified. According to this method, 27 of 32 samples had been identified correct state level.

4.3 Final identification result

In order to validate the method present in this paper, 16 new samples were selected from the original experiments data and the identification result was listed in Table 3.

Table 3. Final identification result of 16 samples

No.	$A(x_i)$	$B(x_i)$		Actual state	Identified level
		$B_1(x_i)$	$B_2(x_i)$		
1	0.5654	0.5153	0.4869	A	B_1
2	0.9942	0.0058		A	A
3	0.9210	0.0790		A	A
4	0.8982	0.1018		A	A
5	0.9893	0.0107		A	A
6	0.0003	0.4889	0.5125	B_2	B_2
7	0.0070	0.4888	0.5123	B_2	B_2
8	0.0085	0.4852	0.5152	B_2	B_2
9	0.0090	0.4865	0.5149	B_2	B_2
10	0.0075	0.5144	0.4895	B_2	B_1
11	0.0077	0.4869	0.5151	B_2	B_2
12	0.0028	0.4873	0.5144	B_1	B_2
13	0.0087	0.9087	0.1297	B_1	B_1
14	0.0086	0.4876	0.5167	B_1	B_2
15	0.0089	0.4874	0.5139	B_1	B_2
16	0.0088	0.5119	0.4908	B_1	B_1

Table3 showed that the method identified 93.75% samples of the alert level correctly, and identified 63.64% samples of the fatigue and transitional phrase level correctly.

5 Discussion

This study tested the applicability of multi-physiological factor fusion for detecting a driver's state level in a simulator driving experiment using set of sensors that do not interfere with the driver's perception of the virtual road. The method was developed based on the multi-physiological signals recorded by the driving simulator experiments. The results showed that the transitional phase and fatigue level data could be identified from all signals well. Due to the actual state level was observed by researcher, the boundary of the three state levels wasn't objective and this directly resulted in the distinguishing rate between transitional phase and fatigue level was not satisfied. In the future, some object criterion such as EEG should be used as the objective criterion of driver's state level, although there were also some insufficiencies, such as the relationship between different fatigue phases and EEG signal factors was not very determinate and the EEG sensors were too intrusive. 5-min intervals of data used in this study were propitious to find out the typical factors of the three state levels, but might neglect some boundary factors of the three state levels. The next phase of our research will consider the transitional phase factor as the most important feature in preventing fatigue related accidents.

Despite of these limitations, this method showed that physiological signals provided a viable method of measuring a driver's state level. With developing of physiological sensing system, this method could be used in future intelligent transportation systems to improve safety and to manage in-vehicle information systems cooperatively with the driver.

Acknowledgement

This study is supported by the "863 Plan" of the Ministry of Science and Technology of China. The supported project name is "The Causation Analysis of Road Traffic Accident Based on Multi-Analysis Method"(NO. 2006AA11Z214)

References

Yang, G. Lin, Y. Bhattacharya, P.(2005) "A driver fatigue recognition model using fusion of multiple features". Systems, Man and Cybernetics, IEEE, Vol. 2, 1777-1784.

Saroj KL Lal & Ashley Craig.(2005) "Reproducibility of the spectral components of the electroencephalogram during driver fatigue". International Journal of Psychophysiology, 55(2005) 137–143.

Bretzner, L.& Krantz, M.(2005) "Towards low-cost systems for measuring visual cues of driver fatigue and inattention in automotive applications." Vehicular Electronics and Safety, 2005. IEEE, 161–164.

Q. Ji, Z. Zhu, and P. Lan.(2004) "Real-time nonintrusive monitoring and prediction of driver fatigue," IEEE Transactions on Vehicular Technology, 53(4), 1052-1068.

Y.S. Yang, Z.Q. Yao, Z.Y. Li, K. Jiao(2003). "Investigation on correlation between ECG indexes and driving fatigue". Machinery Design & Manufacture, (5), 94-95.

S.-S. Ieng, J.-P. Tarel, and R. Labayrade.(2003) "On the design of a single lane-markings detector regardless the on-board camera's position," IEEE Intelligent Vehicles Symposium, 564–569.

B. Fasel, and J. Luettin.(2003) "Automatic facial expression analysis: a survey," Pattern Recognition, 36, 59-275.

M. R. Rosekind, E. L. Co, K. B. Gregory, D. L. Miller.(2000) "Crew Factors in Flight operations XIII: A Survey of Fatigue Factors in Corporate/ Executive Aviation Operations". National Aeronautics and Space Administration, Ames Research Center, Moffett Field, CA, NASA/TM- 2000-209 610.

Roman Bittner, Pavel Smrcka, Petr Vysoký, Karel Hána, Lubomir Pousek, Petr Schreib.(2000)." Detecting of Fatigue States of a Car Driver". Lecture Notes In Computer Science, 1933, 260 – 273.

Appl. Sci. Lab. (1999), "PERCLOS and eyetracking: Challenge and opportunity." Tech. Rep., Appl. Sci. Lab., Bedford, MA.

Philip W Kithil, Roger D.Jones, John MacCuishi.(1998) "Development of driver alertness detection system using overhead capacitive sensor array." SAE Technical Paper Series 982292, SAE International.

EVALUATION MODEL FOR THE URBAN PUBLIC TRANSIT NETWORK CONNECTIVITY BASED ON GRAPH THEORY

YANG, Xiaoguang[1]; ZHANG, Chao[1]; ZHUANG, Bin[2]

[1]School of Transportation Engineering, Tongji University, Shanghai, Postcode:200092, PH (+86) 21-65988372; FAX (+86) 21-65988372; EMAIL: yxgits@vip.sina.com
[2]Urban Transport Center, Ministry of Construction, Beijing, Postcode:100044, PH (+86) 10-58323205; FAX (+86) 1058323200; EMAIL: zhuangbin_01@163.com

Abstract: The connectivity capacity of the topological structure of a road network is based on the structure of the urban Public Transit Network, and the performance of the connectivity capacity can reflect the rationality of the network. This article adopts the graph theory to conduct modeling for the urban Public Transit Network. With the computer simulation software TransCAD, using the space analysis function, the model based on evaluation for the connectivity of urban Public Transit Network and algorithm are provided in this article. In the end, verification for the model and algorithm provided in this article is conducted by an instance.
Key Words: Public Transit Network; Graph Theory; Connectivity degree; TransCAD

1. Introduction

The structure of the urban Public Transit Network is the important factor for determining the comprehensive performance of the urban public transportation system. Planning the layout of the network and depots reasonably will enhance the comprehensive accessibility of urban public transportation, which will improve the degree of public transportation service and meet the traveling needs of the urban residents. At the same time, it will save construction funds and reduce the time and cost of the residents' travel which will have good economic, social, and environmental benefits. The public transportation system of many cities is set up along with the ever increasing speed of the city's development and the expansion of the city scale; however, there is no general consideration for the optimization of the design and the development of the city. As a result, problems such as the performance of whether the urban Public Transit Network can meet the requirements of the layout of the city have become the most urgent for designing the urban public transportation system and enhancing its operational efficiency. We need to evaluate the structural performance of the Public Transit Network to find the degree of service to answer the questions mentioned above.

Graph theory is adopted in this article to conduct modeling for the connectivity of the Public Transit Network. A feasible theoretical model and evaluation method are

provided for the Public Transit Network with the help of computer simulation software TransCAD.

2. Model Assumptions

The performance of the construction network lies in whether it can meet the needs of the society, that is to say, whether it can provide high quality service for the citizens [1]. As to the evaluations for the degree of public transportation service, many scholars, both domestic and overseas, have made comprehensive and deep study and have provided many systems for evaluation. As to the technical aspect, there is network density, non-linear coefficient, the number of repeated routes, etc. and as to the aspect of service, there is simplicity, safety, and rapidity and also the evaluation for efficiency. In order to give prominence to the key points, this article just considers the service degree indicator caused by the topological structure of the route network; viz. making a general evaluation for the Public Transit Network based on the connectivity of the Public Transit Network.

The indicator of connectivity reflects the connection conditions of every point which presents the structural characteristic of the network [2]. Many former calculations are aimed at the road network, [3] [4] and there is almost no study on the Public Transit Network. This article begins at the definition of the connectivity of the road network and provides the algorithm for the connectivity of the Public Transit Network.

The calculation of connectivity involves much basic data, while most of the public transportation design is based on a large quantity of survey data which needs a great labor force and physical resources; even some of that is still difficult to acquire. Making use of the macroscopic simulation software TransCAD and the GIS nature of the Public Transit Network, this problem can be well solved.

3. Model for the Connectivity of the Public Transit Network

As to the Public Transit Network, there are two subnets, line network and route network. According to the graph theory, the route network is a simple graph while the Public Transit Network is a non-simple network. It is more meaningful to study the line network than to study the route network since the line network reflects the layout of the whole transportation network, while the route network only reflects the coverage condition of the public transportation network. The route network, which is the simplified aspect of the line network, is also a simple graph, and as a result, some nature for the graph theory of the route network can be used for the line network.

Commonly, a complex route network can be expressed by sides and nodes. The Public Transit Network system consists of the Public Transit Network, formed by the public transportation route and stations with the stations as the nodes of the network

graph while the lines between the nodes are the public transportation routes. The nodes are connected: there is one or more public transportation routes connecting the two nodes; however, there is no connectivity for some nodes, that is to say, there are no public transportation routes between them.

This article will provide two algorithms for calculating the connectivity of the Public Transit Network by linear structure and topological structure.

(1) Model Based on the Connectivity of the Linear Public Transit Network

Suppose in a transit network system, L is the total mileage (km) of the public transportation route, A is still the acreage of the planned area (km^2), n is the number of the connected stations in this planned area, H is the average linear distance between two adjacent stations (km), ς is the average non-linear coefficient of the Public Transit Network. According to the expression of the connectivity in transportation, the computing formula for the connectivity C of this Public Transit Network is:

$$C = \frac{L/\varsigma}{nH} = \frac{L/\varsigma}{\sqrt{nA}} \qquad (1)$$

(2) Model based on Connectivity of Transit Network with Topological Structure

In order to give a better analysis for the problems and modeling for the Public Transit Network, we regard the Public Transit Network as a connected graph in which every station is treated as a node and the route between two stations is treated as a side. Suppose the average degree of all the nodes in the Public Transit Network is $\mathrm{d}_{\mathrm{ave}}(\mathrm{G})$, viz. the average sides connecting to every node is $\mathrm{d}_{\mathrm{ave}}(\mathrm{G})$. As a result, as to the Public Transit Network with n nodes, there should be $\frac{\mathrm{n}\cdot\mathrm{d}_{\mathrm{ave}}(\mathrm{G})}{2}$ sides. Suppose H is the average distance between the stations, viz. the average distance of every side in the graph of the Public Transit Network and the whole linear length of the L_{Total}, the Public Transit Network in the planned area is:

$$L_{Total} = \frac{\mathrm{n}\cdot\mathrm{d}_{\mathrm{ave}}(\mathrm{G})}{2}\cdot H \qquad (2)$$

Suppose L_{Total} is the whole mileage of the routes in the planned area; ξ is defined as the mileage of the actual route, viz., a non-linear coefficient, and as a result, L_{Total}/ξ means the total linear mileage of the route in the planned area, that is:

$$\frac{\mathrm{n}\cdot \mathrm{d}_{\mathrm{ave}}(\mathrm{G})}{2}\cdot H \approx L_{Total}/\xi \qquad (3)$$

So we can get the approximate computing formula of the connectivity C as follows:

$$C \approx \frac{\mathrm{d}_{\mathrm{ave}}(\mathrm{G})}{2} \qquad (4)$$

According to the definition in the graph theory, G is a connected graph, there will be:

$$\sum_{v\in v(G)} d(v) = 2e(G) \qquad (5)$$

Where $e(G)$ is the number of sides in the connected graph G.

From that we can conclude the average degree of the vertices in the connected graph G is:

$$\mathrm{d}_{\mathrm{ave}}(\mathrm{G}) = \frac{2e(G)}{n(G)} \qquad (6)$$

In this formula, *n (G)*—the numbers of summits of the connected picture *G*.

Therefore, the approximate calculative formula of a public transport network connectivity is:

$$C \approx \frac{e(G)}{n(G)} \qquad (7)$$

When the network is larger, the connectivity degree of a public transport network will be calculated better by this approximate calculative formula.

(3) The judgment standard of the public transport net connectivity

The judgment standard of the public transport net connectivity is in Table 1.

Some of the lines in the public transport networkare the same; therefore, according to the correlative knowledge about graph theory, it can be proved that the connected degree of a public transport network outweighs or equals the connected degree of a transport network. According to this, we should raise the standard when we evaluate a connected degree of a public transport network.

Table 1. Based on the divided standard of public transport net connectivity of a line type model

Span of connectivity	Evaluation of connectivity
$C \le 1.0$	public transport network connectivity is worse
$1.0 \le C \le 2.0$	public transport network connectivity is good
$2.0 \le C \le 3.22$	public transport network connectivity is consummate
$C \ge 3.22$	public transport network connectivity reaches an ideal state

4. Case study

With the discussion on the evaluation model about connectivity of a public transport network, we take the planning program (planning 80 public transport lines on the island) of the XiaMen public transport line network as an example, discuss the evaluation method and its application of public transport network connectivity based on the macroscopic simulation software---TransCAD.

It can be known from the analysis above that if we calculate the connectivity of a public transport network from the line type angle only, it mainly involves five indicators: total mileage L of public transport roads in the planning area, the average non-line coefficient ξ of the public transport lines, the number of stations in planning area n, planning area A, or the average distance between two stations H. If we take the second way to calculate, based on the method of the line network developing the construction, it will refer to two indicators: the sides number of public transport network *e (G)* and the number of summits of the public transport network *n (G)*. In general, these indicators can be difficult to gain through ordinary ways; this article uses the software TransCAD to solve these indicators based on the road network characteristics of traffic space.

TransCAD has a great analysis function for space information, and its space data is organized by the definition of layer. When the public transport network is created, there are road layers, route system layers, station layers, traffic district layers and so on. According to the attribute information entered, TransCAD software has calculated the length of every road; it can be seen in chart 2. The total length of the public transport line network was 834.49 km through calculations. In the same way, XiaMen Island was divided into 25 districts, as shown in table 2. Every acreage of the district is in table 3, plus acreage of all the districts, gaining the planning acreage of the public transport line network which is 128.18 km2.

Table 2. Length of the Route

Route_ID	Route_Name	Length
7	z2-41002	1354
8	z2-41002a	1242
11	z4-41004	517
12	z4-41004a	477
13	z5-41005	800
14	z5-41005a	824
15	z6-41006	1188
16	z6-41006a	1188
19	z8-41009	1069
20	z8-41009a	1299
21	z9-41010	992
22	z9-41010a	992
25	z11-41012	567
26	z11-41012a	567

Table 3. Area of Traffic Region

ID	Area	numble
152	6052.73	131
200	7644.40	132
228	19358.38	133
145	5712.29	134
185	6893.30	135
251	8999.29	137
255	11123.65	138
272	150772.42	139
273	10913.78	141
262	10921.54	144
269	13461.45	145
351	100333.02	147
234	5098.90	148
199	194196.17	149

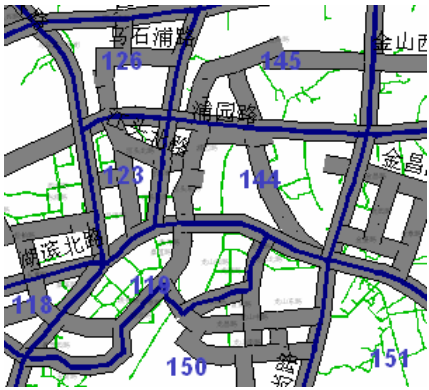


Figure 1. Partition of Traffic Region

Figure 2. The straight distance of a public transportation line

We may obtain each indicator in a connectivity calculated model through the use of the spatial analysis function of the macroscopic simulation software TransCAD.

Table 4. Comparative analysis of indicator value

Name of Indicator	**Indicator Value**	**Connectivity Degree Value**	
Total Mileage of Route	834.49	Method One	Method Two
Planning Area	128.18	3.05	2.84
Average Nonlinear Line Coefficient	1.54	**Connectivity Valuation**	
Numbers of Transit Bus Stops	246	Public Transport Network is Basically Improving and Perfect	Public Transport Network is Basically Improving and Perfect
Numbers of Public Transport Network	698		

Based on the above analysis, we may see that combining TransCAD software and calculating according to method one, the connectivity degree of the public transport network in Xiamen city is 3.05; by using method two, the connectivity degree is 2.84. Compared with the standard, both of the calculating methods show the transit line system is basically improving and perfect.

5. Conclusion

The framework of an urban transit line system decides the connectivity of the topology framework of this line system, whether the connectivity can reflect the rationality of the framework of network. This paper uses the definition of the degree of connectivity in a driveway network, unifies the related knowledge of graph theory, and gives the connectivity model of an urban transport network, in order to determine the connectivity of the public transportation wire. But actually due to the limits of the research method and conditions, connectivity is unable to be obtained or it is hard to gain a large amount of data which is needed in the process of calculating the degree of wire connectivity,. It is also difficult to do accurate appraisal work. By applying the geography document of the TransCAD software spatial analysis tool, we may carry on the computation analysis to the connectivity degree target of a public transport network. Through the calculating analysis of examples, it has proved the rationality and feasibility of the method, and it has provided the theoretical basis for the appraisal of an urban public transport network.

References

Gu Zhenghua, Li Xuhong, Yu Shijun. "The Connectivity degree Research on the Regional Highway network layout framework" *J. Driveway Traffic Technology*, 2005(Feb.) Page: 86-89.

Han Chuanfeng, HU Zhiwei. *"The Network Chart Method Which is About Urban Public Transport Network Performance Evaluation" J.System Engineering,*2003(March） Page: 58-61.

Han Chuanfeng. "Synthesize Calculation Appraisal of Urban Transport Network performance" *J. Harbin Technical University Transaction,* 2005(June) Page: 864-866.

Pu Yuehua. *Graph Theory and Its Application. [M]*. Nanjing：Southeast University Press, 2002.

Qin Xia, Gu Zhenghua, Li Xuhong. "Researching on the Connectivity Appraisal Target of Layout and Programming of the Project of Regional Driveway Network" *J. Civil Engineering Transaction,* 2006(Jan.) Page: 112-116.

Qian Songdi. *Over-All Planning[M]*.Beijing: Qinghua University Press，1993.

TransCAD User Guidance（M），2006.8.

TransCAD Model Handbook（M）for Traffic Requirements，2006.8.

Wan Xujun, Hu Anzhou. "Research the Expression Method of the Transportation Network Connectivity" *J. Planning and Management*, 1999.3 Page: 61-66.

UCF DRIVING SIMULATOR'S RESEARCH AND APPLICATION IN TRAFFIC SAFETY AND OPERATION

X. Yan[1], E. Radwan[2], M. Abdel-Aty[3], X. Wang[4], R. Harb[4], C. Santos[5]

[1]Civil and Environmental Engineering Department, University of Tennessee, Center for Transportation Research, 309 Conference Center Bldg, Knoxville, TN 37996-4133, Email: yxd22222@yahoo.com.
[2, 3, 4, 5]Department of Civil & Environmental Engineering, University of Central Florida, Orlando, FL 32816-2450, Email: aeradwan@mail.ucf.edu

Abstract: The UCF (University of Central Florida) driving simulator housed in the Center for Advanced Transportation Systems Simulation is an I-Sim Mark-III system with a high driving fidelity and immense virtual environments. The UCF driving simulator provides a research platform for multi-disciplinary investigations in the traffic engineering area. This paper describes several simulator studies that aimed at enhancing the quality of traffic safety and operation, including measuring drivers' minimum gap acceptance at stop-controlled intersections, analyzing the effects of dynamic sight distance problem on traffic safety, evaluating a pavement marking countermeasure for reducing red-light running, and assessing the safety benefits of variable speed limits.

Key words: *driving simulator, traffic safety, traffic operation, driving behavior, human factor*

1. Introduction

With the progress of computer science and electronic engineering in recent years, driving simulators used for training and research are being rapidly developed. In recent years, numerous studies have applied driving simulators in the traffic engineering area, such as pavement marking effect (Horberry et al., 2006), traffic signs (Dutta et al., 2004), gap acceptance behavior (Alexander et al., 2002), passing maneuver (Jenkins and Rilett, 2005), crash avoidance study (Smith et al., 2002), driving distraction due to mobile phones (Rakauskas et al., 2004), and so on. The use of an advanced driving simulator has many advantages over similar real world or on-road driving research, including experimental control, efficiency, expense, safety, and ease of data collection. Especially, a simulation experiment has the ability to reproduce dangerous driving conditions and situations in a safe and controlled environment to test driver behaviors.

In Spring 2002, The Center for Advanced Traffic System Simulation (CATSS) of the University of Central Florida (UCF) acquired a reconfigurable driving simulator placed in a high bay lab. The UCF driving simulator is an I-Sim Mark-III system with

a high driving fidelity and immense virtual environments. It provides a research platform capable of supporting studies in driving simulation, driver training, human factors, ITS, and traffic engineering. This paper describes several simulator studies that aimed at enhancing the quality of traffic safety and operation, including measuring drivers' minimum gap acceptance at stop-controlled intersections, analyzing the effects of dynamic sight distance problem on traffic safety, and evaluating a pavement marking countermeasure for reducing red-light running.

2. Features of UCF Driving Simulator

The UCF driving simulator is mounted on a motion base capable of operation with 6 degrees of freedom. The simulator cab is a Saturn model that has an automatic transmission, an air condition, a left back view mirror and a center back view mirror inside the cab. The simulated environment is projected at 180 degrees of field view and at a resolution of 1280×1024 pixels. It includes 5 channels (1 forward, 2 side views and 2 rear view mirrors) of image generation, an audio and vibration system, steering wheel feedback, operator/instructor console with graphical user interface, sophisticated vehicle dynamics models for different vehicle classes, a 3-dimensional road surface model, visual database with rural, suburban and freeway roads plus an assortment of buildings and operational traffic control devices, and a scenario development tool for creating real world driving conditions. The output data include steering wheel, accelerator, brake, every car's speed and coordinates, and a time stamp. The sampling frequency is 60Hz.

Figure 1. UCF driving simulator-outside and inside

3. Research and Applications of UCF Driving Simulator

3.1 Gap Acceptance at a Stop-Controlled Intersection

Because the driver's gap-acceptance maneuver is a complex and risky driving behavior, gap acceptance is a highly-concerned topic for traffic safety and operation. Using the UCF driving simulator, this study evaluated the effects of major traffic speed and driver age and gender on left-turn gap acceptance behaviors at a stop-controlled intersection.

The experiment was a 3 (age) × 2 (gender) × 2 (major traffic speed) within-subject repeated measures design. Sixty-three paid participants in three age groups, ranging in age from 20 to 83 years, were recruited for this research. The young group ranged from 20 to 30 years old, the middle-age group ranged from 31 to 55 years old, and the old group ranged from 56 to 83 years old.

In the gap acceptance experiments, a long straight undivided two-lane highway was used as the major road. Its length is around 3000 m with a lane width of 3.66 m. A driving scenario consists of 3 stages. During the course of the experiment, the driver first drives the simulator for 400 meters along the minor road (stage 1). When the simulator approaches the intersection, all vehicles on the major road begin to enter the scene. Then, according to the voice reminder and indication sign (stop sign and left-turn sign), the driver stops at the intersection and waits for the appropriate gap in the traffic on the major road to make his/her left turn maneuver (stage 2). When the simulated vehicle enters the major road and its speed is caught up with the major road traffic speed (stage 3), the experiment is completed. All subjects were tested in two scenarios, 40.2 km/h (25 mph) major road speed (Scenario A) and 88.5 km/h (55 mph) major road speed (Scenario B). A challenge in designing this experiment was how to make the drivers perform their minimum left-turn gap-acceptance maneuver in the same way accomplished in the real world. For obtaining that goal, the oncoming traffic on the major road from the right was composed of two classes of intermingled gaps to make the traffic appear random. One gap classification had very small gaps (less than 3 sec) that were unlikely to be accepted by the participants. The other class consisted of increasing gaps in which the subsequent gap was one second larger than the previous one. This design assured that the selected gap would be close to the driver's minimum acceptable gap and the major road traffic looked realistic.

The experiment results illustrated that the major road traffic speed contributes to an interactive effect between gap-acceptance drivers and vehicles on the major road. First, drivers are more likely to accept smaller gaps in the higher major-road traffic speed scenario than the lower major-road traffic speed scenario. It implied that drivers show relatively less sensitivity to vehicle approach speed but more sensitivity to on-coming vehicle’s distance or position. On the other hand, due to accepting smaller gaps for the higher traffic speed, drivers have shorter separation from the following vehicle, result in more speed reduction and deceleration rate of the following vehicles to accommodate the turning vehicle, and therefore contribute to more effects on major road traffic operation. These findings implied a connection between high crash risk and high traffic speed at stop-controlled intersections.

It was found that there is an apparent age difference in the left-turn driving behavior, as well as gender difference. Compared to younger drivers and middle-age drivers, the older subjects tend to select larger gaps to make left turns, turn the steering wheel more slowly, and keep a further clearance distance from the following car. The conservative driving attitude of older drivers would compensate driving ability decrease. For the gender factor, the results showed that male drivers accept smaller gaps than female drivers, which reflect a general conservative driving attitude of

female drivers. However, the result also showed that the older female drivers contribute to the most significant influence in the major road vehicle speed. The finding implied that older females could be the most vulnerable group for the relatively complex driving maneuvers because of more reductions in their driving ability.

3.2 Testing Pavement Marking Countermeasures for Reducing Red-Light Running

In this project, a pavement-marking countermeasure is proposed for signalized intersections to assist drivers in making a clear stop/go decision at the onset of yellow-change interval. To test the effectiveness of the pavement marking countermeasure, a driving simulator experiment was conducted to compare driving behaviors at intersections with and without the marking. At a signalized intersection, there is a length of roadway on the intersection approaches, where it is difficult for drivers to decide to cross or stop at the intersection at the onset of the yellow signal. This region of roadway is commonly referred to as the 'dilemma zone'. A pavement marking with a word message 'SIGNAL AHEAD' (see Figure 2) is placed on the pavement of the upstream approach to a signalized intersection and is sufficient to permit vehicles cruising at the speed limit to stop safely before reaching the intersection stop bar. The marking position is related to speed limit and deceleration rate, which is referenced in the ITE's *Traffic Engineering Handbook* (1999). The distance from the marking to the stop bar is calculated by Equation 1,

$$X = Vt + \frac{V^2}{2a + 19.6g} \tag{1}$$

Where

X = distance from the marking to the stop bar (m)
V = 85th percentile speed or speed limit (m/sec)
t = reaction time (1.0 sec)
a = average deceleration rate (3.05 m/sec^2)
g = grade of intersection approach.

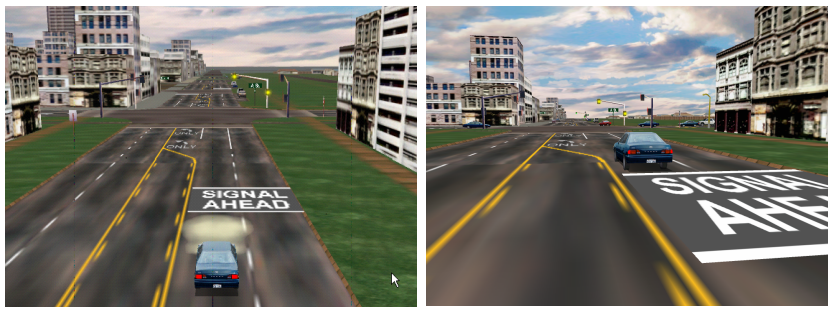


Figure 2. Pavement marking at a signalized intersection

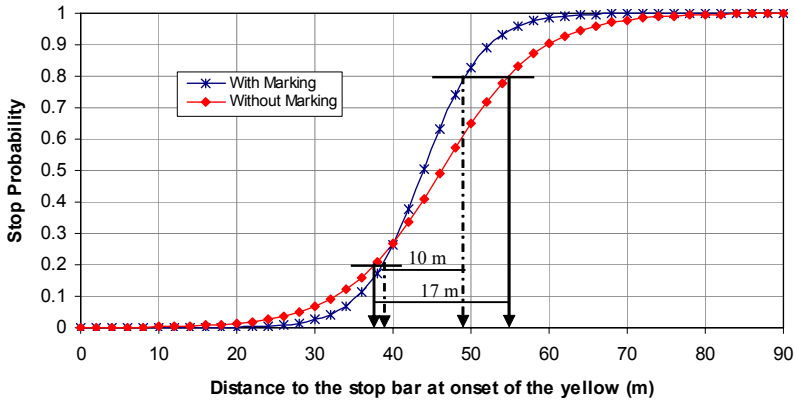
The intent of the marking design is that for drivers located upstream of the marking at the onset of the yellow phase, they are encouraged to stop at the intersection. On the other hand, when drivers are located downstream of the marking at the yellow phase onset, they are encouraged to cross the intersection. Potentially, the marking provides drivers relative-position information: where they can cross intersections safely. Because the markings have a fixed position at the intersection, drivers who adhere to the speed limit will be more likely to benefit from the marking information.

This experiment utilized within-subjects repeated measures to test the effectiveness of the pavement-marking countermeasure. Three treatment factors were speed limit, treatment type, and yellow-phase onset distance. There were two levels for speed limits (48.3 km/h = 30 mph and 72.4 km/h = 45 mph), two levels for treatment types (with and without marking), and eight yellow-phase onset distances for each speed limit. The yellow-onset distance determines the length from the simulated vehicle to the intersection when drivers encounter a yellow-phase intersection. For the 48.3 km/h speed limit, the eight levels for yellow onset distances range from 25 to 85 m with 8.57-m increment; for the 72.4 km/h speed limit, the eight levels range from 55 to 110 meters with 7.86-m increment. Totally, the factorial manipulation of the three factors described above results in 32 unique intersection-approach types. In this experiment, there were 42 paid test subjects in two age groups, including 18 younger subjects (18-26 years) and 24 middle-age subjects (26-55 years). Data analysis was based on the decisions made by the 42 subjects approaching the 32 test intersections. Each subject responded to 16 test signalized intersections with markings and 16 regular signalized intersections without markings for a total of 1344 driver-intersection encounters.

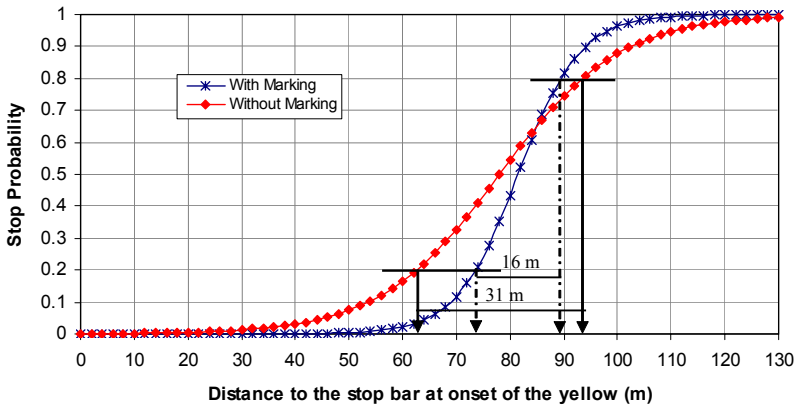
Based on the logistic regression results, the significant variables associated with drivers' stop/go decisions include Distance (yellow-onset distance), Speed (approaching speed), Marking (with or without markings), and the interaction variable between Distance and Marking. The marking effect on the driver stop decision is correlated to the yellow-onset distance. The correlation between Marking and Distance is clearly illustrated in Figures 5-a and 5-b. With the help of marking information, drivers tend to make better stop/go decisions: stop when the signal change is encountered at farther distances, and cross when the signal change is encountered at shorter distances from the intersection. As illustrated in Figure 3, with markings, the uncertainty distance between 20 and 80 percent probability of stopping is 7 m for the 48.3 km/h (17 m vs 10 m) and 15 m for the 72.4 km/h (31 m vs 16 m) shorter than those without marking, respectively. Therefore, the marking information may lead to reducing the region of driver uncertainty that possibly results in higher accident rates.

Comparison of running red-light rates between scenarios with and without markings can directly reflect the effect of the pavement marking countermeasure. Red-light running rates without marking information are higher than those with markings. For 48.3 km/h speed limit without marking, there were 15 red-light running events representing a red-light running rate of 4.5 percent; for 72.4 km/h speed limit without

marking, there were eleven red-light running events representing a rate of 3.3 percent. However, with the help of marking, there were only four red-light running events representing a rate of 1.2 percent for 48.3 km/h speed limit and five red-light running for a rate of 1.5 percent at the 72.4 km/h speed limit. Potentially, the pavement marking could results in a 65.4 percent reduction in red-light running. By the Chi-square test, the reduction in red-light running rate due to the marking is statistically significant at the 0.05 significance level ($\chi_{1,1344}^{2} = 8.478$, p = 0.004).



(a) For the 48.3 km/h speed limit and the approaching-vehicle speed is 48.3 km/h



(b) For the 72.4 km/h speed limit and the approaching-vehicle speed is 72.4 km/h

Figure 3. Probability of stop based on the logistic regression models.

3.3 Effects of dynamic sight distance problem

3.3.1 Horizontal Visibility Blockage

The horizontal view blockage occurs when a driver's visibility is inhibited to his left or/and right at an intersection. This can occur when someone is driving a passenger car behind a Light Truck Vehicle (LTV), such as vans and SUVs. The driver's sight view in the following passenger car may be restricted and the driver may not be able to see and know what is happening beyond the LTV at the intersection. For instance, if a pedestrian invisible to the driver in the following vehicle suddenly crosses the intersection from left to right, the LTV driver would be forced to slam on his brakes. This critical event may leave the following passenger car driver insufficient time to react appropriately and stop, which could lead to a rear-end collision with the leading LTV.

In order to determine whether LTVs contribute to rear-end collisions, two scenarios were designed and analyzed in the driving simulator (that was between-subject). The first scenario serves as a base scenario where the simulator car drives behind a regular passenger car (PC), and the second scenario serves as a test scenario where the simulator car drives behind an LTV. We recruited two matching groups A (following a passenger car) and B (following an LTV) with the gender and age as matching factors. Each group consisted of 20 subjects. For each group, there are two age levels. The young drivers vary between the ages of 18 and 25 and the middle age drivers vary between 26 and 55. Each group consisted of 5 young males and 7 middle age males, 3 young females and 5 middle age females.

The driving course of each scenario can be described in three stages as shown in Figure 4. In the first stage, the driver in the simulator car cruises on a four-lane urban road with a 72.42km/hr (45 mph) posted speed limit. The traffic in the scene is assigned to flow at 72.42 km/hr (45 mph). The purpose of this design is to make the simulator car drivers adapt to relatively higher speed traffic. At the second stage, the simulator car approaches the signalized intersection and stops at the red phase behind the LTV, which is assigned to be there. When the light turns green, the LTV is assigned to cruise at a 56.33 km/hr (35 mph), following the speed limit, while the following vehicle (simulator car), accustomed to a higher speed limit follows the LTV at a speed tending to be greater than the speed limit. Moreover, the two-lane road in the direction of the simulator is dropped to 1 lane to inhibit any passing between vehicles. Therefore, the simulator car driver is forced to drive behind the LTV until the two-way stop intersection in the third stage, in which an opposing left turn vehicle (Veh#3) will turn left in front of the leading vehicle (Vehi#1) with a very short clearance distance. This critical event triggers that the leading vehicle has to make a sudden stop to avoid the potential crash. The assigned deceleration rate for the leading vehicle is 0.8g. Then, the following vehicles (Veh#2) also need to stop consequently in order to avoid a rear-end crash. For the design of the experiment, the width of the leading LTV was 1.88 m (6.17 ft) while the width of the leading PC was 1.70 m (5.58 ft) and the assigned deceleration rate for the leading vehicle is 0.8g. The experiment is

to test if following LTV can contribute to a higher probability of rear-end collision due to restricted visibility of traffic surrounding by the LTV.

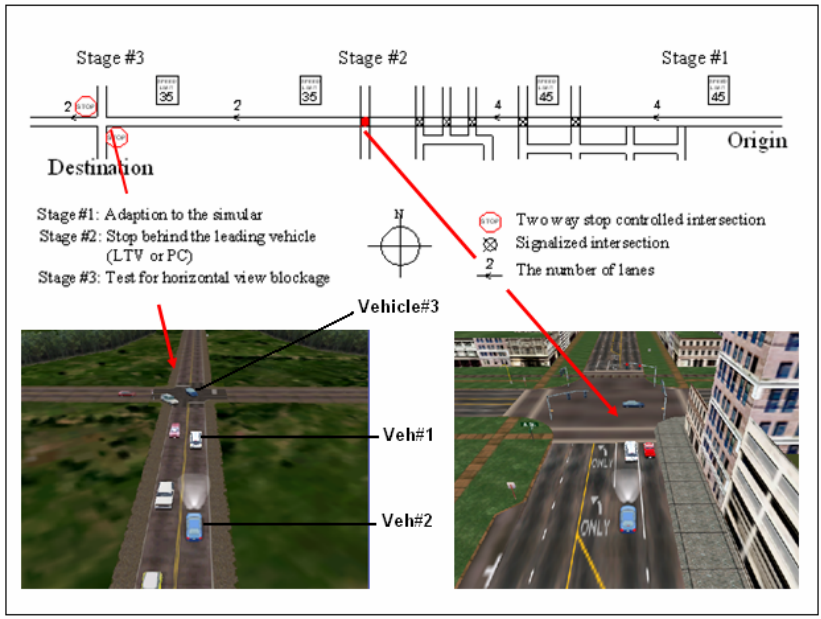


Figure 4. Scenario design for horizontal sight blockage

The experiment result showed that the probability of being involved in a rear-end collision with a PC is 0.1 (P_{PC}= 2/20=0.1) and with an LTV is 0.4 (P_{LTV}= 8/20=0.4). According to the logistic regression, the odds ratio for LTV compared to PC was 7.364. In other words, passenger car drivers behind an LTV are 7.364 times more likely to be involved in a rear-end collision with a leading LTV compared to a leading PC (p=0.0132). According to the summary statistic shown in Table 1 and the three-factor ANOVA results, the mean speed for following a PC (52.37km/hr) is significantly lower ($F_{(1,33)}$=4.91, p=0.034) than the mean speed for following an LTV (55.20km/hr); the mean gap distance for following a PC (34.94m) is significantly larger ($F_{(1,33)}$= 6.17, p=0.018) than the mean gap distance for following an LTV (23.03m); the deceleration rate for following LTV (6.78m/s/s) is significantly larger (P=0.038) than following a PC (5.42m/s/s).

From the conducted analyses, it was confirmed that LTVs contribute to the increase of rear-end collisions probability in the critical traffic event. This result based on the driving simulator experiment complied with the results obtained by previous researchers (Abdel-Aty and Abdelwahab, 2004).

Table 1. Dependent variables versus independent variables statistics summary

DEC RTIME SPEED GAPDIST GAPTIME * PCLTV

Factor	PCLTV		DEC (m/s/s)	RTIME (s)	SPEED(km/h)	GAPDIST(m)	GAPTIME(s)
PC-LTV	PC	Mean	5.42	2.40	52.37	34.94	2.46
		N	20	20	20	20	20
		Std. Deviation	1.51	1.22	4.11	16.95	1.33
	LTV	Mean	6.78	2.23	55.20	23.03	1.51
		N	20	20	20	20	20
		Std. Deviation	1.04	0.50	2.52	9.46	0.60
Gender	Female	Mean	6.44	2.50	53.76	29.89	2.07
		N	16	16	16	16	16
		Std. Deviation	1.40	1.16	4.79	16.32	1.31
	Male	Mean	5.86	2.19	53.80	28.38	1.93
		N	24	24	24	24	24
		Std. Deviation	1.47	0.73	2.77	14.08	1.01
Age	Young	Mean	6.24	2.27	54.20	27.93	1.87
		N	16	16	16	16	16
		Std. Deviation	1.58	0.81	2.58	15.65	1.06
	Middle	Mean	6.00	2.34	53.51	29.69	2.06
		N	24	24	24	24	24
		Std. Deviation	1.40	1.01	4.26	14.56	1.19
Total		Mean	6.10	2.31	53.79	28.98	1.98
		N	40	40	40	40	40
		Std. Deviation	1.46	0.92	3.66	14.83	1.13

3.3.2 Vertical Visibility Blockage

Traffic signals' degraded perception or invisibility could be one of the most important factors that cause red light running. During the yellow change interval at intersections, the following passenger car driver being in the vertical shadow of the blindness caused by the leading the large size vehicles (LSVs) like buses or trucks (moving/dynamic obstacle) may can not see the signal change and have insufficient response time to come to a complete stop, which could lead to red light running. A simulator experiment was designed to identify whether LSVs increase the probability of red-light running due to vertical view blockage, and furthermore, to analyze whether the additional signal pole on the right side of the road decreases red light running resulting from the sight distance issue.

The vertical view blockage scenario consists of three sub-scenarios: a base scenario where the simulator car follows a regular passenger car; a test scenario where the simulator car follows an LSV (school bus); and the suggested FHWA solution scenario where the simulator car follows an LSV with the additional traffic signal pole. We recruited 20 subjects for each sub-scenario, matched by age and gender. The whole experiment course can be described in three stages as shown in Figure 5. In the first stage, the subject drives the simulator car to a T-intersection where he/she is instructed to make a left turn. At the second stage, the simulator car approaches the signalized intersection, where the signal phase has just turned green and where a leading vehicle (a school bus or a passenger car) just started making a right turn slowly. The subject is assigned to make a right turn behind the leading vehicle at that intersection. The speed limit in the second stage is 35 mph and the route consists of one lane per direction to inhibit the following car from any attempt to pass the leading vehicle. Finally, in the third stage, at the time T0, the traffic signal turns yellow. At that time, the leading vehicle is at a maximum safe distance to cross the intersection.

However, the following vehicle is not at a safe distance to clear the intersection. The hypothesis in this study is that the driver's vertical sight view in the following car is more likely to be restricted by a leading LSV compared to a passenger car. Therefore, he/she may not be aware of the traffic signal change when following the LSV. At the time T1, 3.5 seconds (yellow change interval) after T0, the traffic signal turns red. At time T2, this could be prior or subsequent to T1, the following vehicle's driver reacts and is faced with two alternatives. The driver can either suddenly stop or run the red light.

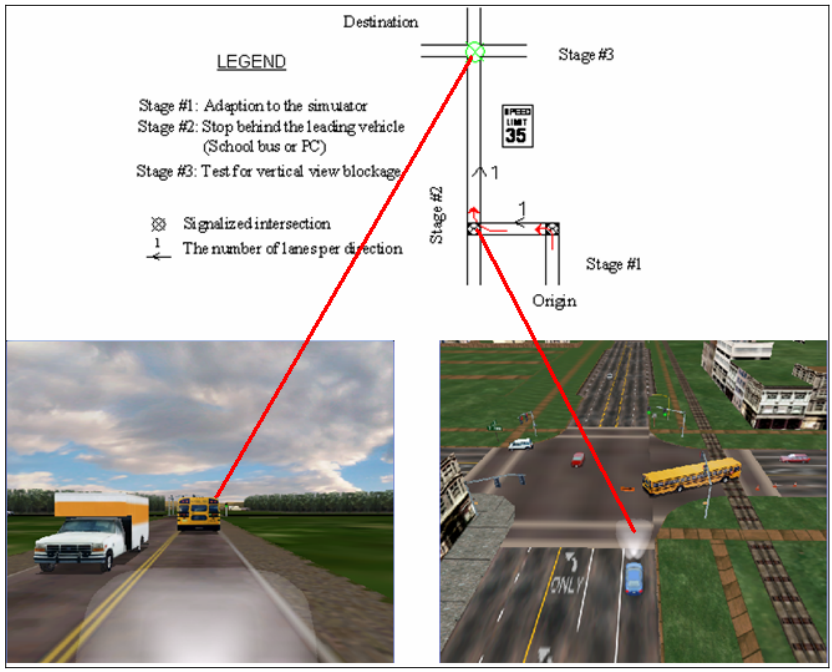


Figure 5. Scenario design for vertical sight blockage

From the collected data, when following the PC, 2 subjects out of 20 subjects driving the simulator behind the PC ran the red light. However, when following the LSV, 10 subjects out of the 20 subjects driving the simulator car behind the LSV ran the red light. A chi-square test showed a significant statistical difference (p=0.006) between the red light running rates (0.1 for following a PC Vs. 0.5 for following an LSV). With the additional traffic signal pole, 4 out of 20 subjects who drove the driving simulator behind the LSV ran the red light (02. red-light running rate). A Chi-Square test showed that light running ratios when following an LSV with and without an additional signal pole at 0.05 significance level are significantly different (p=0.047). The additional signal pole reduced red light running from 50% to 20%.

The collected data demonstrated that the brake response times of following the LSV are greater than the brake response times of following a PC. The mean response time for following a school bus is 3.45 sec and the mean response delay time for following a PC is 2.02 sec. With the additional traffic signal pole, the brake response times of following the LSV reduced to 2.45 sec. This difference is marginally significant (p=0.065). Therefore, the additional signal pole is a potential solution for vertical visibility blockage due to LSVs.

4. Conclusions

This paper introduced four application examples of the UCF driving simulator in traffic engineering. These studies fully took advantages of the advanced driving simulator's accurate data collection, effective traffic arrangement, high operation efficiency, and driver safety. Considering the Chinese traffic characteristics, the driving simulator can developed as an effective research tool to study older drivers' safety issues, reduce crash involvement between pedestrian and vehicles, inspect design drawback in import highway design projects, and enhance the quality of transportation landscapes especially in urban areas.

References:

Abdel-Aty, M. Dilmore, J. and Dhindsa, A., 2006. Evaluation of variable speed limits for real-time freeway safety improvement. Accident Analysis and Prevention, Vol. 38, pp. 335-345.

Abdel-Aty, M. A., Abdelwahab, H. T., 2004. Analysis and prediction of traffic fatalities resulting from angle collisions including the effect of vehicle configuration and Compatibility. Accident Analysis and Prevention, Vol. 36, pp. 457-469.

Alexander J., Barham P. and Black I., 2002. Factors influencing the probability of an incident at a junction: results from an interactive driving simulator, Accident Analysis and Prevention, pg 779–792.

Dutta, A., Fisher, D. and Noyce, D., 2004. Use of a driving simulator to evaluate and optimize factors affecting understandability of variable message signs. Transportation Research Part F: Traffic Psychology and Behavior, Vol. 7, pp. 209-227.

Horberry, T., Anderson, J., and Regan, M. A., 2006. The possible safety benefits of enhanced road markings: A driving simulator evaluation. Transportation Research Part F: Traffic Psychology and Behavior, Vol. 9, pp. 77-87.

Institute of Transportation Engineers (ITE), 1999. Traffic Engineering Handbook, Washington, DC: Institute of Transportation Engineers.

Jenkins, J. and Rilett, L., 2005. Classifying Passing Maneuvers: A Behavioral Approach. Transportation Research Record 1937, pp. 14-21.

Smith, D., Najm, W. and Glassco, R., 2002. Feasibility of Driver Judgment as a Basis for Crash Avoidance Database Development. Transportation Research Record 1784, pp. 9-16.

Rakauskas, M.E., Gugerty L.J., and Ward, N.J., 2004. Effects of naturalistic cell phone conversations on driving performance, Journal of Safety Research, Vol. 35, pp. 453–464.

FURTHER RESEARCH AND APPLICATION OF TRAIN INTERCONNECTING NETWORK SYSTEM

Xi LI [1] , Xiaoning ZHU [2] and Guoqiang CAI[1]

[1]State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University. Beijing, P.R.China; PH 86(10) 5168-3973;email:dllx@dl.cn
[2] School of Traffic and Transportation, Beijing Jiaotong University. Beijing, P.R.China; PH 86(10) 5168-8314;email:zhuxiaoning@jtys.bjtu.edu..cn

Abstract: This paper introduces the buildup process of train interconnecting network base on CAN bus and the process that realizes multiple locomotives network's dynamically buildup .The paper first introduces the functions and then described the inauguration and regular process of the interconnecting network, analyzes its inauguration arithmetic and describes the process of using object Petri net (OPN), finally, this paper checks the feasibility and stability designed above and give a further applications expectation.

Key words: Interconnecting network, Inauguration, Couple, Uncouple, Object petri net

1 Introduction

Train communication network is a data communication network of train-control network. Locomotive and vehicle bus are two rank buses in train communication network prescribed in standard IEC61375-1 and they link the train as a whole. Vehicle information was transmitted by vehicle bus; locomotive information after protocol conversion by gateway will be transmitted to target locomotives by locomotive bus. After the process above the control instructions and message sent out by driver could be transmitted to each locomotive through the train-control network and the states of each locomotive could be transmitted to the driver. The process is useful for the train's safely running.

Along with the development of industrial automation, field bus such as Lonworks, CAN, FF, Profibus are widely used in industrial control network. The paper introduces a design method of train interconnecting network based on CAN bus which realized data communication between locomotive buses. CAN is a serial bus which is realized point to point, one point to multiple points and broadcast communication. CAN was widely used with its high feasibility and stability. The design changed some characteristics of the bus and defined the first two bytes as a special control field. Device on the bus was defined to be a master or slave node and it has its node address by different data setup in the special control field. The format of special field was displayed in figure 1.

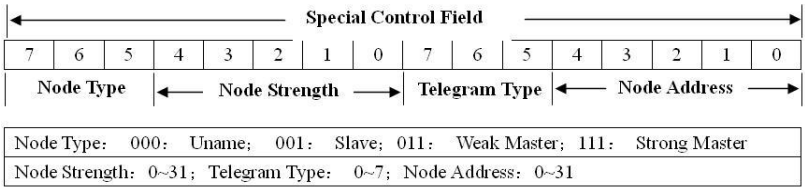


Figure 1. Special Control Field Format

Network interconnecting process includes two aspects, inauguration and regular running process. Inauguration is the initialization of the train network interconnecting process and it realizes dynamically-composed multiple train network. Regular process realizes regular control data exchange process in train control which includes process data, message data, and supervisory data exchange among nodes on the bus. Couple and uncouple are two processes in regular process. Couple realized multiple compositions to be whole and uncouple separates the composition into multiple ones. Both of the operations above need to run inauguration. Network interconnecting displayed in figure 2.

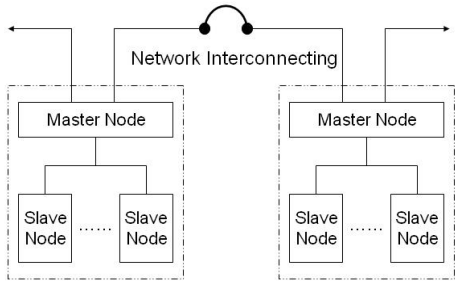


Figure 2. Network interconnecting structure

2Inauguration design

When the composition changes, in particular each time compositions are coupled or uncoupled, the Master reconfigures the bus. This procedure is called inauguration. The application may implement different mastership strategies by ranking a node as either: strong node, weak node or slave node. There is only one node acquires mastership with its node address 1 which is decided by application. A slave node never acquires the mastership.

There are some telegrams on the bus during the inauguration as follows:

1）detection telegram: An end node shall signal its existence to another node through a detection request, to which the other node （if it exists and can answer） shall respond with a detection response.

2）naming telegram: The master shall communicate its allocated address and strength to a slave by a naming request, which the slave shall acknowledge with a naming response.

3）status telegram: The master shall request the status of a node with a status request, to which the slave shall respond with a status response.

4）set to intermediate telegram: The master shall request a slave to put its switches in intermediate setting by sending a set to intermediate request, which the slave shall acknowledge with a set to intermediate response.

5）topology telegram: The master shall communicate its topology to a slave by a topology request, which the slave shall acknowledge with a topology response.

6）set-to-end telegram: The master shall request a slave to switch to end setting and to adopt new composition strength by a set-to-end request, to which the slave shall respond with a set-to-end response.

7）presence telegram: The master shall request an end node to signal its presence and the possible presence of another composition through a presence request, to which the end node shall respond with a presence response.

Inauguration is a preparation for regular process of train interconnecting which links all train compositions and nodes on the bus together to be one composition. Functions of the process are described as follows:

（1）Decides maser node in the composition according to the correlative competitive principle.

（2）Detects unnamed nodes and compositions on the bus and name the new nodes and allocate node address.

（3）Sets all the nodes on the bus to be intermediate node or end node, open bus switches correctly.

（4）Sets bus topology period list according to the desired individual period and node descriptor of each node.

（5）Topology all the node and send the bus topology to each node includes address, position and category.

Master node could be an intermediate node or an end node, it communicates to two directions on the bus, direction 1 and direction 2(which is direction 1 is decided by applications). Master node sets up itself an end node when it's a single node on the bus, opens their bus switches, sends a detection request to direction 1 and 2 by turns and receives detection request sent by other nodes, responds with a detection response if received a detection request. If master is not an end node, it sends a presence request to

an end node, and end node sends a detection request to detect whether there is a new node or composition and sends a presence response to master reports the result.

An unnamed node responds with a detection response after received a detection request. Master sends naming request to the unnamed node after received detection response. Master could name 31 nodes at most on the bus, and its address is 01, names nodes in its direction 1 in descending order from 31(31, 30, etc.), and nodes in its direction 2 in ascending order (2, 3, 4, etc.).An unnamed node responds with a naming response after received naming request. Then master sends a set to intermediate frame to end node and a set to end frame to the new node, after that the formerly end node becomes an intermediate node and the formerly unnamed node becomes the new end node of the composition, it turns off its auxiliary channel and switch on its main channel in the direction of the master, opens the switch in the other direction.

Master sends a status request to the new end node. Master sends a status request to itself if it is an end node. End node sends a status response containing the node descriptor, which indicates the frame size for process data and the desired polling period, and sends a detection request to another direction to detect whether there is another new unnamed nodes or compositions exists.

The inauguration will complete when there aren't unnamed nodes or compositions reported after three status requests and the new composition has finished. Inauguration running process is displayed in figure 3.

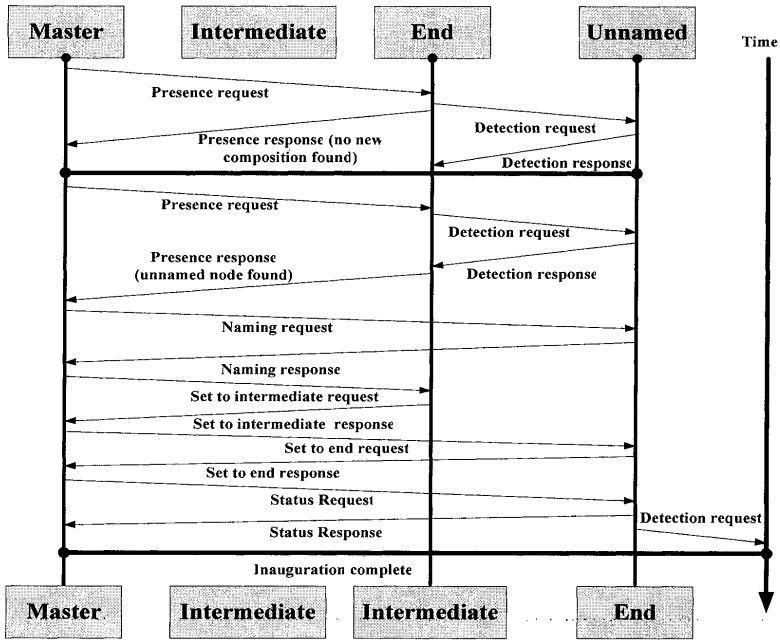


Figure 3. Inauguration running process

Master computes the new periodic list and the individual period of each node is based on the desired period and the frame size of each node. Master builds the topology, a data structure containing for all nodes, the address, the node type, and a master topo uniquely identifying this inauguration. Master then distributes the topology individually to each slave through a topology request, which each slave acknowledges through a topology response.

After all slaves acknowledge reception of the new topology, and after waiting a basic period to allow all nodes to update their interpretation of the process data, the master enters regular operation and starts polling the nodes for process data.

But if an end node sends a status response to master, it reports that a new unnamed node or composition was found. If the detected is an unnamed node, the above operations will repeat; but if the detected is a composition, the two composition need to compare their respective strengths and take a decision of which composition is stronger:

• If the strengths are different, the node of the stronger composition insists and the weaker yields. This ensures that the master of the largest composition retains the control of the bus;
• If the two compositions have the same strength, the end nodes decide among themselves which one yields and which one insists. To break ties if no side is strong, the end node which first receives a detect request from the other "yields" and decreases its strength, the other does not change its strength.

A master starts an inauguration if its end node reports a stronger composition, but only under the condition that the inauguration by all nodes in its composition and by the remote composition is allowed.

The master of the loosing composition will unname explicitly its nodes.

The master of the winning composition also unnamed explicitly its nodes, but it remains master of the bus and starts immediately to rename all nodes. When no more nodes are detected, the master closes the inauguration and it begin to distribute the topology. It resumes regular operation.

Multiple compositions or nodes interconnecting process described by OPN is displayed in figure 4.

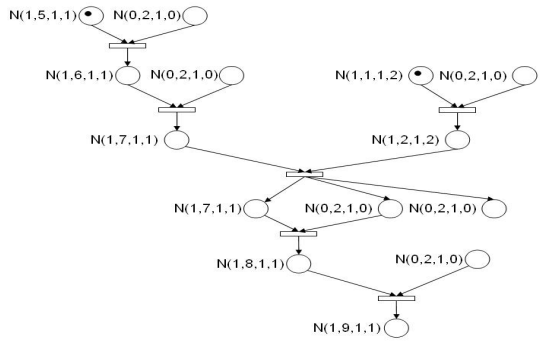


Figure 4. Train interconnecting process

The Petri net above is a quadruple form object Petri net described as follows:
OPN=N (NT，NS，NA，MT):
NT= {0，1}
NS= {1，2，3，…31}
NA= {1，2，3，…31}

MT= {0, 1, 2}

NT indicates node numbers of the composition; NS shows strength of the composition strength; NA figures the node address, if the composition is composed of multiple nodes it shows the address of the master node. MT indicates the master mode type, strong master, weak master or unnamed node.

3 Regular process design

Every basic period after inauguration master sends a presence request to poll an end node, end node responds with a presence response. Master polls another end node in the next basic period. If master is an end node it also polls itself and sends presence response so that all nodes could supervise its existence. End node sends a detect request to the direction opposite to the master. End node should report "no new composition found" in presence response if no detect response received.

And in every basic period after inauguration master sends process data request to the named node and the node sends a process data response. Whether sends message data and supervisory data is decided by applications.

When an end node finds another composition or another unnamed node, it indicates the strength of its composition in the detection request and detection response to the other end node, and at the same time its application enables inauguration. End node tells the master a new composition was found and its strength, master decides whether inauguration is allowed. If it is allowed, inauguration will start and couple the two compositions.

During the running process, composition will be uncoupled because of practical requirements or faults, one composition separated into many.

In the separated composition, the primary master renames the nodes remained. In other compositions, if the master hasn't appointed by applications, one of the nodes maybe the node nearby the separated point could be the master, it starts a new inauguration and resume the configuration.

4 Experimental analysis

For checking the feasibility and stability of the train interconnecting process, we did some tests include three aspects as follows:

1) Computes inauguration cost times;

2) Computes the efficiency of multiple compositions dynamic interconnecting the network;
3) Tests stability of dynamic interconnecting network in multiple interferential factors condition.

In the test we collected 40 inauguration cost times between two nodes which test in normal and multiple interferential factors condition. Test data is displayed in figure 5. Moreover, we tested the inauguration cost times that 16 nodes composed through 16 single nodes interconnected the network dynamically. Test data is displayed in figure 6.

Through the test, the result concludes that:

1) Time each inauguration costs is less than 25 ms;

2) Time 16 nodes composed network dynamically cost is less than 450 ms, so we can estimate that the time 32 nodes composed network will be less than 1 second;

3) The train interconnecting network designed is accorded with standard IEC61375-1, could be aboard used in multiple compositions integration and network dynamic interconnecting.

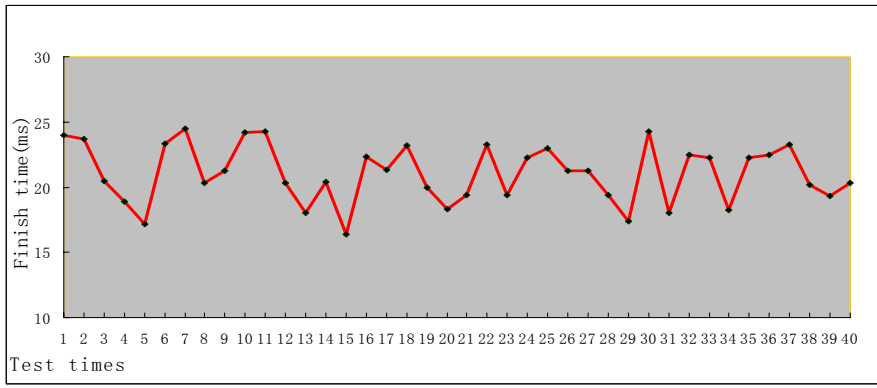


Figure 5. Inauguration cost time

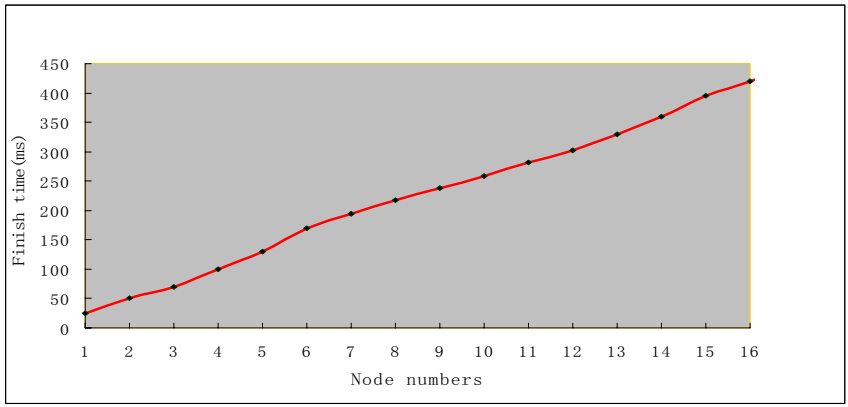


Figure 6. Multiple nodes inauguration cost time

5 Conclusion

This paper introduced a train interconnecting network on CAN bus, particularly introduced the network inauguration process. Master and slave nodes realize the interconnecting between the compositions through sending and receiving status request and response frames，naming request and response frames and topology request and response frames, etc. Next the paper introduced the regular process during network running, finally tested the network performance through experimentation and described the interconnecting process through object Petri net. The experimental test result indicated that the network accorded with standard IEC61375-1 could be widely used in multiple composition integration and network dynamic interconnecting. The interconnecting arithmetic applies to other industrial buses as well as correlative field in other industries.

References

Am Weichselgarten, CAN in Automation, “ErlangenController area network-a serial bus system, not just for vehicles” ,26,D-91058

B.P. Upender, “ Analyzing the Real-Time Characteristics of Class C Communications in CAN through Discrete Event Simulations,” Advanced Digital Systems, United Technologies Research Center

Gerald Schickhuber and Oliver McCarthy, “distributed fieldbus and control network systems” COMPUTING & CONTROL ENGINEERING JOURNAL FEBRUARY 1997

IEC61375-1(1999) ，Part1：Train Communication Network:602-632

IS0 11898: 'Road vehicles-interchange of digital information controller area network (CAN) for high-speed communication', 1993

M. Schwartz, Broadband Integrated Networks, ser. ISBN:0-13-519240-4. New Jersey: Prentice Hall, 1996.

R.M. Daoud, H.H. Amer, and H.M. Elsayed, "Fault-Tolerant Networked Control Systems under Varying Load," IEEE Mid-Summer Workshop on Soft Computing in Industrial Applications, IEEE SMCia/05, Espoo-Finland, June 2005

STEINER, J., Selectron Ltd, "Distributed architectures in automation systems with CAN bus", 1995

V. M. Stathopoulos, I. S. Venieris, "Modelling and Performance Design of Distributed Intelligent Networks" 2001 IEEE

Safety Research during Entrance and Exit of Highway Tunnel Based on Drivers' Eye Movement Experiment

PAN, Xiaodong, DU,Zhigang and YANG,Zhen
(School of Transportation Engineering, Tongji University, Shanghai 200092, Siping ST. PH 086-021-65982276 Fax 086- 021-65982507. email: panxd3@163.com, zhig_du7@hotmail.com, yangzhen2276@263.net)

1 Abstract:

Luminance's sharp variation at highway tunnel's entrance and exit will engender driver's visual adaptation difficulty, and even lead to traffic accident. By using EMR-8B system, a driver's pupillary changes and variety of moving fixation during entrance and exit of highway tunnels are selected as research object, the evaluation index of driving safety based on the variety velocity of pupil area and the evaluation index of visual information processing level based on the moving fixation, are brought forward. The rational range of the variety velocity of pupil areas and the index of moving fixation are also obtained, and can be used in safety evaluation of highway tunnel's entrance and exit.

Key words: *tunnel's entrance and exit, safe index, eye tracking experiment, pupil area, moving fixation*

2 Introduction

The tunnel section is the throat of highway transportation where many traffic accidents happen, and traffic accidents inside the tunnel are difficult to dispose of, extremely easily creating heavy casualty. The investigations of Pan Xiaodong and Ying Ruihua indicated: the tunnel's entrance and exit are always accident blackspots of the road. The sharp variety of lighting during a tunnel's entrance and exit has greatly increases the driver's visual load, bringing out harmful driving behaviors, and even leading to traffic accidents. Scholars have chiefly concentrated on safety design of road route and driving sight-distance theory of highway tunnel research, but lack the support of driving experiments, and seldom have considered the influence of drivers' visual reactions.

Through a large number of driving experiments, this article uses the EMR-8B system to capture drivers' pupillary reaction to the illuminance during tunnel's entrance and exit, finds out the reaction characteristics of driver's dynamic vision, and puts it into use to the illuminative setting during tunnel's entrance and exit.

3 Experimental design and data collection

3.1 Experimental tunnels

The experimental highways are the Yuyuan freeway, the Yuanmo freeway in Yunnan province, and the Shangshan freeway in Zhejiang province. The speed limits of these tunnels are $60km \cdot h^{-1}$ or $80km \cdot h^{-1}$. 10 typical tunnels were chosen, whose lengths are between 155m and 2445m.

3.2 Experimental drivers

6 male drivers, 4 in their 20s, 2 in their 30s, and whose revised visual acuity surpasses 5.0, are selected.

3.3 Experimental vehicle and apparatus

The experimental vehicle is SANTANA 2000 produced by Shanghai Volkswagon Automobile Co.Ltd, and EMR-8B system produced by Japan NAC Co.,Ltd is chosen to measure the driving dynamic vision. The MS6610 luminance meter produced by Shenzhen Huayi Co., Ltd is selected to measure illuminance, acceleration collection meter developed by research group（national state patent number is 200520046765.5） is used to carry on the acceleration test .

3.4 Experimental measurement factors

(1) Tunnel illuminance measurement

Along the driving direction inside the tunnel, measuring location distance from the tunnel port are chosen as -20m(out of the tunnel), 0m, 5m, 10m, 20m, 30m, 50m.

(2) Vehicle acceleration and driver's pupil diameter measurement

In driving process, the acceleration collection meter is used to record the vehicle acceleration during tunnel's entrance and exit, EMR-8B system is used to record the variety of drivers' pupil diameter incessantly, and the results were analyzed and studied by system software.

4 Drivers' dynamic vision characteristics

4.1 Dynamic vision characteristics and driving safety

In the driving process, 80%～90% of drivers' information relies on the vision receival from the external world, and 95% of visual information is dynamic, so dynamic vision characteristics are most closely related to traffic accidents.

The research of Beatty and Vernep etc. indicate that pupil area is the sensitive index to weigh the human body load: the pupil's magnification indicates a high pressure on a person mentally and physically. As for drivers, the variety of pupil area can describe drivers' visual adaptation process and driving load degree. The drivers' moving fixation can denote the choice, duration and transferring of visual attention, namely the drivers' visual information processing. The driving load and driving environment's comfort can be evaluated by the record analysis of drivers' pupillary changes and variety of moving fixation with the EMR－8B system.

4.2 Pupil area's variety.

(1) Pupil diameter's variety

In the daytime, the driver's pupil diameter changes as in Figure 1. Before entering the tunnel port, the driver's pupil diameter is about 1.9mm, the pupil diameter enlarged rapidly after entering the tunnel port, and stabilizes at 6.5mm after 10s' light adaptation. The pupil diameter changes contrarily when the drivers come out of the tunnel port.

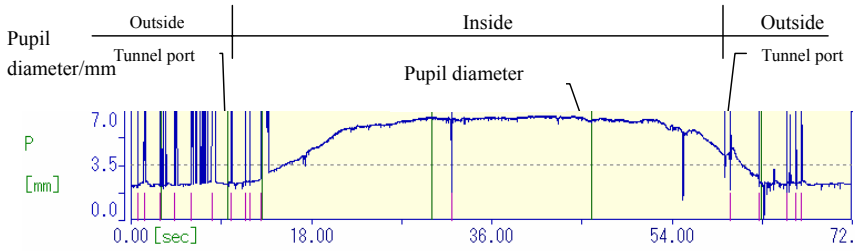


Figure 1. Pupil diameter's variety during passage of tunnel in daytime

(2) Variety velocity of pupil area

During the tunnel process, the drivers' variety velocity of pupil area are shown as Figure 2.

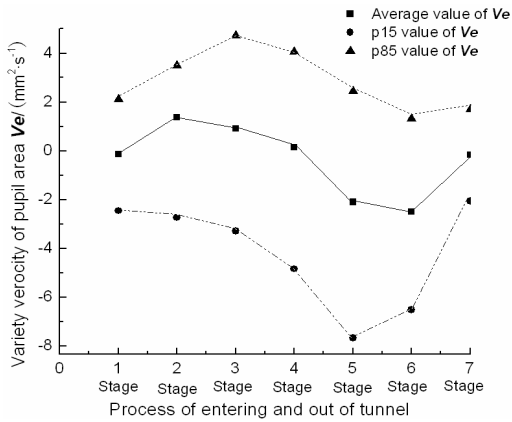


Figure 2 Variety velocity of pupil area during tunnel stage

Notes: ① the measure's signification of abscissa were shown as follows:

Stage 1— The normal highway stage about driving distance of 15 seconds before tunnel's entrance.

Stage 2— The light adaptation process at the tunnel port (during 50m'distance of entrance inside of the tunnel) .

Stage 3— The light adaptation process inside the tunnel (the distance from the tunnel' entrance port to the spot that the finished variety percent of pupillary area is 85% inside the

tunnel) .

Stage 4— The middle stage inside the tunnel (the driving stage between the light adaptation process and the dark adaptation inside of the tunnel).

Stage 5— The dark adaptation process inside the tunnel (the distance from the spot that the finished variety percent of pupillary area is 85% to the tunnel's exit port) .

Stage 6— The dark adaptation process at tunnel port (during 50m'distance of exit inside of the tunnel) .

Stage 7—The normal highway stage about driving distance of 15 seconds after tunnel's exit.

② P15 value of Ve—15% value of variety velocity of pupil area.

P85 value of Ve—85% value of variety velocity of pupil area.

To tunnel's entrance and exit(Stage 2 and Stage 6), the primary study indicated that the driver' s pupil area has a power functional relationship with the illuminance of tunnel's entrance and exit.

(3) Variety velocity index of pupil area

To the varieties of speed of pupil area, Stage 6> Stage 5> Stage 2> Stage 3, denotes light adaptation is faster than dark adaptation, and the varieties of speed of pupil area during tunnel's entrance and exit is wide, and the corresponding driving load is big.

Owing to the sharp light and dark transition during tunnel's entrance and exit, the varieties of speed of pupil area increase rapidly, and driving load also increases greatly. When the variety of speed of pupil area is too large for drivers to adapt, the pupil finds it difficult to focus accurately for a vivid image on the retina, then a serious visual obstacle may happen; the driver even can't distinguish the road and this will lead to traffic accident.

Owing to driver's limit capacity of visual adaptation, when the variety velocity is out of the critical region [V_{ecr},V_{eci}], the visual obstacle may occur, and the driving is not safe. Inside the critical region, the visual obstacle may not occur, and the driving is safe. Taking 85% value of variety velocity of pupil area for critical velocity of pupil area, the critical velocity of pupil area during tunnel's entrance and exit can be obtained from Figure 2:

V_{eci}= 3.5~4.7mm^2 • s^{-1}（tunnel's entrance）and V_{ecr}=-6.5~-7.6mm^2 • s^{-1}（tunnel's exit）, the negative sign indicates the pupil area's decrease.

To the tunnel entrance in the daytime, the enlargement of the pupil area indicates the increase of the visual load mentally physically, both the psychological load and physical visual load have positive affect on the enlargement velocity of pupil area; to the tunnel exit, the decrease of the pupil area makes the psychological load decrease, and the nervousness degree reduce, but the physiological load increase rapidly, and at the leading position. In conclusion, using variety velocity of pupil area to evaluate the degree of visual load during tunnel's entrance and exit is rational and efficient, more efficient than pupil diameter index and pupil area index.

On the tunnel's entrance and exit, decreasing driving speed sign and transition

slope of illuminance can be adopted to reduce the variety velocity, finally ensure driving safety.

4.3 Variety regulation of moving fixation

(1) Index of moving fixation

The Japanese research of Fukuda etc. hold that when the fixation duration exceeds 0.165s, we will watch the same object continuously. Furuichi discovered the relationship between average fixation duration and average fixation times and denoted that the relationship can be adopted as the index to evaluate the driving behaviors and the planar alignment curved section.

So we can define the product of average fixation duration and average fixation times as the index of moving fixation, which shows the correlative degree to the target object; and it can evaluate the distribution regulation of moving fixation, and can be adopted as the evaluating index of information processing level. It can be shown as follows:

$$a = T_a \cdot N_a$$

$$T_a = T_t / N_t$$

$$N_a = N_t / T_l$$

a——Index of moving fixation,

T_a——Average fixation duration of individual (s),

N_a——Average fixation times of unit time (s^{-1}),

T_l——Duration of target stage's passage (s),

T_t——Accumulated fixation duration during the target stage (s),

N_t——Fixation times of the target stage (s^{-1}),

The drivers' processing of traffic information is divided into primary processing （shown as saccade） and profound processing（shown as fixation of movement）. If a is small it denotes the lowness of the information processing level, namely the drivers' fixation duration is short and saccade duration is long; the reason is the shortage of traffic information or the low range of drivers' visual psychology and physiology; and if a is big, it denotes the height of the information processing level, owing to the height of the visual load or the low range of drivers' visual load mentally and physically.

(2) Experimental result of moving fixation

The index of moving fixation during tunnel stage is shown in Figure 3 and Table 1. From Figure 3, drivers' visual information processing level of various stages reads: Stage 2< Stage 3< Stage 6< Stage 5< Stage 4. The results indicate: the visual information processing level of dark adaptation process is lower than the light adaptation inside of the tunnel, the visual information processing level of dark and light adaptation process during tunnel port are lower than inside the tunnel. If traffic information is large and the driver's visual information processing level is low, the traffic accidents are most likely to happen at the tunnel's entrance and exit. The primary study indicates: index of moving fixation has positive relationship with

driving velocity, and negative relationship with luminance of traffic environment.

The parameter's rational range of moving fixation in Table 1 are [P15, P85], if the parameter is not in the range, indicates the driving is not comfortable.

From Figure 3 and Table 1, the index of moving fixation during 50m'distance of tunnel's entrance（Stage 2）is much lower than inside the tunnel and normal highway stage. Experimental results indicates: 45.8% experimental samples have visual shock (namely the sharp fluctuation of pupil diameter that the apparatus can't capture), lead to rapid decline of drivers' visual information processing level. Figure 3 indicates that among the P85-P15 values of various stages, tunnel entrance and tunnel exit are bigger; it also shows the big diversity of experimental samples' visual information processing level, which is the major cause of the high accident rate at the entrance and exit of a tunnel.

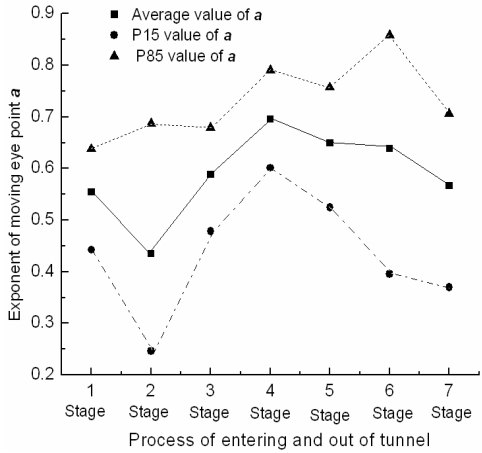


Fig.3 Index variety of moving fixation during tunnel section

Notes: P15 value of ***a***—15% value of index of moving fixation during the stage.

P85 value of ***a***—85% value of index of moving fixation during the stage.

Table 1. Main parameters of moving fixation during tunnel stage

Project / Stage	***Ta***/s			***Na***/s^{-1}			***a***			Sample number ***n***
	Average value	P15 value	P85 value	Average value	P15 value	P85 value	Average value	P15 value	P85 value	
Stage 1	0.314	0.265	0.359	1.806	1.650	2.000	0.555	0.443	0.638	38
Stage 2	0.297	0.222	0.400	1.457	0.967	2.044	0.435	0.246	0.688	36
Stage 3	0.342	0.293	0.397	1.753	1.454	2.026	0.588	0.479	0.680	36
Stage 4	0.370	0.300	0.468	1.943	1.627	2.208	0.696	0.601	0.791	36
Stage 5	0.358	0.290	0.443	1.883	1.379	2.206	0.649	0.524	0.758	38
Stage 6	0.323	0.271	0.388	2.042	1.416	2.722	0.639	0.395	0.858	36
Stage 7	0.316	0.263	0.359	1.787	1.467	2.067	0.566	0.370	0.707	36

5 Conclusions

Here we may draw the following conclusions.

(1) The evaluation index of driving safety based on the variety velocity of pupil area is a rational index to drivers' visual load degree at the entrance and exit of a tunnel.

(2) Index of moving fixation is brought up to evaluate the visual information processing level.

(3) The rational range of the variety velocity of pupil area and the index of moving fixation were obtained.

6 Recommendations for Future Research

In further research, the elderly and women driving samples should be argued, and the night experiments should be reinforced to improve the universality and practicability of the corresponding index.

References

Beatty, J. (1982). "Task-evoked pupillary responses, processing load, and the structure of processing resources." *Psychological Bulletin,* 91, 276-292.

Fukuda, R. (1996). Experimental considerations on the definition of eye fixation points. *Journal of Japan Ergonomics Society,* 28(2), 197-204.

Pan Xiaodong. (2005). "The attentive point of driver's watch and the research of traffic safety experiment." *The 7st Annual Conference 's Academic Collections of The Shanghai Road Academy*, 255-260.（in Chinese）

Tomoki, F. "Eye movement and driving behavior in curved section passages of urban motorway." *TRB 2005 Annual Meeting CD-ROM*.

Vernep, S. P. (2004). "Pupillary responses on the visual backward masking task reflect general cognitive ability." *International journal of Psychophpsiology,* 52, 23-36

Ying Ruihua. (2003). "Discussing the safety design of highway tunnel's alignment." *Highway*, 22(3), 140-142. （in Chinese）

SAFETY ANALYSES AT SIGNALIZED INTERSECTIONS: RESEARCH STRATEGIES, MODELING TECHNIQUES, AND SIGNIFICANT FACTORS

Xuesong WANG[1], Mohamed Abdel-Aty[2], and Xiaohong CHEN[3]

[1] Department of Civil & Environmental Engineering, University of Central Florida, Orlando, FL 32816-2450; PH (407) 823-4902; FAX: (407) 823-4676; email: xu297341@ucf.edu
[2] Department of Civil & Environmental Engineering, University of Central Florida, Orlando, FL 32816-2450; PH (407) 823-5657; FAX: (407) 823-4676; email: mabdel@mail.ucf.edu
[3] Department of Civil & Environmental Engineering, University of Central Florida, Orlando, FL 32816-2450; PH (86-21) 6598-2127; FAX: (86-21) 6598-2274; email: chenxh@mail.tongji.edu.cn

Abstract: Signalized intersections are among the most dangerous locations in a roadway network. The first step towards an effective solution to safety improvement is to identify the primary causes of crash occurrence. Negative Binomial regression provides a common tool for modeling cross-sectional count data like crash frequencies at signalized intersections, which assume the dependent variables are independent. Intersections could be considered as isolated when the distance between them is long. Signalized intersections, especially for those closer ones along a certain corridor, are spatially correlated and will influence each other in many aspects. Generalized Estimating Equations, which is an extension of generalized linear models to correlated data, is applied to account for spatial correlation for the safety analyses at the intersections. Rear-end crashes are the most frequently occurring collision type at signalized intersections. The paper presents a series models developed in the United States for total crashes and rear-end crashes at signalized intersections. Intersection geometric design features, traffic control and operational features, traffic characteristics, land use, and some corridor level factors are collected and their safety effects are tested. The significant factors at the different levels are identified which lead to a better understanding of crash occurrence and eventually to develop efficient safety countermeasures for signalized intersections.

Keywords: *signalized intersection, safety modeling, spatial correlation, total crashes, rear-end crashes, significant factors*

1 Introduction

Intersections are among the most dangerous locations of a roadway network. In the year 2003, Florida had 243,295 crashes recorded in the Florida Department of

Highway Safety and Motor Vehicles (DHSMV) database. Among these, 96,710 crashes (39.75%) occurred at or were influenced by intersections. Intersections also tend to experience more severe crashes. Since the signalized intersections are generally large intersections, the safety status for these intersections is even worse. It is well accepted that traffic crashes at signalized intersections are complicated events which involve the interaction between the driver, vehicle, roadway, traffic, and the environment. However, the relationship between crash occurrence and these contributing factors is still not clear. The first step towards an effective solution to improve safety is to identify the primary causes of crash occurrence. Several approaches have been used to establish the relationship between crash occurrence and intersection characteristics. The most common approach is modeling crash frequencies in a certain period at signalized intersections by geometry and traffic related explanatory variables. Crash frequencies at intersections are count data, which are nonnegative, integer-valued outcomes.

Corridors play a prominent role in the moving of people and goods between major points in both urban and rural areas. Signalized intersections are the most complex locations along corridors, and they are vital to efficiency and safety of corridors. The signalized intersections along a certain corridor will affect each other in the following aspects especially for those nearest ones. Corridors usually serve relatively long trips between major points, and several adjacent signalized intersections along a certain corridor will share a high percentage of the same traffic; signals within 0.5 mile of each other along a corridor would be coordinated in most circumstances (FHWA, 2004), and this coordination in signals will promote platooning of vehicles crossing intersections; and adjacent intersections along a corridor probably have similar types of land use and roadway design. In order to improve both the safety as well as operation of these transportation systems, there is a need to look at the spatial relationship for signalized intersections along a corridor rather than treat each intersection as an isolated entity.

Most of the safety studies at signalized intersections treated intersections as isolated entities. Negative Binomial regression provides a common tool for modeling cross-sectional count data like crash frequencies at signalized intersections. There are serious problems arising when we use traditional modeling methods for clustered data. Basic count data models assume the dependent variables are independent. This is valid for most cross-sectional crash data, but not for spatially correlated data. The use of Negative Binomial for the spatially correlated data would lead to invalid statistical inference due to incorrect reported test statistics and standard errors based on the misspecified variance (Abdel-Aty and Wang, 2006; Wang et al., 2006). Generalized Estimating Equations (GEEs) provide an extension of GLMs to the analysis of longitudinal or clustered data, which account for the correlation among the repeated observations by specifying a working correlation matrix of the observations for each cluster (Liang and Zeger, 1986; Zeger and Liang, 1986). The GEE method give consistent estimates of the regression parameters and of their variance for longitudinal data under mild assumptions about the dependence (Liang and Zeger, 1986). Wu (1998) proved that the GEE is a robust modeling method for spatially correlated data.

It is reasonable to assume that the different crash types have different contributing factors; fitting total crashes at intersection can not reveal the different

effects of geometric, traffic, and environmental features on different crashes. For example, it has been found that installing traffic signal will reduce angle and left-turn crashes but increase rear-end crashes due to frequent sudden stops at signals. Rear-end crashes occur when the front of a vehicle strikes the rear of a leading vehicle. They are the most frequently occurring collision type at signalized intersections; rear-end crashes represent 40.2% of all reported intersection crashes based on the crash history of 1531 signalized intersections in the state of Florida (Abdel-Aty et al., 2005), which means that rear-end crashes are a real problem at signalized intersections.

For the explanatory variables included in intersection crash models, traffic flow is the most powerful variable (Hauer et al., 1988). But intersection geometric design features (i.e. number of through lanes, right-turn lanes, left-turn lanes, etc.) and traffic control and operational features (i.e. signal phase, speed limit, etc) are also important (Poch and Mannering, 1996; Wang et al., 2006). In order to estimate the model consistently and efficiently, there is a need to apply GEE to account for the correlation among the data. Moreover, the improvement of both safety and operation of signalized intersections along corridors is a major concern for many studies. It is widely accepted that the spatial effect is important to the improvement of safety at signalized intersections along corridors, and a variety of treatments have been proposed in corridor-level, such as signal coordination (Antonucci et al., 2005; FHWA, 2004); however, the spatial relationship among intersections along corridors has not been explored in the safety literature.

The primary objectives of this paper are to develop statistically valid models to analyze the spatial effect among signalized intersections along corridors and to identify variables which significantly influence total crashes and rear-end crashes at signalized intersections. In China, intersections are also the most dangerous locations of a roadway network. For example, in Shanghai Lujiazui Area, on the year 2005 crash spot map, around 80% of traffic crashes occurred at intersections; therefore, the research strategies and the modeling techniques summarized and applied and significant factors identified in this study could be also useful in China.

2 Data Preparation

In order to perform intersection safety and spatial analysis, one needs to select a variety of corridors as well as signalized intersections possessing different geometric and traffic characteristics. Geographic information system (GIS) is a powerful tool to visualize and explore spatial data and is used extensively in this study in corridor and signalized intersection selection and intersection sequence identification. Some corridor-wide features like the design aspects of access points near intersections are retrieved using software that combines satellite imagery and road maps.

A total number of 476 signalized intersections along 41 principle and minor arterials are selected from three counties in the state of Florida: Orange, Brevard, and Miami-Dade Counties. Most of the selected arterials are state roads and/or multilane highways with high speed limits and which serve relatively long trips between major points. Figure 1 is the GIS map for Orange County; the grey lines represent the road network and the points represent the selected intersections. Since the distance between some intersections along some corridors is very long and the number of intersections

along certain corridors is extremely large, it is reasonable to divide them into sub-clusters in which intersections group together. In this analysis, the SAS MODECLUS procedure is used to divide intersections along a certain corridor into sub-groups. The distance between intersections is considered for grouping the intersections. The number of intersections in clusters varies from 1 to 13; the data are unbalanced.

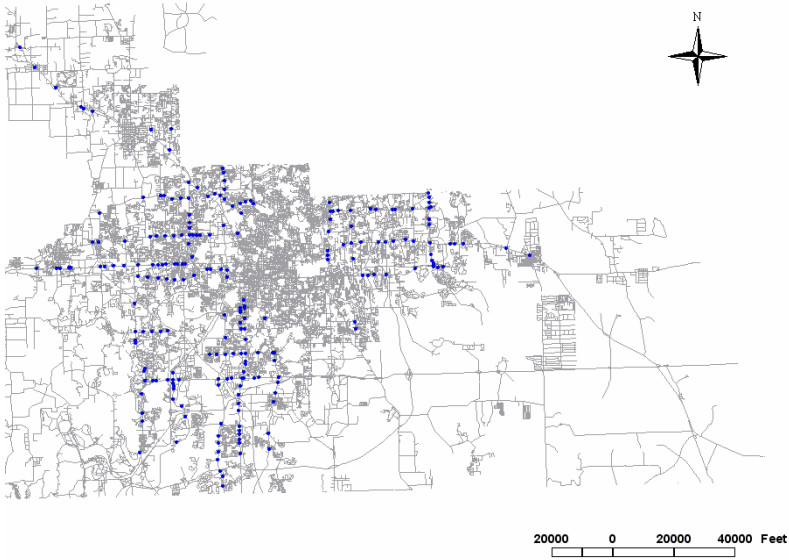


Figure 1. Road networks and geocoded signalized intersections for Orange County.

Crashes that occurred at the intersections for two years (1999 and 2000) were collected by retrieving the Crash Analysis Reporting (CAR) system for 386 state road intersections (at least one intersecting roadway is a state road) and by using the county maintained crash database for 90 county road intersections. Crashes considered in this analysis were ones occurring within 250 feet of the intersection milepost and labeled 'at intersection' or 'influenced by intersection' for crash site location.

Traffic characteristics for intersections included entering traffic flows for each approach. The sum of traffic volumes on each approach is the total entering ADT for the intersection. Dividing the total entering ADT by the total number of lanes (including through, left-turn and right-turn lanes) is ADT per lane (ADTPL), which is an indicator of intensity of traffic at the intersection.

Intersection geometric design features and traffic control and operational features in the study period are extracted from the intersection traffic planning and design diagrams provided by the three counties. Hundreds of drawings are then individually examined and identified. The geometric design features of the intersection include:

number of through, left, and right lanes for each approach; the presence of exclusive turn lanes at each approach; and the presence of a median at each approach. The traffic control and operational features include: speed limit for each approach and signal timing. The intersection location type has three categories: primarily business, primarily residential, and open country.

3 Methodology: Modeling Spatial Correlation Using GEEs

The Generalized Estimating Equations (GEEs) come from specifying a known function of the marginal expectation of the dependent variable as a linear function of covariates, assuming that the variance is a known function of the mean, and in addition, specifying a "working" correlation matrix for the observations for each cluster (Liang and Zeger, 1986; Zeger and Liang, 1986). Let y_{ij} represent the jth observation on the ith subject, for $i = 1,2,\ldots,K$ and $j = 1,2,\ldots,n_i$. In this study, y_{ij} represent the two-year crash frequency (or two-year rear-end crash frequency) for intersection j in corridor i. Define K as the total number of corridors, and n_i is the number of intersections in corridor i; the number of intersections per corridor varies and the data are unbalanced and there are $\sum_{i=1}^{K} n_i$ total intersections.

Let the vector of crash frequency for the ith corridor be $Y_i = \left(y_{i1},\Lambda\ ,y_{in_i}\right)'$ with corresponding means $\mu_i = \left(\mu_{i1},\Lambda\ ,\mu_{in_i}\right)'$ and V_i is an estimator of the covariance matrix of Y_i. Suppose $x_{ij} = \left(x_{ij1},\Lambda\ ,x_{ijp}\right)'$ denote a $p\times 1$ vector of explanatory variables associated with y_{ij}. The GEE for estimating β is an extension of the GLMs to the correlated data. The link function and linear predictor setup is as regular GLMs and is given by

$$S(\beta) = \sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial \beta} V_i^{-1}(Y_i - \mu_i(\beta)) = 0 \tag{1}$$

Since $g\left(\mu_{ij}\right) = x_{ij}'\beta$, where g is the link function. The $p\times n_i$ matrix of partial derivatives of the mean with respect to the regression parameters for the ith subject is given by

$$\frac{\partial \mu_i'}{\partial \beta} = \begin{bmatrix} \frac{x_{i11}}{g'(\mu_{i1})} & \cdots & \frac{x_{in_i1}}{g'(\mu_{in_i})} \\ \mathrm{M} & & \mathrm{M} \\ \frac{x_{i1p}}{g'(\mu_{i1})} & \cdots & \frac{x_{in_ip}}{g'(\mu_{in_i})} \end{bmatrix} \tag{2}$$

The covariate matrix of Y_i is specified as the estimator $V_i = \phi A_i^{\frac{1}{2}} R_i(\alpha) A_i^{\frac{1}{2}}$, where A_i is a $n_i \times n_i$ diagonal matrix with $v\left(\mu_{ij}\right)$ as the jth diagonal element. V_i can be different from subject to another, but generally is to specify the same form of V_i for all subjects.

$R_i(\alpha)$ is a $n_i \times n_i$ working correlation matrix that is fully specified by the vector of parameters α.

Independent, exchangeable, and autoregressive correlation matrix are the common working correlation structures. The goal of selecting an appropriate working correlation structure is to estimate parameter more efficiently (Ballinger, 2004). The independent correlation structure assumes that repeated observations for a subject are independent. The exchangeable working correlation makes constant the correlations between any two observations within a subject. The autoregressive structure weighs the correlation between two observations for a subject by their separated gab (order of measure); as the distance increases the correlation decreases. The model-based estimator of $Cov(\hat{\beta})$ is then given by $S_m(\hat{\beta}) = I_0^{-1}$, where $I_0 = \sum_{i=1}^{K} \frac{\partial \mu_i'}{\partial \beta} V_i^{-1} \frac{\partial \mu_i}{\partial \beta}$.

The GEE estimates are obtained when a quasi-likelihood technique is used; therefore, the goodness-of-fit tests for the traditional negative binomial regression are not valid for a GEE negative binomial. Zheng (2000) introduced a simple extension of R-square statistics for GEE models as "marginal R-square". The marginal R-square is interpreted as the amount of variance in the response variable that is explained by the fitted model.

4 Estimation Results

Intersections within a cluster are assumed to be spatially correlated, and intersections from different clusters are assumed to be statistically independent. The Generalized Estimating Equations (GEEs) with a Negative Binomial link function for total crash and rear-end crash frequency per intersection with the three correlation structures (independent, exchangeable, and autoregression) were fitted separately using SAS GENMOD procedure (SAS Institute Inc., 2004). The marginal R-square statistics were calculated for the GEE models. Associated spatial correlation structures for GEE models were estimated.

4.1 Spatial safety analysis for total crashes

As an initial study, the Negative Binomial regression model was fitted using maximum likelihood estimation (MLE). As shown in Table 1, the estimated coefficients and standard errors for the Negative Binomial regression and the GEE models with different correlation structures are different. For example, the left-turn protection on major roadway is not significant in the Negative Binomial (p-value = 0.2561), while it is significant in the GEE models with the exchangeable or autoregression structures (p-value < 0.10). Based on the misspecified variance, the reported test statistics and standard errors for the Negative Binomial are not valid for the correlated data. The three correlation structures have produced different coefficients and standard errors, which show the effect of different correlation structures in the analysis.

The marginal R-square statistics are reported in. The GEE model with autoregression structure has a higher marginal R-square value (0.391), which indicates that the autoregression structure could be the appropriate structure for spatial

correlation. For the spatial correlation of signalized intersections along corridors, as the spacing between intersections increases, it is reasonable to assume the correlation between them decreases. Since the cluster size varies from 1 to 13, the dimension of the correlation matrix is 13 with one in each diagonal position. The autoregression structure has a maximum correlation of 0.305 for any two successive intersections along a corridor; the correlation between intersections decreases as the spacing between intersections increases, and it reduces to zero after 9 intersections (although it is reduced significantly after 4 intersections). The correlation estimated by exchangeable structure is 0.247.

4.2 Spatial safety analysis for rear-end crashes

The GEE models with a Negative Binomial link function for rear-end crash frequency in two years were fitted for independent, exchangeable, and autoregression correlation structures as reported in Table 2. The three correlation structures have produced different coefficients and standard errors, which show the effect of different correlation structures in the analysis. The GEE model with autoregression structure has a slightly higher marginal R-square value (0.4591), which again indicates that the autoregression structure could be the appropriate structure for spatial correlation. The autoregression structure has a maximum correlation of 0.6313 for any two successive intersections along a corridor; the correlation between intersections decreases as the spacing between intersections increases.

Traffic volumes on major and minor roadways are the most significant variables. Among the intersection geometric design features, the number of legs, the presence of exclusive right-turn lanes on both roadways, and the number of left-turn lane on major roadway are significant to rear-end crash occurrence. The left-turn protection on major roadway will reduce rear-end crashes, while the protection on minor roadway will increase rear-end crash occurrence. The intersection with a median on the minor roadway has a lower number of rear-end crashes. The posted speed limit on the major roadway is significant for rear-end crashes. The logarithm of the average length of upstream and downstream segments of an intersection is identified to be the most significant variable. The negative sign for this factor (coef. = -0.1246) indicates that the effect of both neighboring signals (not just upstream or downstream segment) decreases as the distance increases.

5 Summary

This study is an empirical inquiry of the spatial correlation among the crash data. A massive data collection effort has been completed. The data for 476 signalized intersections which are located along different corridors were collected in the state of Florida. The total crash and rear-end crash frequencies were modeled using the Generalized Estimating Equations (GEEs) with a Negative Binomial link function separately, and the different working correlation structures (independent, exchangeable, and autoregressive) have been explored. The GEE autoregression models assuming that the correlations between the multiple observations for a certain intersection or intersection cluster will decrease as the space gap increases are better for spatially correlated crashes.

It was found that for total crashes the estimated correlation is 0.3054 for two nearest intersections along corridors; for rear-end crashes, it is 0.6313, which is relatively high. The estimates have been modified when considering the spatial correlation. The use of basic models for such correlated data may produce biased estimators and invalid test statistics. One example is that the insignificant variable, the left-turn protection on major roadway, was found significant in the GEE models with the autoregression structure.

Turning to the significant variables, traffic volume is the most significant variable affecting intersection safety. This conclusion has been proven by many studies using panel data methods (Wang et al., 2006) and basic count data methods (Poch and Mannering, 1996). In this study, LogADTPL was used for total crash models and it came to be one of the most significant variables. Traffic volume is just the quantity of vehicles crossing the intersection while the ADT per lane is an indicator of the intensity of vehicles at the intersection. For rear-end crash models, this study showed the logarithm transformation of traffic volumes on major and minor roadways are the better functional forms.

The size of the intersection, the number of legs, and the presence of exclusive right-turn lanes on both roadways are identified significant influencing total crash occurrences. Right-turning traffic can have an adverse effect on intersection safety. The spatial analysis shows that the installation of right-turn lanes on major and minor roadways would be expected to reduce total intersection crash occurrence significantly. In rear-end crash study, it is found that compared to the shared right-turn lane, the channelized and exclusive right-turn lanes on the minor roadway reduce rear-end crashes.

The left-turn protection is found to be significant to affect safety at intersections. Signalized intersections, where a corridor connects with local roads or intersects with other arterials, usually have a high amount of through traffic down the corridor with high speed. Crashes may occur when left-turning drivers become impatient and accept a gap that is smaller than needed to complete a safe maneuver (FHWA, 2004). Protection of left-turning traffic from opposing through traffic on a corridor at intersections will reduce the crashes, which is indicated by a negative sign for the factor of having at least one approach protected on major road. However, the left-turn protection on minor roadway is associated with higher crashes indicated by the positive coefficient 0.2161 in the GEE model. The reason might be that the more approaches with left-turn protection determine the larger number of phases per cycle. The safety advantage of traffic signal control is to reduce the frequency and severity of certain types of crashes, e.g. right-angle, turning and pedestrian/bicycle, which tend to be severe; while the disadvantage is that left protection might cause an increase in rear-end crashes, which is confirmed by the finding in the rear-end study that protecting left-turning movement on minor roadways will increase rear-end crashes.

For signal spacing, the distance to the nearest signals is identified to be the most significant variable to affect the total crash occurrence. The negative sign for this factor indicates that the effect of the nearest signal decreases as the distance between them increases. It was found that the average distance to the neighboring signals along corridors is identified to be significant to affect rear-end crash occurrence. These

findings indicate that intersections along corridors affect each other and should not be considered in isolation.

The results of these studies can be applied to diagnose safety problems and to develop efficient countermeasures for specific deficiencies at signalized intersections. The developed models can be applied to estimate the average total crash and rear-end crash frequencies for an intersection of interest. Its safety status can be evaluated and diagnosed by contrasting its crash history data with the estimated values. Then specific safety problems can be identified by comparing the observed crashes with the average crash frequencies. Based on the significant factors in the models, the factors affecting the crash problem can be identified. The efficient countermeasures can then be developed to improve safety at that intersection based on the identified factors that affect this precise problem.

Table 1: Spatial analyses model estimates for negative binomial and GEE negative binomial with different correlation structures.

Parameter	Negative Binomial Estimates (s.e.)	GEE Estimates (s.e.)		
		Independent	Exchangeable	AR-1
Intercept	-1.4352 (0.9099)	-1.4352 (0.9313)	-0.9682 (1.1034)	-0.7881 (1.0118)
Logarithm of total entering ADT per lane (LogADTPL)	0.4774 (0.1099)	0.4774 (0.1109)	0.3834 (0.1278)	0.3804 (0.1178)
Size of intersection (Total number of lanes)	0.1346 (0.0193)	0.1346 (0.0182)	0.1407 (0.0182)	0.1387 (0.0179)
Intersection configuration				
3-legged	-0.325 (0.1204)	-0.325 (0.1142)	-0.2124 (0.114)	-0.3117 (0.1106)
4-legged	0	0	0	0
Right-turn lanes on Major roadway				
At least one approach has exclusive right-turn lane	-0.2694 (0.1156)	-0.2694 (0.1113)	-0.2172 (0.1133)	-0.2859 (0.1089)
No exclusive right-turn lane	0	0	0	0
Right-turn lanes on Minor roadway				
At least one approach has exclusive right-turn lane	-0.2766 (0.106)	-0.2766 (0.1042)	-0.2463 (0.0996)	-0.2552 (0.095)
No exclusive right-turn lane	0	0	0	0
Left-turn protection on Major roadway				
At least one approach has protected left-turn lanes	-0.1649 (0.1452)	-0.1649 (0.1402)	-0.284 (0.1408)	-0.2493 (0.1371)
No protected left-turn lanes	0	0	0	0
Left-turn protection on Minor roadway				
At least one approach has protected left-turn lanes	0.2102 (0.1078)	0.2102 (0.1026)	0.2232 (0.1001)	0.2161 (0.0977)
No protected left-turn lanes	0	0	0	0
Speed limit on corridor near intersection (mph)	0.0234 (0.0088)	0.0234 (0.008)	0.0168 (0.0082)	0.0189 (0.0079)
Logarithm of distance to the nearest signalized intersection along corridor (ft)	-0.1996 (0.0569)	-0.1996 (0.0551)	-0.1569 (0.0578)	-0.1759 (0.0554)
Intersection location types				
Open country	-0.8943 (0.5113)	-0.8943 (0.4967)	-0.557 (0.4574)	-0.629 (0.3765)
Primarily residential	-0.3125 (0.1187)	-0.3125 (0.113)	-0.2571 (0.113)	-0.249 (0.1122)
Primarily business	0	0	0	0
Scale	0.6245 (0.0499)	0.9623	0.9759	0.9479
Summary Statistics				
Number of corridors		41	41	41
Number of intersections	476	476	476	476
Number of clusters		116	116	116
Minimum cluster size/Maximum cluster size		1/13	1/13	1/13
Sum of initial crashes	8731	8731	8731	8731
Marginal R-square statistics		0.375	0.381	0.391

Table 2: Model estimates for the spatially correlated rear-end crashes.

Parameters	NB Estimate (s.e.)	GEE Negative Binomial Estimate (s.e.)		
		Independent	Exchangeable	Autoregression
Intercept	-7.7926 (1.1152)	-7.7926 (1.2275)	-7.6801 (1.4784)	-4.5202 (1.2874)
Logarithm of ADT on Major roadway	0.48 (0.1265)	0.48 (0.1386)	0.4888 (0.1566)	0.3854 (0.1338)
Logarithm of ADT on Minor roadway	0.5124 (0.1038)	0.5124 (0.1152)	0.4449 (0.114)	0.2826 (0.0848)
Intersection Configuration				
3-legged	-0.427 (0.1163)	-0.427 (0.1263)	-0.3486 (0.1283)	-0.2873 (0.0949)
4-legged	0	0	0	0
Right-turn lanes on Major roadway				
At least one approach has an exclusive right-turn lane	0.2607 (0.1107)	0.2607 (0.1186)	0.2213 (0.1234)	0.2215 (0.0972)
No exclusive right-turn lane	0	0	0	0
Right-turn lanes on Minor roadway				
At least one approach has an exclusive right-turn lane	-0.4243 (0.1028)	-0.4243 (0.108)	-0.3859 (0.1038)	-0.1288 (0.0778)
No exclusive right-turn lane	0	0	0	0
Number of left-turn lanes on Major roadway				
More than 2	0.5804 (0.1517)	0.5804 (0.1587)	0.6263 (0.156)	0.3802 (0.1233)
Less than or equal to 2	0	0	0	0
Left-turn protection on Major roadway				
At least one approach has protected or partially protected left-turn lane	-0.3254 (0.1301)	-0.3254 (0.1461)	-0.3717 (0.1462)	-0.1763 (0.1078)
No protected left-turn lane	0	0	0	0
Left-turn protection on Minor roadway				
At least one approach has protected or partially protected left-turn lane	0.3969 (0.1039)	0.3969 (0.1111)	0.4377 (0.1097)	0.2435 (0.078)
No protected left-turning movement	0	0	0	0
Median on Minor roadway				
With median	-0.2356 (0.1032)	-0.2356 (0.1094)	-0.1828 (0.1099)	-0.1527 (0.083)
Without median	0	0	0	0
Logarithm of the average distance of upstream and downstream segments of an intersection along corridor	-0.1408 (0.063)	-0.1408 (0.0669)	-0.094 (0.0712)	-0.1246 (0.0595)
Speed limit on Major roadway (mph)	0.0196 (0.0097)	0.0196 (0.0099)	0.0216 (0.0102)	0.0131 (0.0081)
Dispersion	0.6044 (0.0527)	1.0792	1.0886	1.058
Summary Statistics				
Number of corridors	41	41	41	41
Number of intersections	476	476	476	476
Number of clusters	-	116	116	116
Minimum cluster size / Maximum cluster size	-	1/13	1/13	1/13
Marginal R-square Statistics	-	0.4069	0.4560	**0.4591**

Reference:

Abdel-Aty, M., Lee, C., Wang, X., Nawathe, P., Keller, J., Kowdla, S., Prasad, H. (2005b). *Identification of Intersections' Crash Profiles/Patterns,* Final Report, Florida Department of Transportation.

Abdel-Aty, M. and Wang, X. Crash Estimation at Signalized Intersections along Corridors: Analyzing Spatial Effect and Identifying Significant Factors. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1953*, TRB, National Research Council, Washington, D.C., 2006, pp. 98-111.

Antonucci, N., Hardy, K., Slack, K., Pfefer, R., and Neuman, T. (2005). *NCHRP Report 500: A Guide for Reducing Collisions at Signalized Intersections.*

Ballinger, G. (2004). "Using Generalized Estimating Equations for Longitudinal Data Analysis." *Organizational Research Methods*, Vol. 7, No. 2, 127 – 150.

Federal Highway Administration [FHWA]. (2004). *Signalized Intersections: Informational Guide* (Rep. No. FHWA-HRT-04-091). Washington, DC: USDOT, FHWA.

Hauer, E., Ng, J., and Lovell, J. (1988). "Estimation of safety at signalized intersections." *Transportation Research Record 1185*, 48–61.

Liang, K. and Zeger, S. (1986). "Longitudinal Data Analysis Using Generalized Linear Models." *Biometrika*, 73, 13-22.

Poch, M. and Mannering, F. (1996). "Negative Binomial Analysis of Intersection-Accident Frequencies." *Journal of Transportation Engineering*, 122, 105-113.

SAS Institute Inc. (2004). *SAS OnlineDoc® 9.1.2.* Cary, NC: SAS Institute Inc.

Wang, X., Abdel-Aty, M., and Brady, P. Crash Estimation at Signalized Intersections: Significant Factors and Temporal Effect. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1953*, TRB, National Research Council, Washington, D.C., 2006, pp. 10-20.

Wu, C.-T. (1998). *Generalized estimating equations for spatially correlated data*. Dissertation, Department of Statistics, North Carolina State University.

Zeger, S. and Liang, K. (1986). "Longitudinal Data Analysis for Discrete and Continuous Outcomes." *Biometrics*, 42, 121-130.

Zheng, B. (2000). "Summarizing the Goodness of Fit on Generalized Linear Models for Longitudinal Data." *Statistics in Medicine*, 19, 1265-1275.

RESEARCH ON TRAIN OPERATING SCHEME OF INTERCITY RAIL TRANSIT

Xingfang XU[1], Jingsong XIANG[2]

[1]Department of Transportation Management Engineering, Tongji University, Shanghai. 201804; email: xfx@mail.tongji.edu.cn

[2]Shenzhen Urban Transport Planning Center, Shenzhen, 518034; email: pinetree001@163.com

Abstract：Based on the investigation and analysis of intercity passenger transportation market, the paper researched on some basic problems of intercity rail transit (IRT), such as the types of intercity train (IT), the ratio of carrying traffic, etc. The paper optimized the existing methods of train operating scheme (TOS), then brought up a new model to work out the scheme gradationally, taking Hu-Ning IRT in building as an example, which validated the model as a feasible and practical one.

Key words: *intercity rail transit, intercity train, train operating scheme*

1 Introduction

With the fast developing economy, city clusters are shaping in China. The frequency of intercity trip is steadily becoming denser and the demands on travel service are higher and higher, so a "large capacity, high quality" transit mode is an urgent need. Intercity Rail Transit (IRT) is a good choice to meet these requirements in regional passenger transportation.

Intercity train (IT) is one of the public transport vehicles for passengers traveling among cities in city cluster, which operates in a high density, higher speed, shorter distance and orderly way.

The research on operating scheme of IT is not only propitious for the operating theory on rail transit, but also for meeting the strong traffic need, for improving transportation service, for enlarging the market ratio of passenger rail transit, etc.

2 Analysis of transit market

2.1 Classification of passenger

According to the investigation and analysis of passengers' characters in Yangtze River delta Region, passengers generally have higher demand on departure time, higher density of IT, higher traveling speed and higher comfort of trip. Meanwhile, having different characters in endurance for ticket price, time value and comfortableness, passengers could be classified into four types whose ratio is 2：4：3：1 (see tab. 1).

Since some passengers have wrong cognition on the difference between soft seat and hard seat, the outcome of statistic may have a little difference from the real need

for soft seat. In order to reflect the real market and reduce the risk of the scheme, the result should be properly revised by means of decreasing the ratio of type Ⅰ while increasing that of type Ⅳ.

Tab.1 The types of passenger and its demand characteristics

type	Demand on speed	Demand on comfort	Sensitivity on change of price	Choice of seat	investigation		Suggest ratio
					statistic	ratio	
Ⅰ	Very high	Very high	low	Soft seat	204	0.277	0.20
Ⅱ	higher	higher	common	hard seat(1st); soft seat(2nd)	287	0.39	0.40
Ⅲ	high	common	high	hard seat	200	0.272	0.30
Ⅳ	low	low	Very high	Hard seat	45	0.061	0.10

2.2 Classification of IT

Based on different passengers and characters of IT, it is necessary to operate different kinds of train to meet subdivided market. The differences between trains are mainly in the priority, the ratio of stops, and the ratio of soft/hard seat.

Priority means the sequence of train when the operating scheme ensured. Higher grade train has higher priority. It could be conveyed as: express train f rapid train f ordinary train.

Ratio of stops γ_S means total stops K of one train compared to total stations S on the journey (see the formula 1). Higher priority train has fewer stops.

$$\gamma_S = K/S \tag{1}$$

So the stops K_i of train type i is (integral number): $K_i = \lfloor S \cdot \gamma_S \rfloor$

Ratio γ means the number of soft coach m_S compared to the number of hard coach m_H of a train. This ratio could be calculated by comparing the passenger numbers of choosing soft or hard seats in the market investigation.

$$r = m_S/m_H = \frac{a_H N_S}{a_S N_H} \tag{2}$$

In the formula, a_S and a_H respectively mean a fixed number of seats in one soft or hard coach. Supposing the cars of IT formation is 10, $a_S = 88$, $a_H = 128$, the formation structure and the fixed seat number in each type of train could be obtained as showing in table 2.

Tab.2 the ratio of stops, formation structure and fixed seat number in each type of IT

Types of IT	priority	Ratio of stops	Ratio of soft/hard coach	Fixed seat number of IT
express train	1	1/5	4∶6	1040
rapid train	2	2/5～3/5	6∶4	1120
ordinary train	3	1	1∶9	1240

2.3 The ratio of carrying traffic by different trains

According to the passenger flow forecast, the volume of traffic between central cities occupies 24% of the total passenger flow, and the volume of traffic between main towns is about 40%, the others 36%. To emphasize the express train because of its importance in guiding tendency, and take ordinary train as fundamental service, the ratio of carrying traffic attracted by each type of train can be adjusted as showing in Table3.

Tab. 3 the ratio of carrying traffic by each type of train

Types of IT	Service range	The fixed seat number (Ai)	Proportion	Ratio（R_i）
express train	Central cities	1040	18％	20%
rapid train	Main cities and towns	1120	36％	30%
ordinary train	All stations	1280	46％	50%

3. The method of formulating the operating scheme of IT

3.1The analysis of existed methods

Method 1: Based on the ratio of carrying traffic by each type of train and on the total quantity of passenger flow, calculate the train number of each type in every section of the line with a special table, then determine the operating plan of all trains on the IRT. This method is simple and easy to carry out. But it does not give a further analysis on specific stop plans of each type of train.

Method 2: Suppose that all passenger flow are organized to run through the train in order to maximize the travel speed. Then according to the order of the O-D flow from long to short and the stop times from less to more, the through trains of all O-D flows are connected if the beginning of one flow can be linked together with the end of another flow, in order to reduce the need number of train-set. The method will get the highest travel speed, but it ignores the diversity of passengers' demand, and the comparison of train capacity with real traffic need is just 1:1, which results in the loss of flexibility for passengers to choose different trains.

3.2 Hierarchy method

Because of the defect of these existing methods, it is necessary to develop a new method to meet travelers' demand, the requirement of operating management as well as the practical condition of IRT. Hierarchy method considers the operating section, types of train, the number of train as well as stop plan, etc.

Based on the train running range, hierarchy method determines the volume of passenger flow in relevant range, and makes sure the numbers and station-stop plan for each type of train according to their characters, then works out operating scheme in each period in accordance with passengers' demand and their distributing rule.

3.2.1 The numbers of IT

(1) The train running range

In order to make full use of capacity and avoid the unnecessary waste, and the different volumes in different sections considered, it is necessary to provide shuttle service in these sections with larger volume besides through service of the line.

• Make the minimum section volume $N_{\min}$ as the initial value;

• Judge the different volume in each section with the ratio β_i ($\beta_i = N_i / N_{\min}$), which is defined as the volume in any section N_i compared with the initial value $N_{\min}$. If the ratio is less then 5%, it could be regarded as no difference. While the ratio is beyond 10%, it will be regarded as an apparent difference.

• While β_i>1.1, it is the necessary condition to consider operating a new shuttle service in the section, but not sufficient condition, some other conditions, such as the department of maintenance and repair, warehouse of vehicles ,etc. must be considered.

•If there is a shuttle service, the residual flow should be treated by using the same method above in cycles.

(2) The train numbers

① the volume of passenger flow N

$$N = \max\{N_i\} \tag{3}$$

In the formula, N_i means the volume of the section without shuttle service.

② the train numbers according to the volume of passenger flow and the ratio by each type of trains.

$$n_i = N \cdot R_i \Big/ A_i \cdot \lambda_i \tag{4}$$

In the formula, λ_i means the rate of the train capacity being used, value it as 1 here.

③ the processing of decimal. Sum up the entire decimal as integer, and then distribute it to each type of train by their priority.

④ the volume of shuttle service. Make sure the volume of shuttle service in the same way after treating the residual volume.

3.2.2 the station-stop plan

According to the facts of different passengers served by different type of trains and the different departure/arrival volumes of passenger flow at each station, as well as to the principle that same type train would better have the same plan, comprehensive analysis should be given to the station-stop plans of different type train.

(1)Express train mainly serves the passengers between central cities along the line. Since there are only a few central cities, but the volume of traffic is mass, express train should stop at every central city station.

(2)Rapid train serves the passengers between major cities and towns. It stops at each of the central cities and main towns along the line.

(3)Ordinary train stops at every station, but some stations can be exception only if its volume is too small and reducing stop can increases the carrying capacity of the line.

4 The Operating Scheme of Hu-Ning IRT Line

4.1 The operating scheme

According to the volume of each section on Hu-Ning (Shanghai-Nanjing) IRT and the calculating method mentioned above, a case study is given as follows (see table 4 and table 5), Then make sure the character of each station on Hu-Ning IRT according to their station-stop ratios, departure/arrival volume at peak hours, etc..

Then, the operating scheme is presented in the Fig.1.

4.2 The analysis of operating rule

(1) Train distributing in each hour

With the result from market investigation, make sure that the ratio r_i of passenger flow generated in each hour and the train numbers D_i in each hour.

$$D_i = n_i \times r_i \tag{5}$$

The decimal is treated in such a principle that "increases the express in peak hours, the rapid train in hypo-peak hours, and the ordinary train in non-peak hours. Then the operating scheme puts up different rules in each period.

Tab. 4 Volume calculating for each section in Hu-Ning IRT (Unit:10,000 p/y)

Section	Nanjing -Zhenjiang	Zhenjiang -Changzhou	Changzhou -Wuxi	Suzhou -Wuxi	Suzhou -Shanghai
$N_{\min}$	2865	2775	3147	3121	3323
β_i	1.03	1	1.13	1.12	1.20
$N_{\min}$	2865				
Remained volume	0		282	256	458
β_i'	—		1.10	1	1.78
$N'_{\min}$	—		282		
$N''_{\min}$	—	—	0		176

Tab. 5 The sum of trains in each running range (Unit: pairs/day)

Running range	Express train	Rapid train	Ordinary train
Shanghai -Nanjing	19	25	38
Shanghai-Wuxi	2	3	3
Shanghai -Suzhou	1	1	2

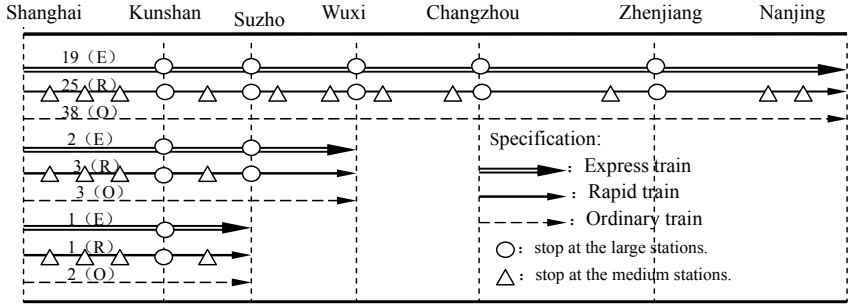


Fig.1 The operating scheme of IT on Hu-Ning IRT

(2) Departure time for each type of train

According to the sum of trains in each hour-- D_i , the running rule of each type of train can be determined. For instance, the express departs at whole or half hour, the rapid train at whole 10 minutes, but for the ordinary train, some factors should be considered, such as utilization ratio of the line capacity and the equilibrium of the trains' distribution.

From the Fig.2, the train distribution of each type shows different orderliness in peak hours and non-peak hours. Adopting the hierarchy method to make sure the operating scheme of IRT, through splitting passengers into different groups, dividing train into different types, then taking Hu-Ning IRT as an example, is verified as a feasible and practical way.

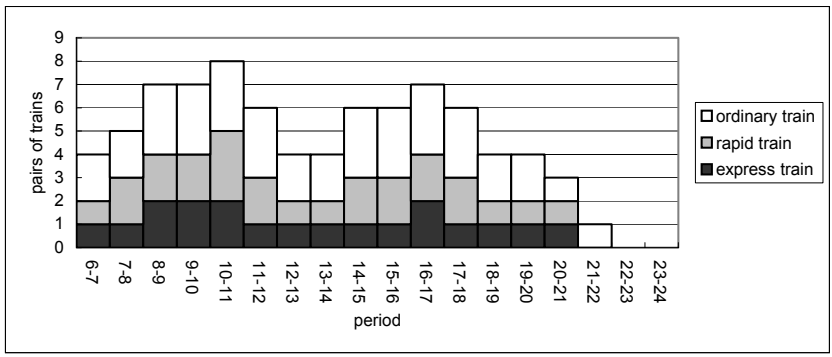


Fig.2 Distributing of each type trains in each hour on Hu-Ning IRT

References

The project team. *The Report for the Feasibility of Hu-Ning Intercity Transit[R]*. Wuhan: The Fourth Survey and Design Institute of Ministry of Railways, 2002.

XIANG Jing-song. The Research and Development on the Operating Scheme System of Intercity Trains, Master's thesis of Tongji University, 2006.3

XU Rui-hua, ZOU Xiao-lei. *The optimization research of the running plan for passenger transportation special line [J]*. Journal of Tongji University, 2005.12.

XU Xing-fang. *Regional Intercity Train Operating Organization: Theory and Practice.* Doctoral dissertation of Tongji University, 2005.2.

XU Xing-fang, XIANG Jing-song, etc. *The market analysis of intercity railway passenger transportation in Hu-Ning-Hang area [J]*. Railway Transportation and Economy, 2006.1.

SURVEY AND ANALYSIS OF DRIVER'S RESPONSE TO DYNAMIC TRAFFIC INFORMATION

Guiyan JIANG[1], Zuduo ZHENG[2], Zhu BAI[3], Leilei DAI[4], Jiaqi ZHAO[5]

[1] Doctor, Prof., College of Transportation Jilin University, 5988 Renmin Street, Changchun City, Jilin Province, P.R.China 130025, PH: 0431-85095505, FAX:0431-85095595, email: jianggy@public.cc.jl.cn

[2] Postgraduate, College of Transportation Jilin University, 5988 Renmin Street, Changchun City, Jilin Province, P.R.China 130025, email: mada7982@yahoo.com.cn

[3] Postgraduate, College of Transportation Jilin University, 5988 Renmin Street, Changchun City, Jilin Province, P.R.China 130025, email: baizhu_2003@sohu.com

[4] Postgraduate, College of Transportation Jilin University, 5988 Renmin Street, Changchun City, Jilin Province, P.R.China 130025, email: alei_9935@126.com

[5] Postgraduate, College of Transportation Jilin University, 5988 Renmin Street, Changchun City, Jilin Province, P.R.China 130025, email: power_zjq@163.com

Abstract

Even the most sophisticated Traffic Information Systems (TIS) might not be successful if driver's response to traffic information cannot be estimated in advance. Overreaction and concentration are often triggered if the estimations of driver's response to the information are not correct. Unfortunately, the analysis of drivers' response to traffic information is also very complex, because the responses of different drivers can be totally different to the same information. This paper presents the results of a study on driver's response to and evaluation of dynamic traffic information from radio and Variable Message Signs (VMS) using qualitative and statistic methods. The analysis of the survey conducted in Changchun city shows that the usefulness of traffic information from radio is admitted by most drivers and that this kind of information hasa significant influence on the drivers' behavior. The analysis also shows that VMS are not dangerous to drivers if designed carefully. The analysis also revealed that congestion information and alternative route recommendations are most welcomed in general regardless of the forms of media used.

Key words: *traffic engineering, information systems, driver's behavior, variable message signs (VMS), traffic information radio (TIR)*

INTRODUCTION

The role of TIS to drivers is not forcible compared with that of the traffic signal control systems. It depends on drivers' response to traffic information whether or not the objectives of TIS could be gained. Even the most sophisticated TIS might not be successful if driver's response to traffic information cannot be estimated rightly in advance. So sufficient understanding of the impact of traveler information on a drivers' behavior is extremely important for TIS. The overreaction and concentration (therefore leading to congestion shifting) are often triggered if the estimations of a driver's response are not correct (Arnott et al. 1991). Unfortunately, the analysis of drivers' response to traffic information is also very complex, because the responses of different drivers can be totally different to the same information. At least it involves the modeling of drivers' behavior (and therefore the psychology of decision-making), the interaction of drivers' behavior with new information technologies, and the interaction among the drivers themselves (Emmerink et al. 1996). Researchers (De Palma and Lefebvre 1983; Emmerink et al. 1995a; Mahmassani and Jayakrishnan 1991) have addressed the difficulties of researching drivers' response to traffic information.

In most countries, TIR and VMS are the most popular channels for drivers to access traffic information currently. In this paper, the responses of drivers to these two types of media is emphatically investigated. Based on the review of the literature in this domain, this paper introduces a survey study conducted in Changchun, China. Driver's response to traffic information from TIR and VMS is analyzed using the qualitative and statistic methods.

2. PROPOSED STUDY

2.1 Review of Study Methods

There are two kinds of methods often adopted by researchers around the world to address traffic information's behavioral impacts on drivers: qualitative techniques and simulation techniques. The former includes focus groups, interviews, and participant-observer methods to help fill the gaps left by quantitative techniques. Qualitative methods, long used in other disciplines, have caught the attention of a growing number of travel behavior researchers. The reason is that "qualitative research is vital to understand the complexity of transportation behavior, which rests upon the subjective beliefs and behaviors of the individual person" (Poulenez-Donovan et al. 1994).

The Stated Preference (SP) method, one of the most popular qualitative techniques, is a powerful tool for helping us understand the complexities of drivers' behavior to dynamic traffic information. So it has been extensively applied to analyze driver's behavior (Bradley et al. 1986; Hensher et al. 1990; Brocken and Van der Vlist

1991; Wardman et al. 1997; Ortuzar et al. 1994; Abdel-Aty et al. 1997).

The attractions of the SP approach are largely due to its ability to control the choice context and the independent variables that will enter the demand model (Kroes and Sheldon 1988). These features are particularly important to the analysis of the impacts of dynamic traffic information from TIR and VMS on drivers' behavior.

At the same time, its indisputable shortcomings are the lack of scientific rigor and the threat of subjective interpretation (Clifton and Handy 2001).

The simulation techniques are used relatively less often, mainly due to the costliness and complexity of building a realistic replica of a road network (Chatterjee et al. 2002).

On the consideration mentioned above, one of SP methods is adopted in the combination of statistic methods in our study.

2.2 Objective and Data Collection

The objective of this study is to investigate drivers' demands for traffic information and their evaluation on the performance of current available methods offering traffic information to them. A questionnaire survey was conducted between Oct.5 and Oct.15, 2004 in Changchun, China. The questionnaire included four parts: overall review and information demands about traffic information, evaluation on VMS traffic information, evaluation of radio traffic information and respondents' socio-economic characteristics.

A small-scaled survey was organized as a preliminary one in order to ensure the formal survey's quality so the questionnaire could be revised according to the feedback. More importantly, some trivial but useful skills were obtained for the later formal survey. It was proved that such preliminary survey was quite necessary before the formal one.

Finally the formal survey was conducted. To avoid the low return rate encountered by some researchers, the questionnaires were finished by face to face contact even though it would take much time. The survey sites were mainly the places where many vehicles often gathered together and the sites covered most locations where lots of vehicles parked in Changchun.

The respondents consisted of four categories: the taxi drivers, the private car drivers, the truck drivers and the car drivers for official business. Their proportions are almost consistent with the proportions of the four class drivers in Changchun.

2.3 Analysis

2.3.1 General Analysis

98.3% respondents deemed that it's necessary to provide traffic information, 70.4% deemed that the traffic information they could receive currently was less than they need. But to our surprise, 30% respondents thought that the traffic information provided currently was neither more nor less than what they need. The main reasons

attributed to the results may be: (1) the attitude of the respondents to the survey was not serious-minded, which is unavoidable in most questionnaire surveys; (2) the main reason might be that many respondents thought the current traffic information they receive could exactly satisfy their need, because the congestion in peak hour is not serious currently in Changchun city and the demand for traffic information is not as urgent as the counterparts in other big cities; (3) the drivers have not realized the importance of traffic information because only few media are used in Changchun for them to obtain traffic information.

The survey showed that the traffic information that drivers most wanted was: road conditions (67.8%), congestion information (67.0%), weather information (53.9%) and traffic flow status ahead (50.4%), listed in descending order according to the percentage. Unexpectedly, only 28.7% respondents chose incident information. The possible reasons may be: (1) some respondents might misunderstand the question because congestion information and incident information appeared in the questionnaire simultaneously; (2) comparing with the occurrence of incident, drivers might pay more attention to the congestion of the traffic ahead.

Under incident condition, the information drivers most wanted was: the severity of incident and congestion (53.0%), the alternative routes (53.0%), incident location (52.2%). However, only 16.5% respondents were interested in the duration information of an incident and only 3.5% in the information of speed limit variation. This is mainly because the respondents were all the drivers using the urban road network, and the importance of the variation of speed limit due to an incident is much less than that in freeways.

In the survey, 80% respondents admitted that they knew VMS. 86% respondents thought that VMS was a bit useful, useful or very useful. 45.5% and 20% respondents thought that VMS did not or might not bring drivers danger respectively. This is consistent with the conclusions mentioned in the literature. So it can be asserted that VMS provides a relatively cheap alternative for the very sophisticated and expensive on-board navigation systems and that VMS cannot be dangerous to drivers if designed carefully.

About the style of traffic information displayed on VMS, 45.9% respondents selected “displayed in combination of character and map”, 36.9% selected “displayed using character only”, only 15.3% selected “displayed using map only” and 2 respondents skipped it. . Note that this conclusion may have limited value for reference because VMS is still one of the burgeoning media in Changchun.

The survey showed that the traffic information that drivers expect VMS to provide include: the traffic condition of main streets (71.2%), location of incident (37%), the violation record information (36%), the traffic condition of intersections (29.7%) and weather information (29.7%). Surprisingly, only 10.8% and 9%

respondents selected the expected delay and expected travel time respectively, which is a bit different from other similar research (Ran et al. 2002; Wardman et al. 1997; Chatterjee et al. 2002). The reason also might be that VMS is still a new medium in Changchun and most of the respondents have little experience in using it.

Most of the respondents thought that the VMS installed in Changchun was valuable, which is consistent with the conclusion mentioned above.

Among the respondents, 67.5% often listen to the Traffic Information Radio (TIR) and 27.2% listen sometimes and only 1.8% never listen.

Seen from the survey, TIR was highly praised by most of the respondents because it's the most primary and popular method in Changchun city providing traffic information currently. 38% respondents deemed that TIR was very useful, and correspondingly, 38% had changed their route choices frequently according to the radio traffic information. 56% respondents thought that TIR was useful, but only 38% had changed their route choices sometimes according to TIR. This is not ambivalent because the radio broadcast includes not only traffic information but also other information such as news, music, entertainment, advertisements and so on. Therefore in the respondents' opinion, the usefulness of TIR does not necessarily imply that the information might make them change their driving behavior.

According to the survey, the main problems in TIR of Changchun city are: (1) sometimes the information provided is neither timely nor accurate; (2) sometimes the useful traffic information is difficult to be used effectively because there is too much redundant and commercial information.

2.3.2 Further analysis

Unlike some literatures where the logit models are employed to analyze the SP data, this paper adopts one of the statistic methods, correlation analysis, to complete the same job because it's difficult to validate the parameters of logit models, which is vital to models' performance.

In the proposed survey, the variables which might have influence on drivers' behavior include: vehicle type (represented by A), evaluation on current available information (represented by B), evaluation on VMS (represented by C), frequency of listening to TIR (represented by D), evaluation on TIR (represented by E), driving behavior changes due to radio traffic information (represented by F) and driver's education level (represented by G）.The former two variables have seldom been considered in the literature. According to the empirical, all or some of these variables might have an impact on drivers' behavior and the bivariate correlations may exist. To testify the hypothesis, correlation analysis is adopted. The results are presented in Table 1.

Derived from Table 1, Table 2 contains only those variables that correlate significantly.

Table 2 reveals that vehicle type has significant correlation with education level. This means that drivers of trucks and official business cars are educated less than those of private cars and taxis. It also reveals that frequency of listening to the radio has significant correlations with vehicle type, evaluation on VMS, evaluation on TIR, driving behavior changes due to radio traffic information and education level respectively, which indicates the following interesting conclusions:

(1) The more frequently the driver listens to TIR, the lower evaluation on VMS he makes. However, Emmerink et al (1998) had a reverse conclusion.

(2) The more frequently the driver listens to TIR, the higher evaluation on radio he makes.

(3) The more frequently the driver listens to TIR, the more route changes he makes due to the information. This agrees with the result in Caplice and Mahmassani(1992).

(4) The less the driver is educated, the more frequently he listens to TIR.

Seen from Table 2, the three variables of education level, vehicle type and frequency of listening to the radio are correlated to each other. It should be noted that Pearson's correlation coefficient is a measurement of linear association between two variables and that the ties among these three variables are ignored. So the Kendall's correlation coefficient is employed in the case to testify if the correlations between variables are generated by the third one. The results are depicted in Table 3.

Table 1. Results of Correlation Analysis between Variables

	item	A	B	C	D	E	F	G
A	Pearson Correlation	1	-.034	.014	-.201(*)	-.179	-.056	.286(**)
	Sig. (2-tailed)	0	.719	.879	.032	.057	.554	.002
	N	114	114	114	114	114	114	114
B	Pearson Correlation	-.034	1	.104	.098	.063	-.011	.065
	Sig. (2-tailed)	.719	0	.269	.298	.508	.910	.494
	N	114	114	114	114	114	114	114
C	Pearson Correlation	.014	.104	1	.189(*)	.116	-.015	-.098
	Sig. (2-tailed)	.879	.269	0	.044	.221	.877	.301
	N	114	114	114	114	114	114	114
D	Pearson Correlation	-.201(*)	.098	.189(*)	1	.447(**)	.192(*)	-.243(**)
	Sig. (2-tailed)	.032	.298	.044	0	.000	.041	.009
	N	114	114	114	114	114	114	114
E	Pearson Correlation	-.179	.063	.116	.447(**)	1	.143	-.173
	Sig. (2-tailed)	.057	.508	.221	.000	0	.129	.065
	N	114	114	114	114	114	114	114
F	Pearson Correlation	-.056	-.011	-.015	.192(*)	.143	1	-.108
	Sig. (2-tailed)	.554	.910	.877	.041	.129	0	.253
	N	114	114	114	114	114	114	114
G	Pearson Correlation	.286(**)	.065	-.098	-.243(**)	-.173	-.108	1
	Sig. (2-tailed)	.002	.494	.301	.009	.065	.253	0
	N	114	114	114	114	114	114	114

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Table 2. Results of Significant Correlation between Variables

item		A	C	D	E	F	G
A	Pearson Correlation	1	.014	-.201(*)	-.179	-.056	.286(**)
	Sig. (2-tailed)	0	.879	.032	.057	.554	.002
	N	114	114	114	114	114	114
D	Pearson Correlation	-.201(*)	.189(*)	1	.447(**)	.192(*)	-.243(**)
	Sig. (2-tailed)	.032	.044	0	.000	.041	.009
	N	114	114	114	114	114	114

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Table 3. Results of Kendall Correlation Analysis

item		A	D	G
A	Correlation Coefficient	1.000	-.175(*)	.204(**)
	Sig. (2-tailed)	0	.040	.010
	N	114	114	114
D	Correlation Coefficient	-.175(*)	1.000	-.248(**)
	Sig. (2-tailed)	.040	0	.003
	N	114	114	114
G	Correlation Coefficient	.204(**)	-.248(**)	1.000
	Sig. (2-tailed)	.010	.003	0
	N	114	114	114

* Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).

Then input the values of 0.204, -0.175 and -0.248 in Table 3 into the equation (1):

$$\tau_{AB,C} = \frac{\tau_{AB} - \tau_{CB} \times \tau_{AC}}{\sqrt{(1-\tau_{CB}^{\ 2})(1-\tau_{CA}^{\ 2})}} \tag{1}$$

When τ_{AB}=-0.175, τ_{CB}=0.204 and τ_{AC}=-0.248, $\tau_{AB,C}$ is -0.131, which is different from τ_{AB} considerably. This implies that the significant correlation of vehicle type and frequency of listening to TIR is derived from the influence of education level.

3. CONCLUSIONS

Based on the analysis of the survey data conducted in Changchun, China, this paper analyzed the driving behavioral impact of traffic information provided by two kinds of the most primary and popular media currently in China: TIR and VMS. The analysis shows that:

(1) Although not very strong, the correlations are significant between frequency of listening to TIR and vehicle type, evaluation on VMS, evaluations on TIR, driving behavior changes due to radio traffic information and education level. Its correlation with vehicle type is derived from the influence of education level. And interestingly, vehicle type and education level also have a significant correlation.

(2) In Changchun city, TIR is the most important method providing traffic information to drivers at present. It is accepted extensively by drivers and influences their driving behavior to some extent.

(3) Although there are not many VMS in Changchun currently, the survey implies that VMS has great potential to guide drivers' driving behavior, because most respondents showed great interest in this new technology and said yes to its value with no hesitation. More importantly, this survey confirms again that VMS won't put drivers in danger if it is designed carefully, which is consistent with the literature.

(4) The congestion information and recommended alternative routes are welcomed mostly in general regardless of what kinds of media are adopted. But the expected delay and travel time are not as important as reported in some literature.

The gender's role in drivers' behavior is not analyzed because most of drivers are male in Changchun. The age's impact is also ignored because most respondents refused to give such information. But gender and age do have significant a impact as mentioned in the literature. So these two factors should be focused on in further study.

More importantly, the sample size should be extended greatly. If there was enormous survey data, the correlations found in this survey should be testified further.

Finally, it should be pointed out that this study was conducted in Changchun, it is uncertain whether the findings are applicable to other cities.

ACKNOWLEDGEMENT

This study is funded by National Natural Science Foundation of China (NSFC) (50338030) and 863 projects of China (2006AA11Z228).

NOTATION

The following symbols are used in this paper.

τ_{AB} = the correlation coefficient between variables A and B

τ_{CB} = the correlation coefficient between variables C and B

τ_{AC} = the correlation coefficient between variables A and C

$\tau_{AB,C}$ = the measure of influence of C to the correlation between A and B

REFERENCES

Abdel-Aty, M. A., Kitamura, R., and Jovanis, P. P. (1997). *Using stated preference data for studying the effect of advanced traffic information on drivers' route choice*. Transportation Research, 5C, 39-50.

Arnott, R., et al. (1991). *Does providing information to drivers reduces traffic congestion.* Transportation Research, 25A:309-318.

Bradley, M. A., Marks, P., and Wardman, M. (1986). *A summary of four studies into the value of travel time savings*. Proceedings of Seminar M, PTRC Summer Annual Meeting, Brighton.

Brocken, M. G. M., and Van der Vlist (1991). *Traffic control with variable message signs*. Proceedings of VNIS, Part 1, IEE.

Caplice, C., and Mahmassani, H. S. (1992). *Aspects of commuting behavior: Preferred arrival time, use of information and switching propensity*, Transportation Research, 26 A,409-418.

Chatterjee, K., Hounsell, N.B., Firmin, P. E., and Bonsall, P.W. (2002). *Driver response to variable message sign information in London*, Transportation Research Part C 10 149-169.

Clifton, K.J., Handy, S.L. (2001). *qualitative methods in travel behavior research*，the Conference on Transport Survey Quality and Innovation.

De Palma, A., and Lefevre, C. (1983). *Individual decision-making in dynamic collective systems*. J. Moth1 Social. 9, 103-124.

Emmerink, R. H. M. (1998). *Information and pricing in road transportation.* Springer-Verlag Berlin .Heidelberg .

Emmerink, R. H. M., Axhausen, K. W., Nijkamp, P., and Rietveld, P. (1995a). *Effects of information in road transport networks with recurrent congestion.* Transportation 22, 21-53.

Emmerink, R. H. M., Nijkamp, P., et al. (1996). *Variable message signs and radio traffic information: An integrated empirical analysis of drivers' route choice behavior*. Transpn. Res.-A. Vol.30, No. 2. pp. 135-153.

Hensher, D. A., Milthorpe, F. W., Smith, N. C., and Barnard, P. (1990). *Urban tolled roads and the value of travel time savings.* Economic Record 66, 146-156.

Kroes, E. P., and Sheldon, R. J. (1988). *Stated preference methods: an introduction.* Journal of Transport Economics and Policy, 22, 11-25.

Mahmassani, H.S., and Jayakrishnan, R. (1991). *System performance and user response under real-time information in a congested traffic corridor.* Transpn Res. 25A, 293-308.

Ortuzar, J. de D., Iacobelli, A., Ivelic, A. M., and Munizaga, M. (1994). *Evaluating private highway projects under concession: route choice modelling with mixed RP/SP data.* Proceedings of Seminar G, PTRC Summer Annual Meeting. PTRC, London.

Poulenez-Donovan, Craig Jesus and Cy Ulberg (1994). *Seeing the Trees and Missing the Forest:Qualitative Versus Quantitative Research Findings in a Model Transportation DemandManagement Program Evaluation.* Transportation Research Record, No. 1459, pp.1-6.

Ran, B. et al. (2002). *Evaluation of Variable Message Signs in Wisconsin: Driver Survey.* University of Wisconsin at Madison.

Wardman, M., Bonsall, P.W., and Shires, J.D. (1997). *Driver response to variable message signs*: *A stated preference investigation*, Transpn Res.-C, Vol.5, No., pp.389-405.

Development of an Application Programming Interface (API) in Simulation to Evaluate Real-Time Signal Control

Fang Clara Fang

Fang Clara Fang, Ph.D., Assistant Professor, Department of Civil Engineering, College of Engineering, Technology, and Architecture, University of Hartford, West Hartford, CT, 06117, phone: (860) 768-4845, fax: (860) 768-5037, email: fang@hartford.edu.

Abstract： Application Programming Interface (API) is a feature available in some traffic simulation programs to allow users to customize changes in simulation such as driver behavior, ramp metering control strategies, and other ITS applications. This paper presents a method to develop an API in AIMSUN, a stochastic and microscopic simulation model, to evaluate a previously developed adaptive signal control algorithm, i.e., DP algorithm. The algorithm is coded in C++ by combing the functions and five routines available from the GETRAM extension module embedded in AIMSUN. The developed API allows dynamic link between the DP algorithm and the simulation environment. During simulation run time, the real-time traffic measurements provided by detectors in the network feed the DP algorithm that, after processing, makes the decision of signal controls. These decisions are then transferred back to the simulated network, which emulates their operations through the signal controllers. This research has demonstrated the API function is a useful tool to assess external ITS applications when these innovative ides are not supported by the simulation program.

Key words: *Application Programming Interface (API), Traffic Simulation, Real-time Signal Control*

1 Introduction

Simulation has been widely used for the evaluation of various new improvement proposals in our transportation system. Particularly, the simulation has become an increasing valuable tool for assessing new ITS based traffic control strategies. However, these innovative ideas usually are difficult to be simulated because of being lack of this feature in many common simulation models. An Application Programming Interface (API) is then available as an advanced feature in some simulation models such as Paramics and AIMSUN to provide an interface connection between user-defined applications with the simulation environment.

The authors have previously developed a method namely dynamic programming (DP) algorithm (Fang and Elefteriadou, 2005) to provide optimal and real-time signal

control of diamond interchanges. This innovative signal control generates and implements the signal plan dynamically based upon real time traffic conditions, which are measured through a traffic detection system. It differs widely from common fixed time and actuated signal control and cannot be supported by the existing simulation programs. This paper will discuss how an API is developed using the GETRAM extension module in AIMSUN to dynamically link the algorithm to the simulation in real time. The developed API will allow the DP algorithm to access internal data such as detected traffic demand. Based on it, the algorithm optimizes signal control and communicates back to the simulation which consequently controls the signal operations in the running time.

This paper is organized as follows. Section II provides a review of the applications of API in simulating external applications. Section III presents a procedure for conducting the simulation evaluation study of the DP algorithm including the selection and calibration of the model. The method for developing an API to simulate the DP algorithm in AIMSUN is discussed in Section IV. In this section, the function of GETRAM extension module is introduced, DP implementation schemes including detectors capabilities and vehicle projection is described next and the flow chart of how the DP is coded is discussed finally. This is followed by some results analysis and conclusions.

2 Literature Review

When a traffic simulation package is used to evaluate the DP optimal signal control algorithm, an API interface connecting from the simulation to the algorithm is necessary because available microscopic simulation programs do not support this customized control. The connection has to allow the DP algorithm to access traffic simulation internal data while optimize signal timing in real time on user-defined simulated network provided by the simulation package.

An API's function should have two folds, one is to override the simulator default models, such as car following, lane changing, the other is to allow them to interface external complementary modules to the simulator. Complementary modules could be any ITS application, such as advanced ramp metering, incident management, etc. Nearly every published research on the application of API is within Paramics (Paramics, 2001), a microscopic traffic simulation tool, to test some traffic control strategies. The California PATH project team (Liu et al, 2001, Liu and Recker, 2002, and Gardes, 2003) perhaps was the first group to use the API function in Paramics. They developed plug-in API modules to modify elements of driver behaviors for modeling high occupancy vehicle lane operations on I-680 and to test actuated signal control, time-based ramp meter control and path-based routing. The research has also been disseminated during some Intelligent Transportation Systems ITS meetings (Liu and Recker, 2002, and Kim, Gardes and May, 2002). Followed by PAHT's work, a few other researchers have also applied API in Paramics for other purposes. They are described as follows,

Bartin, Ozbay and Yanmaz (2005) implemented a gap acceptance and rejection binary probit model in Paramics API based on field data for simulation and modeling the Collingwood and Brooklawn circles in New Jersey. The differences between the simulation results of the API enhanced model and the default Paramics model were presented. Park, Yun and Choi (2004) developed an API for coordinated actuated signal control logic for the Paramics microscopic simulation program. The developed API for the coordinated actuated signal control allows phase skip, gap-out and force-off, which were generally not well supported by the existing functions in the Paramics. Recently, Smith, Tudge and Wilson (2005) have written API module for simulating toll payment types, toll plaza driver behavior and tidal lane management at Sydney Harbor.

Although API has been demonstrated to be feasible for enhancing Paramics capability, very few research have explored the use of API in other microscopic models and no comparison study has been done among APIs in different simulation programs.

3 Using Simulation for Evaluation

3.1 Overview

An optimization methodology and its algorithm have been developed by authors [1] to provide optimal signal control of diamond interchanges in response to real-time traffic fluctuations. The problem is formulated as to find a phase sequencing decision with a phase duration that makes a pre-specified performance measure minimized over a finite horizon that rolls forward. The problem is solved by a forward DP method. The signal timing is optimized for each 10 seconds. The optimal signal switches over each 2.5-second interval are found for each horizon of 10 seconds. The optimization process is based on the advanced vehicle information obtained from loop detectors set back a certain distance from the stop-line. Vehicle trajectories from detections till future arrivals and departures are modeled at the microscopic level to estimate the traffic flows at the stop-line for each horizon. Fang's dissertation (2004) provides more detailed description including methodology formulation, optimization solution and vehicle prediction models, etc.

The following presents a procedure for conducting the simulation evaluation study of the DP algorithm. First, a microscopic simulation model that can simulate networks and adaptive signal control is selected as an evaluation model. Secondly, the selected model is calibrated using diamond interchange field data. Thirdly, and most importantly, an API is developed to provide real-time connection between the DP algorithm and calibrated simulation environment. The DP algorithm is coded in computer language based on simulation program functional routines. Once the real-time signal control is fully integrated, the calibrated diamond interchange is then used to conduct sensitivity analysis of some parameters in the algorithm and compare

of the performance among the real-time signal control and other signal optimization tools.

3.2 Selection of a Simulation Model

To select a simulation model for evaluation purposes, it is necessary to exam its characteristics, functions and flexibilities of each candidate model, with special respect to the potential being used to simulate the proposed adaptive signal control. Several common microsimualtion models including Paramics (2001), CORSIM (1997), AIMSUN (2002) or VISSIM (2002) have been considered after the capabilities of each model are examined. Finally AIMSUN has been selected as the simulation model in this study because a) it provides a API function, called GETRAM Extension module, to implement advanced traffic control applications; and b) it can assign a controller to a set of movements in two or move nodes/intersections, which is essential to simulate signal controls at a diamond interchange. The API module enables users to develop external applications that may need access to internal data of AIMSUN during simulation run time. The user can program them either in C/C++, as a DLL library, or using Python scripting language. The extension module can be used to develop an API interface to other external control applications.

Some characteristics of AIMSUN and its application are briefly described as follows: AIMSUN is a microscopic traffic simulator that evaluates a wide range of functionally classified roadways, such as freeways, urban streets, ring roads and any combination of thereof. The model, based on Gipps car-following logic, can simulate various types of traffic controls (e.g., signs, signals, VMSs, metering), measurable traffic detector (e.g., counts, occupancy, presence, speed and density), vehicle and driver behaviors, public transportation and incidents.

3.3 Calibration of the Selected Simulation Model

The selected simulation model has been calibrated using field data to ensure that it can accurately replicate field conditions. The traffic simulation model usually provides default values of some parameters that represent average conditions. These parameters describe vehicles, drivers and traffic model (i.e., car-following, lane-change) characteristics. Rather than accepting these default values, the calibration process adjusts the model by quantifying these default values with site-specific data to the extent practical. The calibration of the diamond interchange for evaluating the DP algorithm and conducting comparison study in this study is based on the field data of a diamond interchange connecting I-17 with Indian School road in Arizona (Elefteriadou, et al, 2005). The data available include a video-image of the vehicle arrivals; geometric layout as shown in Figure 1; actual signal timing plan; and traffic demand. The field measure of control delay for each movement is also provided.

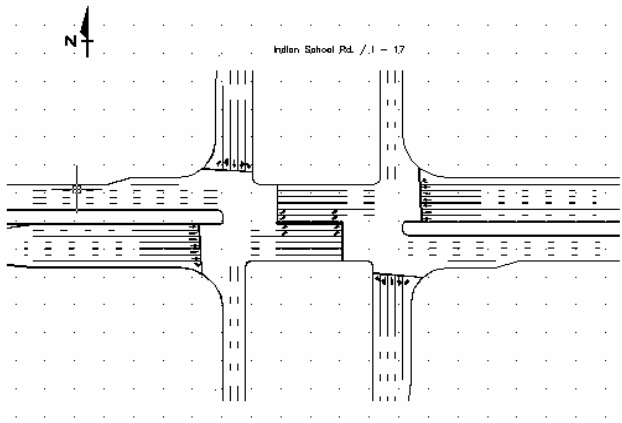


Figure 1. Geometric Layout for the Diamond Interchange (Indian School/I-17, Arizona)

Several parameters have been modified from their default values for calibration purpose, including driver's reaction time; response time at stop; vehicle maximum desired speeds; vehicle maximum acceleration; vehicle speed limit acceptance parameter; speed limit of links (sections); and speed limit of turning. Among them, driver's reaction time is the most sensitive parameter. This is the time a driver takes to react to speed changes in the preceding vehicle. It is used in the car-following model. This driver's reaction time is also taken as the simulation step. This study uses a 0.5s of the simulation step to increase the accuracy of obtaining traffic information for simulating the DP algorithm. In addition, vehicle information can be updated in each 0.5s for a DP interval of 2.5s rather than fractional data are retrieved.

4. Development of an API to Simulate the DP Algorithm in AIMSUN

4.1 GETRAM Extension Module

The simulation of the DP algorithm is carried out by the GETRAM extension module embedded in AIMSUN. The module serves as an interface to link AIMSUN to an external application (i.e., DP algorithm) that requires to access to some internal data of AIMSUN such as vehicle information during simulation run time. Figure 2 shows the process of the information exchange between AIMSUN and the DP algorithm via GETRAM extension module. First, the simulated interchange network equipped with detectors is modeled in AIMSUN. Next, during simulation, the real-time traffic measurements provided by those detectors feed the DP algorithm that, after processing, makes the decision of signal controls, e.g., implement phase for arterials

for next 7.5s. Finally, these decisions are transferred back to the simulated network, which emulates their operations through the signal controllers.

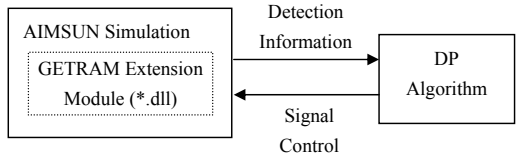


Figure 2 Simulation of the DP Algorithm in AIMSUN via GETRAM Extension Module

In order to achieve this exchange between the simulator and the DP algorithm, and have the network dynamically implementing the DP signal plan, the external algorithm is required to be compiled into a DLL (dynamic link library) file. In this research, Microsoft Visual C++6.0 is used to code the algorithm by combining the functions available from GETRAM extension module. The module has a number of functions available for users to obtain detection information from AIMSUN model, such as the number of vehicles that have crossed a detector during a simulation step or an aggregated time interval, vehicle speed, and detector occupancy, etc.

4.2 DP Algorithm Simulation Scheme

To implement the DP algorithm in AIMSUN, detectors are placed 656 ft upstream of the stop-line for each approach. How to determine the location of detectors has been discussed in authors' paper [1]. These detectors can generate four types of vehicle data at every simulation step. They are:

Count: number of vehicles; Speed: mean speed for vehicles crossing the detector during the step; Occupancy: percentage of time step the detector is pressed; and Presence: whether a vehicle is over the detector or not. Among them, count and speed are the required data gathered for implementing the DP algorithm.

A simulation step of 0.5s is used to collect vehicle detection information because smaller values means more accurate detected data, and a DP interval 2.5s is 5 (an integer) times 0.5s. As shown in Figure 3, detection data are collected in each simulation step of 0.5s for extreme detection range. The simulation time horizon is arranged so that the simulation starts at time 0 and the interchange is controlled in a fixed time signal control, when the time goes to 300 seconds (5 minutes), the signal switches to adaptive signal control. Figure 3 also shows the simulation scheme during this transition period. The first DP horizon is located at time 300 to 310s. Its corresponding detection period is 16 to 2 seconds in advance, i.e., detection time is 284 to 298 seconds. The arrival projection and DP calculation starts to be processed immediately after the detection period, but has to be completed before the time 300. Initial queue length and signal status for the beginning of DP horizon is obtained from

the simulation time at 298s. To increase the measurement precision, two consecutive DP optimization horizons are designed to have one interval of 2.5s overlap. Consequently, there is also an overlap between detection periods. Figure 4 describes the detection overlap for two consecutive optimization periods in terms of simulation time. The detailed relationship between optimization horizon and detection range is addressed in reference Fang [11].

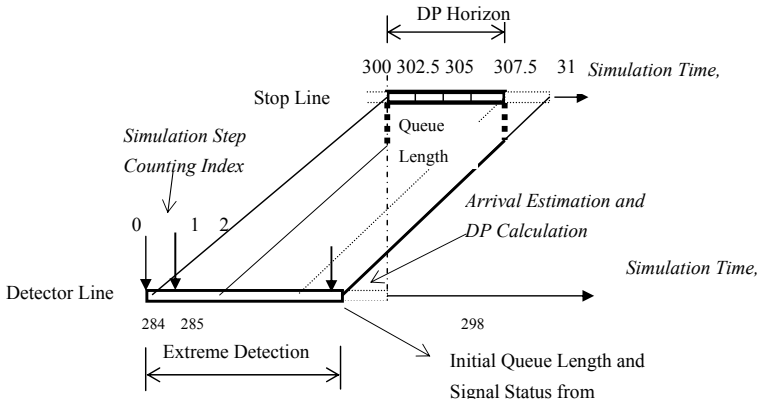


Figure 3 Arrival Detection and DP Computing Time Scheme

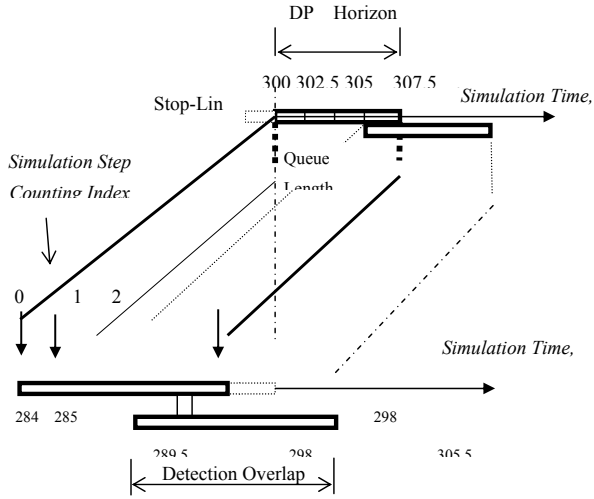


Figure 4 Overlaps for Arrival Detection and DP Horizon

4.3 Development of API for the DP Algorithm

The DP adaptive signal control algorithm is programmed according to five routines defined by the GETRAM extension module to build the API which allows communication between AIMSUN and the DP algorithm in real time. As illustrated in Figure 5, they are GetExtLoad, GetExtInit, GetExtManage, GetExtFinish and GetExtUnLoad. Their functions are described as follows,

- GetExtLoad:

It is called when the GETRAM extension is loaded by AIMSUN

- GetExtInit

It is called when AIMSUN starts the simulation and initialize some parameters, such as the simulation/detection step, DP rolling horizon and DP interval counts, as well as find displays the current signal plan (fixed time control)

- GetExtManage

It is called in every simulation stop at the beginning of the cycle. It is used to request and update detector measures, vehicle information and interaction with junctions. The DP optimization, vehicle arrivals-discharge projection model and signal implementation are embedded in this routine.

- GetExtFinish

It is called when AIMSUN finishes the simulation

- GetExtUnLoad

It is called when the GETRAM extension is unloaded.

Among five routines, GetExtManage is the key one that primarily contains the code for implementing the algorithm. The flow structure of the algorithm in GetExtManage consists of four blocks, as presented in Figure 5. The following illustrates each block function:

Block One – Vehicle Detection and Initial Status

During this block, detectors start to gather information for every simulation step 0.5 second for each approach, initial queue for each DP horizon is calculated, and the signal control at time of 2 seconds in advance is used for the initial signal plan for DP calculation. If it is the first DP cycle, vehicles arriving from time 284 – 298s at detectors are saved, otherwise, the detection overlap is considered.

Block Two – Forward DP Value Calculation

During this block, the DP forward recurrence method is applied. Based on the optimal state values of the previous stage, for all nodes in current stage, the value of the Performance Measure Index (PMI), optimization objective in the DP algorithm, for each node is calculated, and the optimal decision made at each node is obtained in current stage which results in the total PMI value minimal from the DP beginning to the current node. These optimal state values also serve as the initial PMI value of the next stage. Block Two also includes a subroutine developed for the vehicle projection

and DP value calculation over a single DP interval (2.5s). The subroutine, firstly, calculates the vehicle travel time T required to reach the back of queue at the stop-line for each approach in terms of three scenarios: a) no queue and signal green; b) no queue and signal red and c) initial queues present; Secondly, it locates each vehicle detection starting time point (T0) at the detector line; Thirdly, it computes the number of vehicles projected without taking into account the projection interactive dynamics. This is defined as nominal "qprojected" vehicle projection (it projects 5 simulation steps of the detected vehicles into the corresponding DP interval 2.5s starting from T0 for each approach); This step is followed by calculating the number of vehicles discharged (qdischarged). Next, qprojected is adjusted from nominal "qprojected" by considering projection interactive dynamics as follows,

- If qprojected > qdischarged, the queue becomes longer, i.e., the detection line projecting point moves to the right
- If qprojected < qdischarged, the queue becomes shorter, i.e., the detection line projecting point moves to the left

Every added or reduced queued vehicle increases or decreases 0.5s for detection time at the detector, i.e., the projection point moves to the right or the left. Finally, the subroutines calculates different PMI including the sum of final queues, qfinal, for all approaches, the sum of average delay per lane of all approaches, the sum of total delays and the sum of storage ratios.

Block Three – Optimal DP Signal Declaration

Block three finds the optimal phase in the last interval that will be the initial phase for a new round DP horizon and declare the optimal signals to be implemented.

Block Four – Optimal DP Signal Implementation

During this block, at time 300s, fixed time control is disabled and the DP control is implemented with the time rolling forward. Step by step, simulation and calculation of the DP algorithm is moving forward and controlled by three variables. They are "idp", "idprolling" and "isimustep". "idp", updated by counting, is to advance the DP interval. "idprolling", updated by counting is to advance DP horizon. "isimustep", updated at every simulation step, is to advance the detection of vehicle arrival. When the time reaches to 7200s, fixed time control is switched back.

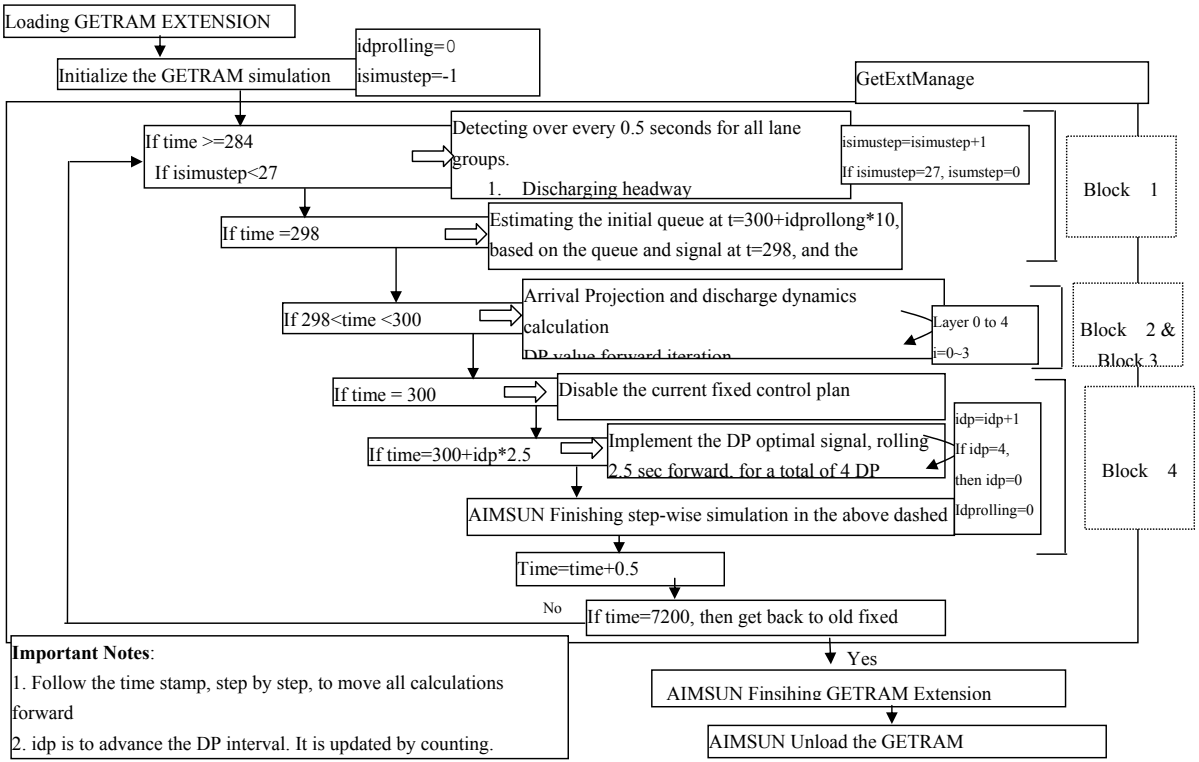


Figure 5 Development of the API for Real-Time Simulation of the DP Algorith

5. Results Analysis

In the API connection for the DP algorithm, the simulation is arranged so that when simulation time is less than 300s (5 minutes), the interchange is operated as fixed time control; when time reaches to 300s, fixed time control is disabled and the DP adaptive control is applied; when 300 < Time < 7200 (2 hours), the interchange is operated as DP control and when time > = 7200 (2 hours), fixed time control is enabled. This provides flexibility in switching back and forth between signal controls. Figure 6 is a snapshot during the simulation run time. The computing of the DP algorithm is updated continuously (in each simulation step) and can be visualized through a Local X console pop-up window as shown in the figure.

The execution of the DP signal control with the time rolling forward is illustrated as follows,

When 284s <= time <= 298s

Detectors start to gather information for every simulation step 0.5 seconds for each eight approaches including the number of vehicles crossing the detectors and vehicles traveling speed.

When time = 298

Estimate the initial queue at t = 300 which is the initial queue for optimization cycle based on the queues and signal at t = 298.

When 298 <time <300

Conduct arrival projection and discharge dynamic calculation, DP value forward iteration and DP optimal signal backward declaration

*When 300 < time < 310 (4*2.5 = 10)*

Implement the DP optimal signal by rolling 2.5 sec forward for 7.5s by adopting majority signal rolling [1]. Time is rolling forward until 2 hours when it switches back to fixed time control.

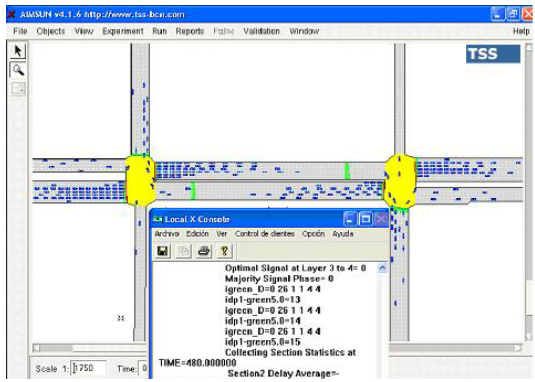


Figure 6 Implementation of the API in Simulation

After successful implementation of the DP algorithm using API in AIMSUN, authors have also conducted sensitivity analysis of the DP algorithm and comparison studies on the performance of three different optimization methods, i.e., the real-time DP algorithm, PASSER III and TRANSYT-7F using the calibrated interchange under varying demand scenarios. Detailed results are provided in the reference (Fang and Elefteriadou, 2005, Fang 2004). In brevity, the research studied how the interchange delay is affected by different definitions of performance measure index and weights on each approach. The simulation results have exhibited that the DP algorithm is superior to PASSER III and TRANSYT-7F in handling demand fluctuations for medium to high flow scenarios when the field demand is increased from the one used in off-line optimization.

6. Conclusions and Recommendations

The research has developed an API to evaluate an adaptive real-time signal control algorithm in AIMSUN. The algorithm is complied by C++ using functions and routines in GETRAM extension module. The developed API allows real time communication between the simulation and the algorithm. The DP algorithm retrieves detected traffic information from the simulation. Based on it, the algorithm optimizes signal control and sends it back to the simulation which consequently controls the signal operations at the diamond interchange. The development of the API in this research has demonstrated the API function is a significantly useful tool to assess external ITS applications when these innovative ides are not supported by the simulation program.

This study is one of the earliest research projects to build an API through AIMSUN simulation. The method developed here provides a reference and some guidelines for future work in this area. It is also believed that comparison studies on

the use of the API in different simulation environment, such as AIMSUN, Paramics, and establishment of a procedure for API development will be necessary to advance the simulation evaluation of ITS applications.

References

AIMSUN (2002) User Manual, Version 4.1, TSS-Transportation Simulation System, Barcelona, Spain

Bartin, B., K. Ozbay, O. Yanmaz. (2005). "Modeling and Simulation of Unconventional Traffic Circles", presented in the 12th World Congress on Intelligent Transport Systems, ITS America, San Francisco, 13p

Chu, L., H. X., Liu, and W. Recker. (2002). **"**Capability-Enhanced Paramics Simulation With Developed API Library", presented in the 9th World Congress on Intelligent Transport Systems, ITS America, Chicago, Illinois, 9p.

CORSIM (1997), Traffic Software Integrated System 97 User's Guide, FHWA - Federal Highway Administration.

Elefteriadou, L., F. C. Fang, J. Mason, R. Roess & E. Prassas. (2005) "Capacity and Quality of Service of Interchange Ramp Terminals", Final Report, National Cooperative Highway Research Program (NCHRP) – 60, Washington, D.C, 85pp.

Fang, F. C. (2004) "Optimal Adaptive Signal Control for Diamond Interchanges Using Dynamic Programming", Ph.D. thesis, 2004 Pennsylvania State University.

Fang, C. F. and L. Elefteriadou. (2005) "Optimal and Adaptive Signal Control of Diamond Interchanges", in Proceedings (CD-ROM) of the 84th Annual Meeting of the Transportation Research Board (TRB), Washington, D.C..

Gardes, Y., A. Kim, and A. D. May. (2003) "Bay Area Simulation And Ramp Metering Study : Year 2 Report", in PATH research report; UCB-ITS-PRR-2003-9, 106p.

Kim, A. M., Y. Gardes, and A. D. May. (2002). "Application Of The Paramics Model In High Occupancy Vehicle Lane Operations, presented in the 9th World Congress on Intelligent Transport Systems, ITS America, Chicago, Illinois, 14p.

Liu, H. X., L. Chu, and W. Recker. (2001) "Paramics Api Development Document For Actuated Signal, Signal Coordination And Ramp Control", in PATH working paper, HCB-ITS-PWP-2001-11, 37p.

Liu, H. X., X. Recker, and W. Wilfred. (2002) "Enhancements Of ATMIS Using Artificial Intelligence", in PATH research report ; UCB-ITS-PRR-2002-30, 50p.

Paramics User Guide (2001), QUADSTONE Ltd., Edinburgh, UK, 2001.

Park, B., I. Yun and L. Choi. (2004) " Development Of An Api For Coordinated Actutated Signal Controller In Paramics", Presented at the ITS America 2004, 14th Annual Meeting and Exposition, San Antonio, Texas, 13p

Smith, I., R. Tudge, and C. Wilson. (2005) "Electronic Tolling on the Sydney Harbour Crossing", ITE 2005 Annual Meeting and Exhibit Compendium of Technical Papers, Melbourne, Australia, 15p.

VISSIM (2000) 3.5 User Manual, VISSIM PTV- Planung Transport Verkehr AG, Germany, 2000.

NON-MOTOR VEHICLE TRIP CHARACTERISTICS AND ITS APPLICATION FOR DISABUSING RELATED POLICIES IN SHANGHAI

Wen XIONG [1] and Xiaohong CHEN [2]

[1] Ph.D Student, Tongji Univ., 1239 Siping Road, Shanghai, 200092, P. R. China, M-Tel) +86-13636460813, Fax) +86-21-65982274, Email) xwart@126.com
[2] Professor, Tongji Univ., 1239 Siping Road, Shanghai, 200092, P. R. China, Tel) +86-21-65982127, Fax) +86-21-65982274, Email) tongjicxh@163.com

Abstract: The insufficient road resource has been unable to satisfy the vehicle demand which has been growing rapidly in Shanghai in recent years. Policy-makers, who are impressed with many cyclists' much higher traffic violations compared with motor vehicle drivers, have tried to carry out measures to forbid non-motor vehicle (NMV) traffic driving on some motor vehicle (MV) lanes of main roads to expand MV's right-of-way. However, the fairness and effectiveness of the measures have been questioned by a huge number of cyclists. From the surveys we organized and the analysis of historical Statistical data from 1986 to 2004, we conclude that in the near future, it is impossible for trips by NMV to be diverted into bus vehicles, and that cyclists should enjoy equal and equitable right-of-way because the efficiency of transporting passengers by NMV is much higher than automobiles in Shanghai. The conflicts between MVs and NMVs should be addressed by proper supervision rather than by forced administrative means.

Keywords: *non-motor vehicle traffic, traffic investigation, transportation policy*

1 Introduction

During the past ten years, Shanghai's non-motor vehicle (NMV) which travels on NMV lane has evolved from traditional bicycle to bicycle and moped powered by electricity, gas or fuel (fuel powered mopeds are no longer used today).

In 2002, "Shanghai Metropolitan Transport White Paper" indicated that "with the development of economy and transit, the operating area and the frequency of utilization of bicycle will gradually decrease and mopeds will be phased out". However, in recent years there is no sign that mopeds are being used less than ever. What's more, with the increase of motor vehicle (MV), the conflicts between MV and NMV appear more serious.

On the other hand, compared with MV drivers, the traffic violations of cyclist are more common, which causes prejudice against bicycles by the decision-makers. They try to carry out measures to separate the MV traffic from NMV, which means through modifying and identifying optional branches for the main roads to serve as

NMV passages so that the main roads can be available only for MV. However, some new constructed NMV lanes with deficient network connectivity, many zigzags, and long-distance round links, always fail to attain the expected efficiency and impartiality of traffic resource allocation. During peak hours, because no other NMV lane can be driven on or the extra detoured distance is not acceptable, some cyclists always ride on the MV exclusive roads.

Aimed at some possible insufficiencies about NMV policy in Shanghai, Four key questions need to be answered:

Question 1: With the public transportation development, have NMV's trips in Shanghai been changed?

Question 2: Based on passengers transporting efficiency, what are the advantages of NMV mode to other modes in Shanghai?

Question 3: In Shanghai, what are the highlighted conflicts between MV and NMV?

Question 4: Regarding to the "Motor vehicle only" policy, what are the most important issues focused on by the concerned cyclists?

2 The change of NMV trip characteristics in Shanghai

Accurately grasping the bicycle traffic volume and its features is the prerequisite which the bicycle transportation policy stipulated for. For the sake of obtaining the change of actual using NMV's characteristics, we analyzed available Shanghai's statistical data from 1986 to 2004 and made supplement field surveys in Shanghai.

2.1 The ownership of NMV in Shanghai

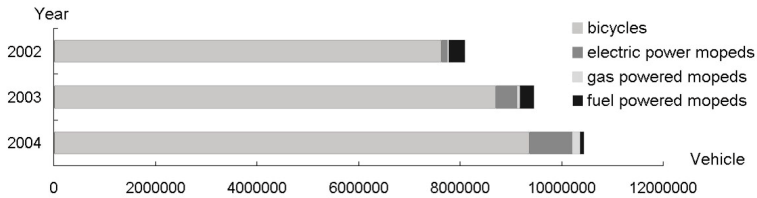


Figure 1. The ownership of NMV (Non Motor Vehicle) in Shanghai

As indicated in Figure 1, by the end of 2004, the ownership of NMV on hand has reached 10444100 in Shanghai, of which the bicycle accounts for 89.7%, electric power moped and gas powered moped accounts for 8.0% and 1.6% respectively, and fuel powered moped which has not been eliminated completely still accounts for 0.7%. Compared to 2003, in 2004 the ownership of bicycle grew by 7.6%, while the ownership of electric powered moped and gas powered moped grew by 104.6% and

135.8% respectively. However, one thing specially needs to be mentioned is that because of a lack of strict bicycle license registration and management rules, the number of bicycles in actual use tends to be smaller than the ownership. According to some estimation, the total number of actual using bicycles approximates 4.7 million in Shanghai.

2.2 Predominant travel model

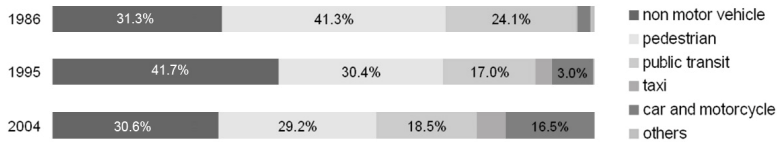


Figure 2. The traffic mode composition of Shanghai residents

Figure 2 reveals that from 1986 to 2004, the trip proportion of NMV in Shanghai increased from 31.3% to 41.7%, and then decreased from 41.7% to 30.6%, while among all the trip modes, the trip proportion of NMV was still the highest. Since the trip ratio of MV had increased quickly (to 40% from 28%), the trip proportion of bus had decreased from 24.1% to 16 %(the trip proportion of rail transit was 2.5% in 2004). From 1986 to 1995, with the urban expansion and the cancellation of monthly bus ticket, a huge number of residents choose the bicycle instead of the bus due to economic consideration. However, from 1995 to 2004, a partiality of residents changed from the bicycle and bus to the automobile, taxi and rail transit because of comfort, convenience and on time records. From 2004 until now, with the deterioration of urban transportation and the transit fare rising, some residents choose the moped as their basic transportation mode instead of mass transit or the bicycle (Figure 1, 2).

2.3 Bicycle transportation volume on roads

Some historical NMV data on several typical roads in Shanghai have been collected, additionally, for the sake of comparing the change of NVM volume, we organized field surveys at the same roads and at a similar period of time corresponding to the historical data.

As indicated in Figure 3, from 2000 to 2005, the NMV volume on main roads has been reduced in diverse degrees, but among which there also are exceptions, such as Siping Road and Datong Road. On Siping road, even the width of NMV lane has been reduced from 5m in 1990 to 2.5m at present, the NMV volume during the morning peak hour has been stable for 16 years due to a lack of alternative bicycle passages available in the survey area. On Datong Road, which is the only bicycle exclusive

passage crossing railway in the survey area, the one-direction NMV volume raised rapidly from 6810 in 1990 to 8477(including 4,076 mopeds) in 2006, which has grown by 24.5% for the last 16 years.

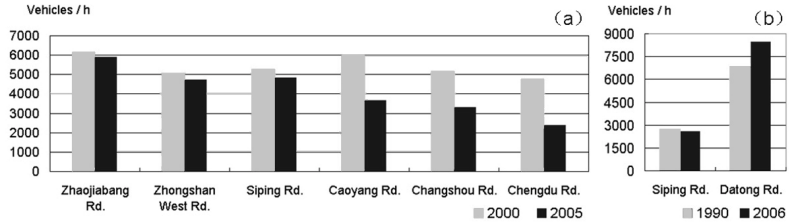


Figure 3. The historical comparison of morning peak hour NMV volume on some roads in Shanghai central city (a) two-direction (b)one-direction

2.4 Trip composition for NMV

In 2004, Shanghai daily bicycle trips reached 12.9 million person trips, of which the moped approximately occupied 17.7 %(17.1% in central city). However, we find that, as indicated in Figure 4, due to the "disordered" development of the moped and the decline of bicycle use, the proportion of mopeds in NMV is far beyond 17% at present, the average proportion has reached 47.8%. Obviously these changes happened, because moped prices declined and rail transit ticket price increasd. Since the improvement in bus service has been inefficient, the ownership of mopeds has not been reduced, instead, it has shown a rapid growth tendency in recent years.

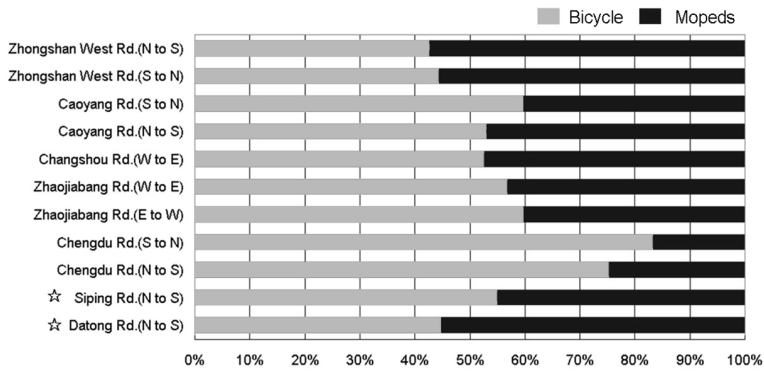


Figure 4. The peak hour composition of NMV on some roads in Shanghai

2.5 Speed and travel time

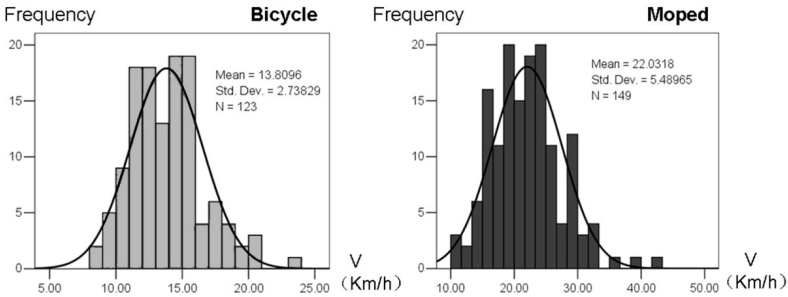


Figure 5. The distribution of NMV speed during non-peak period on Caoyang Rd. (10:30-11:00, 2006.4.18)

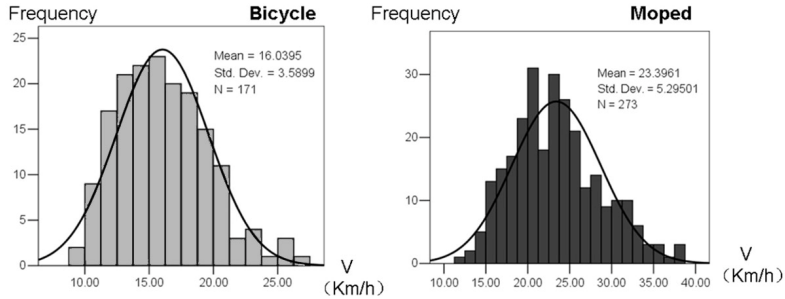


Figure 6. The distribution of NMV speed during peak period on Wuning Rd. (07:30-08:00, 2006.4.25)

As shown in Figure 5 and 6, the Free Flow Speed (FFS) of the moped is much higher than that of the bicycle on NMV lane (4m width). During non-peak hours, the average speed of the moped and bicycle are 22.0 km/h and 13.8 km/h respectively, while during peak hours, the average speed of the moped and bicycle raised to 23.4 km/h and 16.0 km/h. The moped and the bicycle, which are different in speed obviously, both driving on common NMV lanes may cause serious safety problems.

As indicated in Figure 7, from 1986 to 1995, due to urban expansion and central city inhabitants moved out, the travel time consumption of bicycles, pedestrians and bus passengers all increased rapidly. From 1995 to 2004, the travel time of all transportation modes was stable. In 2004, the bus passenger's average travel time consumed and travel distance were 58 minutes and 6.2 km respectively, while the

bicycle and moped's average were 21.7 minutes and 24.1 minutes. Considering the average speed of bicycle and moped, the ideal travel distance of bicycles and mopeds are 1.5 to 5 km and 1.5 to 9 km.

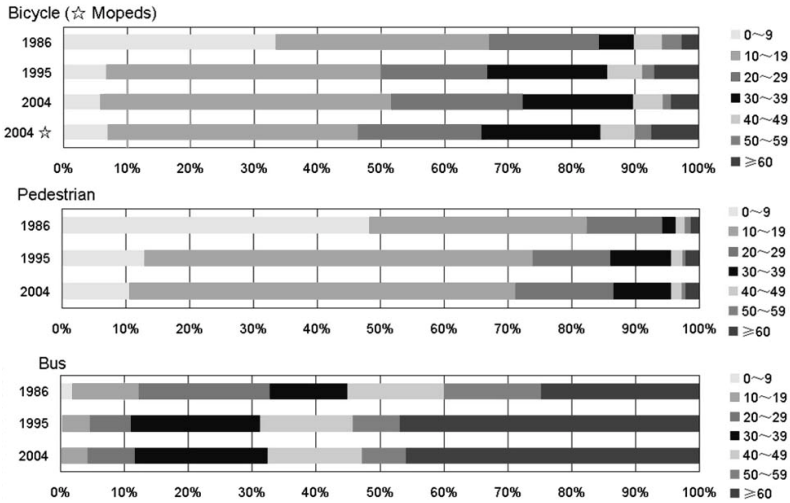


Figure 7. The comparison of Shanghai residents' travel time

In Shanghai, since many resident chose the bicycle instead of walking 9 years ago, the mopeds may become the most common short-distance transportation mode instead of the bicycle and even attract some passengers in the near future, although more and more noise and accidents were brought by mopeds.

3 Transportation efficiency based on carrying passengers

Urban transportation should not be vehicle-centralized but man centralized; we insist that it be improper to evaluate the efficiency of the road only by Passenger Car Unit (PCU) but not including "passengers who have been transported". Two typical roads, where the conflicts between MV and NMV are very common, have been chosen for observation.

Caoyang Road is an important secondary road and many communities and more than twelve bus routes have been assigned near it. For the sake of obtaining the actual volume of cyclists, pedestrians and passengers by bus, car and taxi, consecutive observation was organized at east side of Caoyang Road on January 20th, 2006.

As indicated in Figure 8, during 18 hours, the number of passengers (from north

to south) are 86558 on Caoyang Rd., of which the number of passengers by bus was about 42975(49.7%), the number of passengers by automobile, NMV and taxi, was 14660(16.9%),13241(15.3%) and 5480(6.3%) respectively, the number of pedestrians was 8062(9.3%), and others are about 2140(2.5%). For the efficiency of transiting passengers per lane, the bus is more efficient than NMV, while the automobile is less efficient than NMV, and the taxi is the least efficient mode.

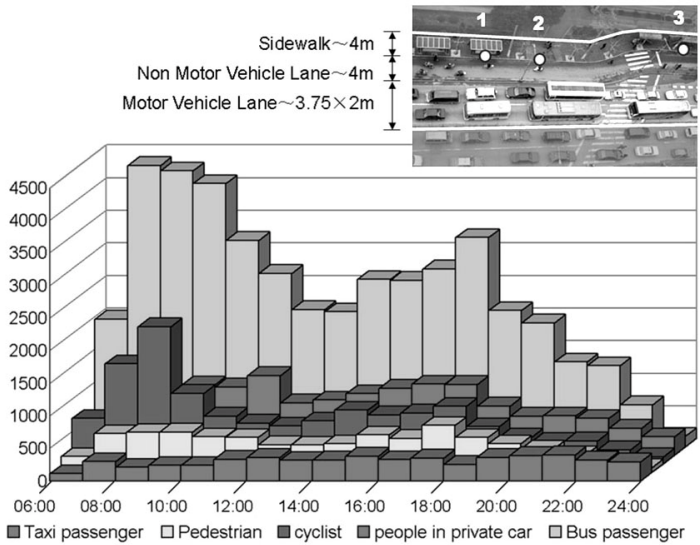


Figure 8. The number of passengers by each mode on Caoyang Rd. (2006.1.20)

Chifeng Road, which connects Zhongshan expressway and Siping Road, is a branch way, including Tongji University, many communities and many bus routes have been placed near it. A similar survey of Caoyang Road had been organized.

As indicated in Figure 9, during 18 hours, the number of passengers (from west to east) are 69964 on Chifeng Road, of which the number of bus passengers was the highest, about 42195(60.3%), the number of passengers by NMV, automobile and taxi, was 11717(16.7%), 3541(5.1%) and 1823(2.6%) respectively, the number of pedestrians was 9050(12.9%), others are about 1638(2.3%). For the efficiency of transiting passengers per lane, the bus is more efficient than NMV, while automobile is much less efficient than NMV, and taxi is the least efficient mode.

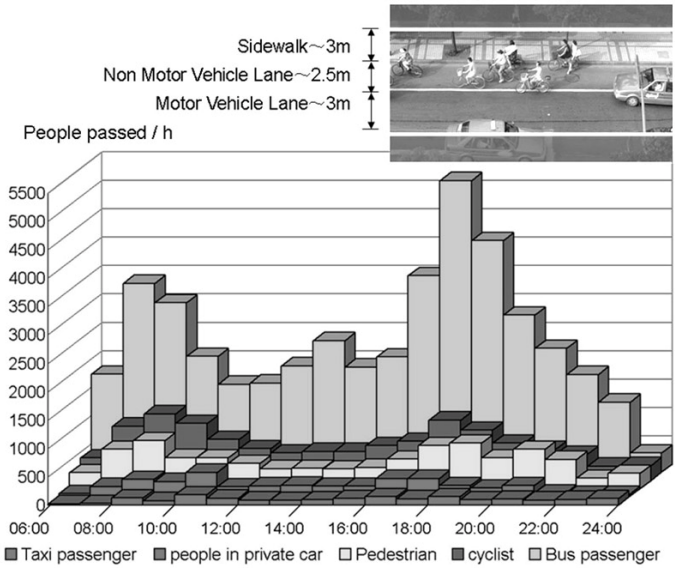


Figure 9. The number of passengers by each mode on Chifeng Rd. (2006.7.7)

From the above surveys, researchers conclude that: for the efficiency of transporting passengers, the bus is much more efficient than both NMV and automobile, while the automobile is less efficient than NMV. Undoubtedly, the bus should have priority of Right-of-way. Compared with automobile, NMV should have the equal Right-of-way at least. In short, "bus and bicycle should be given priority over the automobile", on the basis of the efficiency of transporting passenger may have universal significance to other roads in Shanghai. As revealed in Figure 10, another survey of main roads may prove the above right-of-way allocation principle.

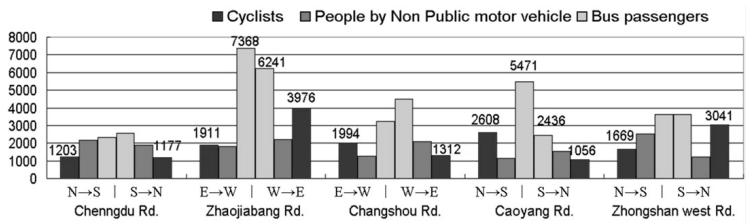


Figure 10. The number of passengers on five arterial roads during morning peak hour in Shanghai central city (2005.12)

4 Conflict analysis

As one of main characteristics of mixed traffic in China, the conflict between MV and NMV is generally one of the factors which influences road capacity for MV. Based on video gathering and analysis, the conflicts between MV and NMV at intersections and road sections have been described and summarized in this paper. The conflicts of Right-Of-Way between MV and NMV also have been discussed.

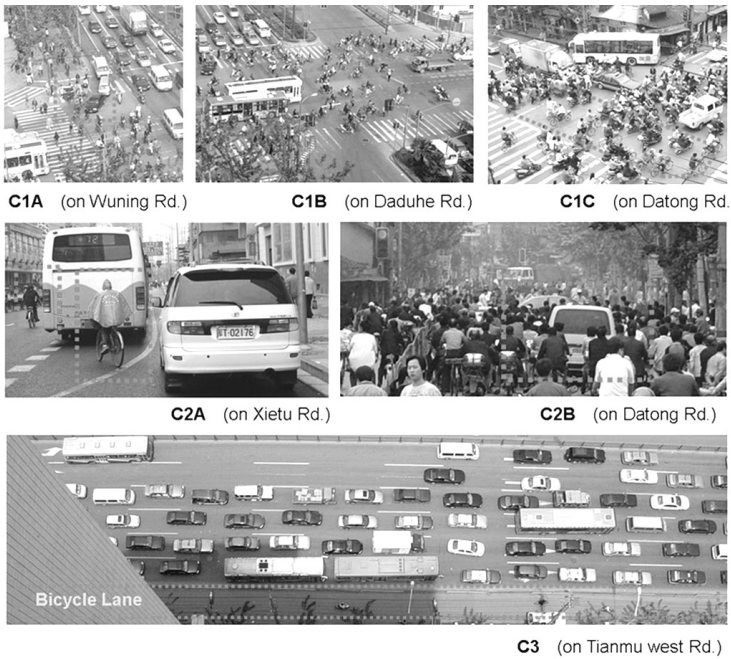


Figure 11. Three categories of conflicts between the MV and NMV in Shanghai

4.1 The conflicts between MV and NMV at intersections

Due to a physical median between the MV lane and NMV lane on arterial roads, there are few conflicts on the links, most of conflicts occur at intersections. Those intersection conflicts could be divided into three kinds(Figure 11, from C1A to C1C).

The first kind of conflict is the same phase period conflict between through NMV and right-turn vehicle. As C1A indicates, the volume from the west approach is about 3342 NMVs at morning peak hour, and the average delay of right-turn vehicle which is caused by through NMV traffic is about 3.7 seconds.

The second kind of conflict is that for one inbound traffic of the same arm, the through NMV traffic competes with though MV traffic at the initial green-phase, for example C1B. When the light turns green, if the NMV volume is very heavy and has not been supervised properly, some NMV traffic will be swelling swiftly into the lanes for MV traffic, which causes mixed traffic at the intersection.

The third kind of conflict is that NMV traffic at the end of red-phase competes with MV traffic at the initial green-phase, for example C1C. When the light turns red, if NMV volume is very big and NMV has not been supervised properly, partial NMV would also run across intersection during red-phase, which causes the delay of MV.

4.2 The conflicts between MV and NMV on mixed-traffic links

Vehicle and NMV always be separated only by markings and signs on branch roads in Shanghai, but these markings and signs are nominal only, conflicts between MV and NMV are very common, as indicated in Figure 11, from C2A to C2B.

The first kind of conflict appears as NMV drives on MV lane (Figure 11, C2A). This may be caused by a small number of cyclists who ignore traffic rules or by other vehicles, especially Transit buses, which may park on NMV lane.

The second kind of conflict can be named "Reef effect", which means that a few vehicles drive into a huge number of NMVs (Figure 11, C2B) .

4.3 Basic contradiction between MV and NMV: Right-of-way

Aforementioned "The conflicts between MV and NMV" is just an appearance even an excuse; the essence of conflict between MV and NMV is that the road space resource has not been assigned properly, much more right-of-way has been allocated to MV.

From 1996 to 2004, since the trip ratio of MV has been increased to 40% from 28%, large-scale construction of arterial roads can hardly meet the demand of increasing MV traffic. The traffic conditions of some road sections are very serious, and the conflicts between MV and NMV has also increasingly risen. Based on the scarcity of main roads and the prejudice of policy-makers towards bicycles, the NMV lane whose utilization rate is low during non-peak hours has been considered by policy-maker to be reconstructed for MV.

As indicated in Figure 11, C3, the utilization rate of NMV lane on Tianmuxi Road is low during non-peak hours. However, refer to "passenger transporting efficiency", the efficiency of NMV is equal to automobile's. We insist that it would be better to use one MV lane but not have an NMV lane being rebuilt as bus lane, and the level of service of bus would be enhanced.

Since many roads are so narrow that it's impossible to increase one bus exclusive lane, the bus exclusive lane planning in Shanghai central city is very difficult to implement. Bus passengers who thirst for speed and efficiency compete with cyclists who insist that accessibility and convenience are their rights.

5 Social surveys for cyclists

As indicated in Figure 8, the width of a one direction MV lane is only 7.5m, due to the roadsides high-intensity land development, it cannot be widened. Along with transportation conditions worsening, the bus lane needs to constructed urgently, however, if policy-makers decide to construct bus exclusive lane with NMV lane, cyclists will have to detour 600m or even more to the other parallel roads in Shanghai. The Policy-makers have planned to gradually convert the NMV lanes into bus exclusive lanes on the three vertical and three horizontal main roads, and to link and rectify the parallel branches serving as NMV passages; however, it could still bring inconvenient for cyclists. Besides efficiency analysis, some research for social equity needs to be carried out immediately.

A survey of cyclists' opinions to "MV only" was organized by us in Shanghai, 2005. Among 122 effective samples, the bicycle cyclists account for 74%.

70 inhabitants (58%) would like to choose a parallel lane which is clearer and safer if the distance between the existing road and parallel lane is short enough. 32 inhabitants are disgusted with the dialogue, 15 inhabitants hate noise, 13 inhabitants can't suffer long red phases along main roads.

Whether the residents can accept the policy of "MV only" or not is related to increased distance and time consumption. As drivers of bicycles, the acceptable increased distance is 247.3 m, and the acceptable increased time is 3.49 minutes. In the survey, when being told that the NMV lane would be reconstructed as a bus exclusive lane, the acceptable increased distance and time is 365 m and 5.06 minutes respectively. As drivers of mopeds, the acceptable increased distance is 325 m, and the acceptable increased time is 2.52 minutes. In the survey, when being told that the NMV lane would be reconstructed for bus only, the acceptable increased distance and time is 393 m and 3.67 minutes respectively.

If the inhabitants cannot accept the extra detoured distance, 8% of them would choose bus instead of NMV, 78% of them would have to choose parallel NMV lane, while 14% of inhabitants would choose to drive on the MV lane.

Obviously, if NMV lane are reconstructed for bus only lanes, the extra detoured distance must be lower than acceptable increased distance of most cyclists. From the conclusions we can also explain why so many cyclists still drive on the lane for MV only in Shanghai central city.

6 Conclusions

Considering physical strength, income levels, riding behavior habits and other personal factors, policy-maker can't forcibly alter inhabitants' choice on transportation modes and trip routes but can only moderately induce inhabitants' riding behavior to

change. Urban transportation policies should be "man-centralized" rather than "motor-centralized". In terms of the above analysis, taking retrospect on the four questions raised in the introduction, we try to disabuse the policies on cyclists in Shanghai as follows:

Answer1: In the near future, to divert those trips by NMV to bus is impossible if the commonwealth and Level of Service (LOS) of public transportation have not being enhanced quickly. Correspondingly, without proper policy, the mopeds may develop out of order, which would bring some problems about security and environment.

Answer2: The right-of-way allocation principle on the basis of the efficiency of transporting passenger should be determinedly implemented. The efficiency of transporting passengers by NMV is no less than automobiles on many main roads, so cyclists should enjoy equal and equitable right-of-way as car drivers there. The number of passengers by bus is almost equal to the total number of passengers by other modes, so Bus passengers should undoubtedly have preferential right-of-way on main roads.

Answer3: It is very common that NMV drives at initial red-phase or on MV lane in Shanghai, but the interference of MV should not be neglected. The conflict is just being transferred to branch roads but not resolved if policy-makers forbid cyclists to drive on NMV lane along main roads. NMV management and transportation education is needed indeed.

Answer4: The inhabitants could accept the policy of "MV only" if NMV lane is used for bus and the increased distance is not great. In the survey, some inhabitants told us emotionally that "they have ridden on some main roads for many years", it would impossible for them to change their travel behavior habits in a short time.

References

The people's government of Shanghai Municipality.(2002). "Shanghai Metropolitan Transport White Paper in 2002"

Bureau of the leading group for the second comprehensive transportation survey of Shanghai.(2004). "The third comprehensive transportation survey of Shanghai"

Bureau of the leading group for the second comprehensive transportation survey of Shanghai.(1998). "The second comprehensive transportation survey of Shanghai"

Shenghong CHEN. (1998). "The Analysis and Forecast of Shanghai Urban Transportation". Shanghai Science and Technology Press, Shanghai,.

Shanghai Transportation Planning net, "the ownership of NMV in 2003 and 2004", http://www.scctpi.gov.cn/chn/content/

The Application of Combinational Evaluation in Traffic Transfer System Based on Multi-objective Programming

Shuai，Bin　　Zhang，XiCheng　　Wang，JiangHong

School of Transportation Engineering, Southwest Jiaotong University, Chengdu China, P.O. Box 610031; PH(086)02887603464; email: shuaibin@home.swjtu.edu.cn

Abstract: Traffic Transfer System is an important subsystem of urban public transportation System and its evaluation can help us to find out its operational problems and solutions. In order to evaluate the layout, operation and service level of it and to optimize its planning schemes; this paper presents five evaluation indices. In practical engineering, the indices were defined; their calculating methods and grade values were given. Aiming at the fault of the traditional objective evaluation and subjective evaluation, this paper presents a combinatorial weighting comprehensive evaluation method which is an amalgamation of objective and subjective evaluation. The proposed method uses *Theil* coefficient to assess the relativity of the sequence and makes the combinational evaluating method to be a multi-objective programming problem. An evaluation model for a traffic transfer system is built up by means of this evaluation method and an example is given to validate the practicability of it.

Key word：traffic transfer system; combinational evaluation；theil coefficient ； multi-objective programming

1 Introduction

Since the 17th century , urban public transportation has become one of the most important parts of the whole urban transportation system, its development level is not only a significant symbol of the national modernization, but also reflects the general economic strength of the country, the development of an urban economy and the citizens' living standard directly as well. The transfer system is an important subsystem of urban transportation system, the guarantee of the city transport priority, and the key to the integration, which is significant for the perfection of the city structure and rationalization of land use.

In this paper, the main factors and indices of evaluating the traffic transfer system were studied. According to the basic principles of subjective and objective weighting methods, we put forward a combinational evaluation method based on Multi-objective Programming and use an example to show the reason for the method.

2 Defining the Traffic Transfer System of Public Transportation Evaluation Index

The traffic transfer system evaluation index is such a synthetic index, that a lot of

factors have to be taken into account. Specific content and the quantities method of the assessing indexes are as follows:

Average transfer times

Average transfer times reflect frequency of vehicle changing route when people are out. In the course of studying average vehicle changing times[1], the literature has put forward the quantities reflection of average vehicle changing times upon road network performance structure, as table 1. The index is an index of cost.

Table 1. Quantities reflection table of average transfer times upon road network performance structure

Average transfer times	0-0.5	0.5-1	1-1.5	1.5-2	above 2
Road network performance	best	better	medium	worse	worst

Vehicle changing facility area per capita

The index means the bus vehicle changing area per capita in the studied area, assessing the bus vehicle changing facility's input index and general supply level in a macro way. The larger the value is, the more the bus system could meet the efficient function of the whole bus system function. Its unit is (㎡/person).

Hub parking facility rate

The index reflects the level of offering staying vehicle changing to mobile and non-mobile in the passenger hub in an area. It is represented by the proportion of area of all mobile and non-mobiles in the studied area to the total area of all passenger hubs. In China, the bicycle's proportion is quite huge, so optimizing bus-bicycle vehicle changing is very helpful for improving public traffic. With the development of urban modernization, the number of private cars has greatly increased, so the vehicle changing of vehicle in traveling is much more frequent and the stop-change—one of the outing means adopted by developed countries –has come up in China. Under such circumstances, the non-mobile and private car parking lots should be increased for the convenience of passengers' vehicle changing. The index is represented by percentage rate.

Vehicle changing service's comfort degree

The index reflects the comfort and satisfaction degree of passengers obtaining the service from the hub and also reflects the service level of various services offered to passengers by the hub. The index belongs to a nature-defining index, and during the assessing process, its general assessing value is a comprehensive reflection in terms of information supply, seat supply or automatic driving equipment supply. Experts should make assessments in the above three aspects and the score is in the manner of percentage.

Rationality of structure

The index is a certain index too, mainly taking three aspects into consideration, including, first, the rationality of the bus change hub station, i.e. the bus change hub

facility should have not only a certain covering area, but also economical factors, the number of hubs shouldn't be too many, and the hub number should meet both outing and economy; second, the matching degree of the change hub and land utilization, the construction of the hub should be surveyed to see whether it combines with the land design according to the urban general planning. Also the following should be surveyed: controlling detailed planning and specific planning, and whether it could maximize the external effect in the place where the hub locates; third, the influence of the hub upon environment pollution, mainly considering the noise and exhaust release into the atmosphere. Experts should give scores to the above three factors, with scores calculated in percentage.

3 A Model of Combinational Evaluation Method Based on Multi-objective Programming

In this paper, we analyze first the basic principles of subjective and objective weighting methods. Then, we put forward a combinational evaluation method resulting in consistency between subjective and objective evaluation in the least squares sense and use an example to show the reason of the method.

There are hundreds of comprehensive evaluation methods at home and abroad[2]. The key to comprehensive evaluation is to determine the weights of indexes.

Currently there are two methods of determining weights: subjective and objective weighting methods. It does not matter which method we use, one is not the better one. It is necessary to combine the advantage of the above two weighting methods[3].

This article first uses multi-objective Programming to get a combinational evaluation method. The synthetic weight is a combination of the subjective weight and objective weight. Therefore, more rationality is possible to make the evaluation results go better in line with the subjective and objective facts.

Let Multi-index evaluation have m evaluating targets, n indexes and k evaluating methods. The weight vector of kth method is:

$$H_k = (h_1^{(k)}, h_2^{(k)}, \Lambda\ , h_m^{(k)})^T, (k = 1,2,\Lambda\ , s)$$

$$\sum_{i=1}^{m} h_i^{(k)} = 1$$

The weight vector of combinational evaluation is:

$$H_0 = (h_1^{(0)}, h_2^{(0)}, \Lambda\ , h_m^{(0)})^T\ ,$$

The research on correlation has been broadly applied in forecast and decision fields *Theil* coefficient could be used to evaluate correlation of serial[4]. *Theil* coefficient of H_k and H_0 is

$$\theta_k = \frac{\sqrt{\frac{1}{m}\sum_{i=1}^{m}(h_i^{(k)} - h_i^{(0)})^2}}{\sqrt{\frac{1}{m}\sum_{i=1}^{m} h_i^{(k)^2}} + \sqrt{\frac{1}{m}\sum_{i=1}^{m} h_i^{(0)^2}}} \quad (k = 1,2,\Lambda\ ,s) \qquad (1)$$

The less of *Theil* coefficient, the more related of H_k and H_0. Let evaluating vector of k kind evaluating methods are $H_1, H_2 \Lambda\ , H_k$ and *Theil* coefficient between H_0 and them are objective function of multi-objective Programming. In order to get the best evaluating effect, we build optimal models as follows:

$$\begin{cases} \min(\theta_1, \theta_2, \Lambda\ \theta_k)^T \\ \sum_{i=1}^{m} h_i^{(0)} = 1 \\ h_i^{(0)} \geq 0, i = 1,2,\Lambda\ ,m \end{cases} \qquad (2)$$

The above evaluation model is a multi-objective Programming problem and it could be solved by related algorithm. A function named Fgoalattain in matlab could be used to solve above multi-objective Programming problem.

4 An Analysis of Practical Examples Based on Calculations

Suppose we start a combinational evaluation of the traffic transfer system of five cities. The index values corresponding to these cities are listed in Table 2. The nature of different indexes determines that Index 1 is negative polarity, while the other four are indexes of positive polarity. We first use the topsis and entropy method to evaluate above cities, which belong to the subjective weighting method and objective weighting method separately, and then the comprehensive evaluating result can be obtained by above combinational evaluation.

Table2 index values corresponding to different cities

Cities	Average transfer time	Vehicle changing facility area per capita	Hub parking facility rate	Vehicle changing service's comfort degree	Rationality of structure
1	0.8	85	8	74	91
2	1.3	68	14	88	86
3	1.1	74	15	84	73
4	0.9	77	10	78	79
5	1.4	82	12	72	82

Obtain evaluating weight vector (H_1) by topsis

from table 2 and evaluating steps of topsis

$H_1 = (0.2136,0.2159,0.2456,0.178,0.1469)^T$

Obtain evaluating weight vector (H_2) by entropy method

from table 2 and evaluating steps of entropy method

$H_2 = (0.1983,0.2016,0.219,0.1994,0.1817)^T$

Obtain evaluating weight vector (H_0) by combinational evaluation

from table 2 and evaluating steps of formula (2)

$H_0 = (0.2059,0.2087,0.2323,0.1888,0.1644)^T$

Evaluating results of three methods are listed as table 3:

Table 3 Evaluating results of three methods

Method	Result Ranking of city 1	Result Ranking of city 2	Result Ranking of city 3	Result Ranking of city 4	Result Ranking of city 5
TOPSIS	3	2	1	4	5
entropy method	3	2	1	4	5
combinational evaluation	3	2	1	4	5

From table 3，evaluating result of above three methods is consistent . H_0 is amalgamation of objective evaluation and subjective evaluation which is efficient and reparative.

5 Conclusion

After establishing the traffic transfer system index system, the article has made a simple summary of the current condition of study of the evaluation. The article has formed the combination evaluation model on the basis of multi-objective planning, combined the subjective evaluation with the objective evaluation. In the article, their *Theil* coefficient is often used to assess the relativity of the sequence, and also, the operators, such as the degree of fuzzy gray, related coefficient, vector angle cosine, etc, are also could be applied in above model. The method sufficiently uses many evaluation methods' useful information to make the traffic transfer system more systematic and rational, and at the same time, has enriched the combination evaluation theory, with its organism and application for further development and improvement.

References

Chen,GuoHong; Cheng,YanTai. The research progress & development trend of comprehensive evaluation methods[A].Proceedings of 2002' International Conference on Management Science & Engineering [C].Harbin:Harbin Institute of Technology(HIT) press,2002

Shi,GuiLan; Deng,Wei; Ge,Liang. Public-transit Evaluation Based on Average Transfer Times[J]. Journal of Luoyang University. 2003，18（4）：26-31

Wang,YingMing. Research on the Methods of Combining Forecasts Based on Correlativity [J]. Rorecasting,2002,21(2):58-62

Xu,Qiang. Research on combinational evaluation [J] Jiangsu Statistics. 2002,10：10-12

Performance Measurement on Arterial Roads in Real Time

James Y. K. Luk[1]

Abstract: Real-time information on how a road network performs is useful to a road manager and a road user. A road manager makes use of the information to manage congestion, e.g. balancing congestion in different parts of a rod network. A road user can also benefit from the same travel information to avoid congestion by re-routing or varying trip start times. Travel times are now displayed on most freeways in many cities in real time. The estimation of arterial travel times is a more challenging task because arterials are interrupted facilities with signals and other control devices interrupting traffic flow. This paper describes the ARRB Travel Time Model for estimating arterial travel times for general traffic and buses. The model retrieves traffic data from the signal control system, SCATS, each minute and has been implemented since May 2006. This paper describes the calibration and validation processes and highlights challenges for estimating travel times without identifying individual vehicles.

Keywords: Performance, Travel time, Delay, Arterial roads, Measurement, Models

Introduction

The performance of a road network in terms of travel time, delay, queue length or other indices has traditionally been measured through surveys which typically use the floating-car technique. This manual method is both costly and labour intensive. Road authorities usually undertake this task several times a year. An example at a national level in Australia is the Austroads National Performance Indicators Program (see http://www.algin.net/austroads).

Congestion in large urban centres has been increasing and the road networks of some capital cities often operate at near- or oversaturated conditions. Traffic incidents such as vehicle crashes, truck load spills, or simply vehicle breakdowns are common occurrences, and planned events such as sporting events, parades and road works add further delay. In such conditions, it becomes critical for road managers to have a clear and up-to-date picture of how the road network is performing in real time. A road manger can then manage the road system more effectively, through changes to incident management plans or signal control strategies and setting appropriate priorities for road improvement projects. The public will also benefit from better traveller information so a road user can re-route, change trip start times, or change mode of travel.

[1]Chief Scientist, Transport Operations, ARRB Group Ltd, 500 Burwood Highway, Vermont South, Victoria Australia 3133, e-mail: james.luk@arrb.com.au

This paper reviews the data sources for monitoring the performance of an urban traffic system in real time. Some past and present system monitoring studies on arterial roads in Melbourne, Australia, are described. These studies include the use of the ARRB Travel Time Model (ATTM) for the estimation of travel times for general traffic and buses. This paper also highlights the challenges for estimating travel times without identifying individual vehicles.

Background on Performance Monitoring

There are a range of data sources for the monitoring of the performance of a road network. These include the following:

- automatic vehicle identification or location devices
- freeway management systems
- advanced signal control systems such as SCATS (Cornwell, Luk and Negus 1986; Lowrie 1992; Luk 1984).

Vehicle Location Technologies

Automatic vehicle location (AVL) technologies identify an individual vehicle in time and space. They allow the calculation of the vehicle's travel time from one location to another. Some key location technologies include:

Signposting and transponders – a network of signposts or beacons is installed to provide location codes to a vehicle (e.g. a bus or tram) that passes by. The transponder in the vehicle receives the code and transmits this code, bus identification (ID) and other data to a central point, which can dispatch route strategies back to the bus.

Satellite navigation – satellites have long been used for navigation. The US Global Positioning System (GPS) is available 24 hours per day free of charge. Other systems include the Russian-built GLONASS and the Euro Galileo system (Khoo and Luk 2002). The maximum accuracy of a GPS system without correction is about 10 m using the code-based technique, and is usable for monitoring traffic performance (higher accuracy can be achieved using correction and phase-coded techniques). Most taxis in Australia are GPS-equipped and can potentially provide travel times. In Singapore, GPS-equipped taxis allowed the local road authority to poll their speeds for the purpose of system monitoring. The Smart Bus in Melbourne also employs the GPS technique along some routes but requires a location fix using a signpost, in the form of a tag embedded in the roadway.

Cellular phone – a mobile phone company knows where a cellular phone is when it is switched on. The location is somewhere within a cell defined by the phone company. The accuracy is dependent on the size of a cell which is typically 26 km^2 and is therefore insufficient for most location applications. It could still be sufficient when the location requirement is not critical, e.g. a logistics company may only need to know whether a vehicle is on the correct side of town as expected. A cellular phone

can now be integrated with a GPS receiver and its location capability is significantly enhanced. The cost of a GPS-cellular phone and data transmission is a concern, and there is the need for a service provider to log, process, and transmit value-added information to users or subscribers.

Tolling systems – another potential source of vehicle location is the current generation of highway speed tolling technologies, or Dedicated Short Range Communications (DSRC). Each vehicle is equipped with a vehicle identification (ID) tag and is identified for billing and other purposes as it passes through a gantry of readers. By knowing the times of passage, vehicle travel times can be calculated.

Note that the stopped times on a route cannot be easily accounted for in a vehicle tagging system and travel time data needs to be carefully screened. The number of tagged vehicles will also determine the accuracy of traffic information and the lack of tagged vehicles in off-peak periods may give the wrong traffic information (Longfoot and Quail 1990). A signal or freeway system operates 24-hour 7-day and provides traffic flow parameters without identifying individual vehicles. They can therefore form the backbone of a performance monitoring system in large cities.

Freeway Management Systems

A freeway management system such as the Drive Time System in Melbourne is a valuable source of system monitoring data. Drive Time relies on a network of inductive loop detector stations at a separation of 500 m. Each detector station consist of two loops per lane. Another key detection parameter is the size of the time slice for aggregating data, which was set to 20 s. These two parameters were originally designed to control the response of a system in detecting incidents on freeways, but have been found useful also for travel time calculation.

The three traffic parameters from a detector station are speed, occupancy and flow per lane. They are usually averaged over a measurement time slice to obtain, e.g. the average spot speed at a station, from which travel time between two stations can be derived and displayed to motorists. Because the spot speeds are time mean speeds that are higher than space mean speeds, there is generally a bias in the estimated travel time.

Another issue is the way a segment or route travel time is calculated. Displayed travel times are based on prevailing traffic conditions. The conditions experienced by a driver passing by a sign are not necessarily the same conditions when the driver arrives at a destination. The displayed travel times often show a lag during shoulder peak periods. For example, a displayed time of 40 min at the city fringe end at 9 a.m. is likely an overestimation of the actual travel time because the same driver will experience less congestion at the town centre-end by the time the driver arrives there. This lag can be overcome by making use of historical travel times, although the use of historical data negates the benefit of a real-time system, especially when non-recurrent congestion or incidents occur.

Signal Control Systems

An advanced signal system such as SCATS employs 16 - 32 loop detectors per intersection. This large quantity of detectors can and should provide valuable data for traffic monitoring. SCATS does have its own performance monitoring facilities at an operational level. It is also possible to process the monitoring data for road managers and road users. In particular, SCATS provides the following information per cycle for intersections designated as critical in the System Monitoring (SM) data:

- cycle time
- green times of all phases
- traffic counts of lanes installed with detectors
- DS values of detectors (representing the degree of saturation).

Other information includes the maximum or saturation flows and fixed parameters embedded in the SCATS database. With these data types, it is possible to determine delay and queue lengths using fundamental signal theory (Akcelik 1983). The basic diagram is as shown in Fig. 1. The concept was exploited to calculate speeds (Doughty et al. 1986), queue lengths (Luk 1992) and, more recently, travel times (Luk and Karl 2004; Luk et al. 2006).

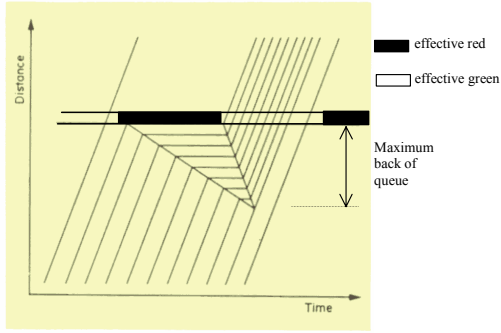


Fig. 1. Delay and Queue Formation at a Traffic Signal

Managing Speeds

In the 1980s, Doughty et al. (1986) reported the use of SCATS data to determine the appropriate driving speed on an arterial road link such that stops could be minimised. The aim was to reduce fuel consumption. The approach was to predict the start of a green phase and the queue length at a downstream intersection by regression. The system was installed along a section of Canterbury Road in Melbourne. The public response was mixed because of the difficulty for drivers not to drive at free speeds or speed limits when traffic volume was low during off-peak periods and on weekends.

The number of stops under advisory speeds was about half of non-advised test runs. The system was dismantled due to changes in SCATS and the lack of funding for on-going maintenance.

The speed-advisory system represents some pioneering work in driver information systems. Much can be learnt from the experience. In particular, it has since been recognised that travel times do not fluctuate as much as speed values and drivers appear more ready to accept travel time than speed information. However, advisory speeds for truck drivers to minimise noise is still a useful application of the concept.

Monitoring Queue Lengths

Instead of forecasting signal timings and queue lengths, Luk (1992) suggested that estimating delay and queue length (and hence travel time) at the end of a cycle with known signal settings was not yet part of the daily operation of a signal control system. Performance monitoring for refining daily signal operation must first be established. From the results of studies in SCATS and TRACS (now known as STREAMS, which is the signal system adopted by Queensland Department of Main Roads), the accuracy of queue length estimation is shown in Table 1.

Table 1. Comparing Observed and Estimated Queue Lengths in Signal Control Systems

	Mean obs. queue (veh)	Mean est. queue (veh)	% diff.	Root mean sq. diff. (veh)	R^2 -terms
SCATS peak period					
Queue at green start	19	16	- 16	8	0.70
Maximum back of queue	22	23	2	12	0.64
SCATS off-peak period					
Queue at green start	11	9	- 23	4	0.79
Maximum back of queue	13	15	9	7	0.53
TRACS through movements					
Queue at green start	3.0	3.1	5	1.6	0.73
Maximum back of queue	3.6	4.5	27	3.5	0.64
TRACS side-streets					
Queue at green start	6.7	4.9	-27	3.4	0.40
Maximum back of queue	7.3	5.0	-31	4.0	0.42

The two empirical studies confirmed the practicality of on-line queue monitoring as part of a signal control system. The level of accuracy at the link level is not expected to be high – the difference between the observed and estimated mean values was about 10 to 30 per cent and the R^2-term was about 0.50. This should not be surprising if traffic counts from loop sensors are currently measured with an error of about ± 5 per cent, and there is no simple way to consider the effect of heavy vehicles on-line.

From a road user's point of view, the lesson therefore is to provide travel times to the public at the level of a route or segment (a route has several segments and a segment consists of several links). From a road agency's point of view, detailed estimates of delay, queue length and travel time will remain invaluable for its operation. The estimation and monitoring of arterial travel times is described in the following sections.

ARRB Travel Time Model (ATTM)

ATTM was developed to meet the challenge of extending the basic theory (Fig. 1) to consider cases of non-uniform arrival and oversaturation. ATTM has the following structure:

Model Travel Time = (Progression Factor x Uniform Delay) + Overflow Delay + Cruise Time Model

This formula states that travel time along a signalised road is made up of the following components:

Uniform Delay – this is defined as the delay due to traffic arriving at a traffic signal with a uniform distribution. This delay is due to the red phase at a traffic signal. Proper signal coordination improves traffic progression and hence reduces uniform delay, which is given by:

$$\text{Uniform delay (sec/veh)} = \frac{0.5 \times r^2}{c(1 - \frac{q}{s})}$$

where r is the effective red time (sec), c is the cycle time (sec), q is traffic flow (veh/sec) and s is the saturation flow (veh/sec).

Progression Factor (PF) – this is in the range 0.0 to 2.6 depending on traffic conditions and how well adjacent signals are coordinated (Highway Capacity Manual of the Transportation Research Board, TRB 2000). If the factor is 0, there is perfect progression. If it is 1, the signal is assumed operating as an isolated intersection.

Overflow delay – Traffic seldom arrives uniformly at a traffic signal so it is necessary to consider an extra delay due to random arrivals and oversaturation due to arrival traffic exceeding the capacity (degree of saturation x > 1). The overflow delay in Akcelik (1983) is given by:

$$\text{Overflow delay (sec/veh)} = 0.25\, T_p \left[z + \left(z^2 + 6(x - x_o) / (Q\, T_p)\right)^{0.5}\right]$$

where

T_p = duration of the analysis period (e.g. T_p = 0.25 h or 15 min)

q_a = arrival flows (veh/h) and q_a can be less or more than the capacity Q

Q = capacity (veh/h) and equal to sg/c

x = degree of saturation, x = arrival flow / capacity = q_a / Q

x_o = degree of saturation below which the traffic delay is zero and speed equals free-flow speed

z = (x - 1).

Cruise time – This is the travel time less stopped delay on a link. It is dependent on traffic and road geometric conditions. Cruise times should be calibrated for each link. A link cruise time increases with congestion due to more vehicle-to-vehicle interactions.

DS in SCATS – The DS value in SCATS is a measurement on how well a green time period is utilised and is available cycle-by-cycle. It is not identical with the x-value in the traditional sense of the degree of saturation over a time period of, say, 15 min. It does correlate quite well with x in undersaturated conditions. DS is structured also to indicate oversaturated conditions at x > 1. The definition of DS is as follows:

$$\text{DS (\%)} = (GT - S_{act} + n\, S_{MF}) \times 100\% \;/\; GT$$

where GT is the green period (sec); S_{act} is the total space time in that green period (sec); n is the number of spaces counted; and S_{MF} is the space at saturation or Maximum Flow (MF, veh/sec). DS can exceed 100% through the choice of an appropriate loop length (about 4.5 m) and when $S_{act} < n\, S_{MF}$. DS is also used to estimate the reconstituted volume (V_k) when oversaturation occurs as follows:

$$V_k = DS \times GT \times MF \quad \text{(veh)}$$

For example, if DS = 105%, GT = 50 sec and MF = 1720 veh/h, then Vk = 25 veh, which will be more than the actual counts because a DS of 105% would imply that vehicle are not moving well in the green phase for this lane. The reconstituted volume represents the real traffic demand in oversaturation conditions, i.e. what the volume could have been if there is enough road capacity. It is used together with DS in ATTM to estimate oversaturation delay.

The following sections outline recent studies on estimating arterial travel times in real time. These studies were funded by the local road authority (VicRoads). The first application of ATTM was carried out on a radial route in Melbourne (Princes Highway East, PHE; see Fig. 2) in 2004. The model was then also applied to Hoddle/Nepean Highway and Springvale Road in 2005. It was extended further for estimating bus

travel times on Hoddle Street and Victoria Parade in 2006 (Luk et al. 2006). The discussions below will focus on the results from the most recent studies.

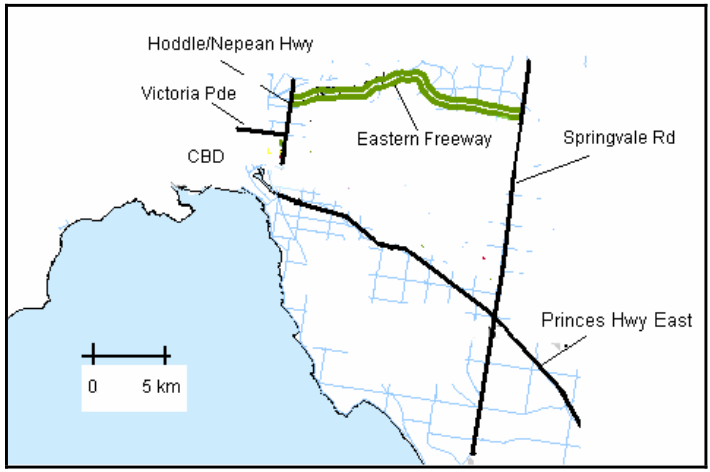


Fig. 2. Travel Time Studies on Arterial Roads in Melbourne, Australia

Arterial Travel Times for General Traffic

The same basic model parameters were found to apply to all study routes. The only difference between the models for the two routes was the different break points to account for SCATS regional boundaries, railway crossings and other situations when platoon progression cannot be maintained and delay could be quite high. A Progression Factor (PF) of 2 was used for the relevant link at such breakpoints.

Fig. 3 shows a typical diagram of the observed and modelled travel time results for Springvale Road in the northbound direction in the a.m. peak. The calibration process requires a certain level of subjective assessment – mainly because the model travel time is based on the information of a signal cycle and represents an average travel time for vehicles departing in the green period of that cycle. On the other hand, the observed travel time in a diagram such as Fig. 3 represents only a single run. Calibrating the model travel time to match a single run is unnecessary and may not be correct – hence the need to calibrate at least six runs and the use of engineering judgement.

Fig. 3 also illustrates a major issue identified from the Springvale Road runs in the northbound direction near a railway level crossing at a suburb called Nunawading. The SCATS detectors cannot properly account for the queues formed south of a railway crossing when the boom gate is activated. A long queue is formed south of the crossing but the SCATS detectors are stop line detectors at a distance away from (and north of) the boom gate.

Possible solutions to this issue include:

- time-of-day, day-of-week factor to account for extra delay
- queue detector to indicate excessive delay in real time
- extra SCATS controller/emulator to provide more Strategic Approach detector information.

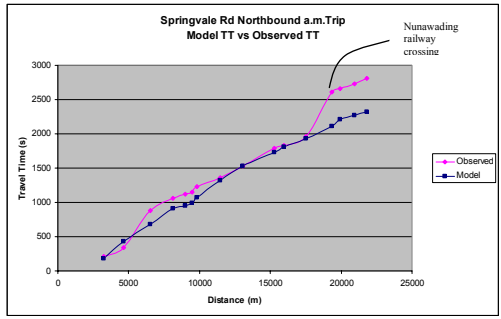


Fig. 3. Travel Times on Springvale Road, Northbound a.m.

Another railway crossing at the Springvale railway station did not create any problem for travel time estimation because the detectors in the northbound direction are close to and south of the crossing. These detectors therefore do provide correct indication of congestion at (and south of) the railway crossing.

The ARRB model was validated with data collected by the local road authority using floating-car surveys or number-plate matching technique at various times for the Springvale Road and Hoddle/Nepean Highway routes. The prevailing SCATS System Monitoring data sets were also available.

Table 2 summarises the statistics for the comparison between observed and modelled travel times. The results for the southbound direction of Hoddle/Nepean Highway are also shown in Fig. 4 for illustration. Because one model was found to apply to both routes (apart from different break points), the validation statistics for both routes are combined at the segment and route levels. Link level validation data is only available for Springvale Road. The results prove that the model developed initially for PHE is robust and can be applied easily to other routes, whether radial or circumferential route types.

Table 2. Comparison Between Observed and Modelled Travel Times

Level of aggregation	Sample size n	Mean % diff.	95% Conf. intervals	R^2-term
Links *from Springvale Rd*	96	17%	± 14%	0.44
Segments *from Springvale Rd and Hoddle/Nepean Hwy*	32	-3.2%	± 8.3%	0.69
Routes *from Springvale Rd and Hoddle/Nepean Hwy*	8	-3.8%	± 8.1%	0.92

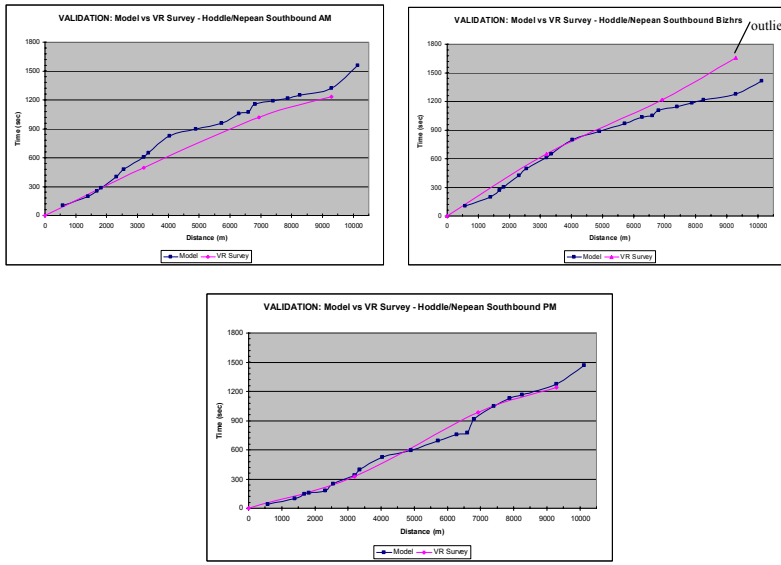


Fig. 4. Validation Results for Hoddle/Nepean Highway, Southbound

Arterial Travel Times for Buses

ARRB was further commissioned to provide car (general traffic) and bus travel times from Hoddle Hwy to St Vincent's Plaza on Victoria Parade in East Melbourne from 7:00 a.m. to 10:30 a.m. in the inbound direction. This project will enable VicRoads to display bus and car travel times from its Variable Message Signs (VMS) on Eastern Freeway to an end point closer to Melbourne CBD as a means to encourage modal shift from car use to bus travel (Fig. 2 and Fig. 5).

Extra parameters and factors such as bus stop delays, bus priority lanes and signal priority have to be taken into consideration when developing ATTM for bus travel. The same ATTM could have been used for bus travel time if detector data for the bus lane was available at all traffic control sites along the route. However, at the beginning of the project, it was found that there was no data for the bus lane for the entire route because of the way the SCATS database was set up for signal operation in that road segment.

Following ARRB's request, VicRoads set up SCATS for bus lane detection at three intersections along Hoddle Street. At the end, only two intersections provided useful data and the data from the third intersection was not useable due to parking and other roadside friction.

Fig. 5. Variable Message Sign Showing Bus and Car Travel Times in Melbourne

As only limited information is available from SCATS, ATTM for bus travel times used a modified algorithm compared to ATTM for cars. The ATTM model for buses consists of two time zones to represent the different prevailing traffic conditions during those periods:

- 7:00 a.m. to 9:30 a.m. - bus lane and signal priority in operation

  ATTM (bus) = function of bus lane traffic volume

- 9:30 a.m. to 10:30 a.m. - no bus lane and signal priority

  ATTM (bus) = ATTM (car) + bus stop delays

Fig. 6 illustrates the operating characteristic between the bus travel times and SCATS vehicle counts on the bus lane in the period when the bus lane was in operation. It consists of two portions. The first portion (an intercept) represents a minimal bus travel time irrespective of the number of buses on the bus lane. The second portion represents the gradient that shows the increases in bus travel times as a function of increasing bus volumes. Both the intercept and the gradient were adjusted

to achieve the best possible correlation between observed and modelled bus travel times.

After 9:30 a.m. buses travel on all normal traffic lanes, and SCATS System Monitoring data can be retrieved at the intersections the same way as for cars. During this period, a reasonable bus travel time model would be based upon the use of the car travel time plus additional delays from the seven bus stops along the route (estimated to be 63 seconds in total). The model is able to produce travel times both at link level and at route level.

The bus model was validated as shown in Fig. 6. The mean percentage difference between model and observed travel times was found to be 3.4%, with an R^2-term of 0.39 and a root-mean squared difference of 1.4 min. The accuracy is reasonable but Fig. 7 shows some biases in estimation. Travel times are generally overestimated for low values and underestimated for higher values (more congestion). The gradient and intercept in Fig. 5 can be fine-tuned. The performance of the bus travel time model is also likely to be improved with more information – extra SCATS detectors on the bus lane.

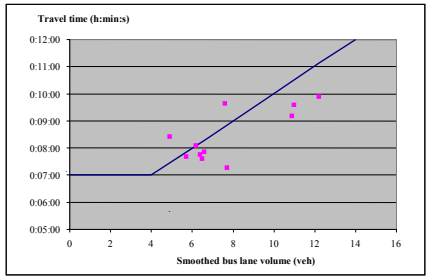


Fig. 6. Bus Travel Time Operating Characteristic from 7:00 to 9:30 a.m.

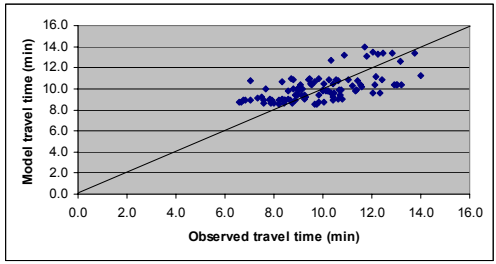


Fig. 7. Bus Model and Observed Travel Times

Conclusions

This paper has demonstrated the feasibility of utilising traffic data from a signal system for the purpose of monitoring the performance of a road network. The data coming from a signal system may not be immediately useful for the purpose of system monitoring. It is necessary to further process the data to produce metrics meaningful to both road managers and road users, e.g. delay, queue length or travel time. The ability for a signal system to produce 24-hour 7-day performance metrics will allow the system to be a solid foundation on which other types of data (e.g. from tagged vehicles) can be added or fused to indicate accurate performance of a road network in real-time.

The ARRB Travel Time Model aims to give accurate minute-by-minute indication of network performance and thus requires data to be available from appropriately located detectors in a road network. This paper highlights the need to install supplementary detectors at sites where the current set of detectors cannot produce the required information for performance calculation. Other approaches for the calculation of performance metrics are possible, e.g. using a time frame of 15 min instead of the 1 min response time in ATTM. This has the advantage of less fluctuation in the calculation of performance metrics. The loss of responsiveness should however be noted – in 15 min a vehicle would have travelled a distance of 15 km at a speed of 60 km/h.

REFERENCES

Akcelik, R. (1983). "Traffic signals: capacity and timing analysis". *Research Report ARR No. 123*. Australian Road Research Board Ltd, Vermont South, Victoria, Australia.

Cornwell, P.R., Luk, J.Y.K. and Negus, B. (1986). "Tram priority in SCATS". *Traffic Engineering and Control*, Vol. 27, No. 11, pp. 561-565.

Doughty, B.W., Nash, D., Van der Vouw, J.W. and Timischenko, H. (1986). "Real-time prediction of adaptive traffic signal timings and queue lengths". *Proc. 13th ARRB/5th REAAA Conf.* Vol. 13, No. 7, pp. 92-106. Australian Road Research Board Ltd, Vermont South, Victoria, Australia.

Khoo, V.H.S. and Luk, J.Y.K. (2002). "GPS technologies and applications in road transport". *Road and Transport Research* Vol. 11, No. 3, pp. 34-45, ARRB Transport Research Ltd, Vermont South, Victoria, Australia.

Longfoot, J.E. and Quail, D. (1990). "Automatic network travel time measuring system – ANTTS". *Proc. 15th ARRB Conf.* Vol. 15, No. 6, pp. 147-170. Australian Road Research Board Ltd, Vermont South, Victoria, Australia.

Lowrie, P. (1992). SCATS: "Sydney Coordinated Adaptive Traffic System: a traffic responsive method of controlling urban traffic". Roads and Traffic Authority, NSW, Sydney, Australia.

Luk, J.Y.K. (1984). "Two traffic responsive area traffic control methods: SCAT and SCOOT". *Traffic Engineering and Control*. Vol. 25, No. 1, pp. 14-18.

Luk, J.Y.K. (1992). "Queue management and monitoring in urban traffic control systems". *Proc. 7th Road Engineering Association of Asia and Australasia Conf. (REAAA) June 22-26, 1992, Singapore.* Vol. 1, pp. 20-28. *Road Engineering Association of Asia and Australasia, Kuala Lumpur, Malaysia.*

Luk, J.Y.K., Karl, C.A., Su M. and Bennett, P. (2006). "Real-time estimation of travel times on arterial roads in Melbourne". *Proc. 22nd ARRB Conf. Oct 29 – Nov 02, 2006, Canberra (14 pages)*. ARRB Group Ltd, Vermont South, Victoria, Australia.

Luk, J.Y.K. and Karl, C.A. (2004). "On-line performance monitoring of urban traffic". *Proc 26th Conf. of Australian Institutes of Transport Research (CAITR), December 8-10, 2004, Melbourne (17 pages)*.

Transportation Research Board (2000). Highway Capacity Manual US Department of Transport. TRB, Washington DC.

THE APPLICATION OF VISUAL INFORMATION INTO FREEWAY OPERATING SPEED PREDICTION MODELING

Zhiqing YANG[1], Xiaoli DU[2] and Zhongyin GUO[3]

[1]School of Transportation Engineering, Tongji University, Shanghai, 4800 Cao'an ST; PH (86) 21-65988812; FAX (86) 21-65981430; email: zhiqing_yang@sina.com
[2]School of Electronics and Information Engineering, Tongji University, Shanghai, 4800 Cao'an ST; PH (86) 21-69584621; FAX (86) 21-65981430; e-mail: du_xiaoli@163.com
[3]School of Transportation Engineering, Tongji University, Shanghai, 4800 Cao'an ST; PH (86) 21-65988812; FAX (86) 21-65981430; e-mail: zhongyin@tongji.edu.cn

Abstract: Operating speed is one of the most important branches in field of Road Safety. Many operating speed prediction models have been presented in the research community, while few take visual information into consideration. It is common sense that the driver's visual information would influence the speed, but it is difficult to be quantified. Based on the features of human vision, this paper proposes a novel concept-- Spatial Sight-Distance (SCSD) to quantify the visual information of drivers. An algorithm is presented to compute the SCSD value abstracted from the visual information of drivers on any location of the road. The algorithm mainly takes into account three factors including the horizontal deviation angle limits, vertical deviation angle limits and terrain undulation. Based on the experimental data, the nonlinear relationship between the expecting speed and SCSD is obtained through regression procedure, and finally the operating speed prediction model is established by the point-by-point grade calibration of nonlinear relationship through TWOPAS model. The experiment results show that the operating speed prediction model based on the visual information matches the reality pretty well in the case of free flow. This operating speed prediction model based on visual information will be very useful to the research work on road safety, road safety audit, alignment assessment, accident black-pot discrimination and accident cause analysis etc.
Key words: *visual information, Spatial Sight-Distance, operating speed, road safety, speed prediction*

1 Introduction

Operating speed, as the final result and external exhibition of the dynamic system- "Person- Vehicle- Road", has been reported in much of the research that it has direct or indirect influence on almost 30 percent of fatal accidents and 12 percent of all accidents. Therefore, it is essential examine the operating speed. Nowadays, there are

many operating speed prediction models, but most of them establish a regression equation between speed and alignment indices by statistic method. For example, the German researchers mainly take the curvature change rate (CCR) and grade into account, and the curve radius and grade are the main factors considered in the Australian research field. Dr Fambro modified the design speed through statistic methods to obtain the operating speed (Du, Boying, 2003). The IHSDM strongly recommended by American researchers integrated all sorts of alignment indices to establish regression equations, and then predicted the operating speed under diversified conditions (Federal Highway Administration, Turner-Fairbank Highway Reserch Center, 1999). Humans, however, who are the dominant factor of dynamic driving system, have rarely been considered in the operating speed prediction, with no mention of the relationship between the driver's visual information and speed.

The literature (Wang, Wuhong, 2000) shows that the available range of vision of the driver varies inversely as the vehicle's speed. But it is difficult to quantify visual information. Although no one denies that there are certain relationships between speed and visual information, predicting operating speed cooperated with visual information is hardly seen in the field.

Based on the characteristics of vision, this paper proposes a novel concept -- Spatial Sight-Distance (SCSD), which is totally different from the former concept of Sight-Distance, to quantify visual information. The former Sight-Distances are generally determined by motion principles, and take the response time of the driver, the speed of vehicle, and the friction coefficient of the pavement as parameters. While Spatial Sight-Distance (SCSD) is determined by vision principles, and take the characters of vision as parameters. In this paper, the relationship between speed and SCSD is established through experiments, followed by application and verification, and then an operating speed prediction model based on visual information is proposed.

The remainder of this paper is organized as follows: Section 2 proposes the concept of Spatial Sight-Distance (SCSD).The algorithm of SCSD is presented in Section 3. The prediction model is outlined in Section 4. Section 5 verifies the model through some applications and shows the test results. Section 6 summaries this paper.

2 Spatial Sight-Distance notion

2.1 Definition

We learn from the eyesight checking that the discrimination of the eyes on focus points is the best, and will decrease quickly with the deflecting of focus. In order to discern the objects clearly, people have to keep turning their heads and eyes to aim their focus on the objects.

As shown in literature (Wang, Wuhong, 2000), objects within 12° could be seen clearly. When sight deflects from focus, the eyesight will decrease accordingly. During driving, the driver usually keeps moving his sight to ensure the objects he cares about in the clearest eyeshot. The object that the driver cares about most during driving is

the road ahead, so he will always focus on road rather than anything else.

Through the above analyses, the longest distance that can be seen clearly in road ahead is defined as Spatial Sight-Distance (SCSD) in this paper. Furthermore, SCSD is defined as the distance between eye-point and the farthest point in lane that can be seen from above H(m) of pavement and within $\pm \alpha°$ along road. As shown in Figure 1 and Figure 2, H refers to the height of eye; α refers to horizontal angle, which is used to identify the focus of eye; and β refers to vertical angle, which is correlated to speed.

By Spatial Sight-Distance, the unimportant factors in visual information would be neglected, and the factors tightly related to driving could be abstracted out and quantified, and then served for operating speed prediction.

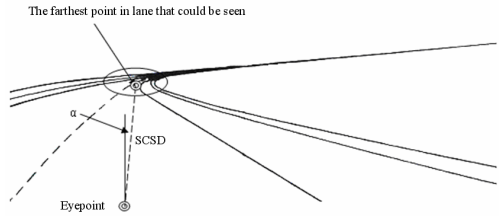


Figure 1 Horizontal Draft of SCSD

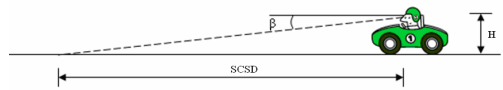


Figure 2 Vertical Draft of SCSD

2.2 Hypothesis

SCSD, however, is inadequate to be used to predict operating speed. There is an extreme example. When a vehicle running in 5 km/h, which is approximately equal to the speed of foot passenger, there seems to be no correlations between SCSD and vehicle speed. The reason is that the driver could keep that speed in spite of the variety of SCSD. However, the road should not only provide accessibility, but also satisfies the economical and fast requirements, which would be inaccurate whenthe vehicleis running at low speed during congestion or malfunction. Although these situations are ordinary phenomena, they should not be considered when focusing on the correlations between alignment and speed. So we assume that drivers always drive vehicles under the fastest and most comfortable speed the vehicle could reach in free flow.

Under this hypothesis, the relationship between SCSD and the speed could be established explicitly. If the SCSD fluctuates heavily and frequently during driving, it will put a heavy mental influence on the driver, and results in speeding up or slowing down the vehicle repeatedly. On the contrary, if the SCSD is stable or rarely fluctuates, the driver usually feels comfortable and will keep the vehicle running at a smooth

speed. So, the correlation between SCSD and speed could be used to predict the operating speed.

3. The Algorithm of SCSD

According to the strip-like characteristic of road and the vehicle running direction, the road could be abstracted as a directed 3D space curve C, which is shown in equation (1).

$$\begin{cases} F(x,y,z)=0 \\ G(x,y,z)=0 \end{cases} \tag{1}$$

Suppose there is a car located at point A:(x_0, y_0, z_0), where $(x_0, y_0, z_0) \in C$, then the position of eye-point of the driver is B:$(x_0, y_0, z_0$+H). Because the sight of the driver is restrained by three factors such as horizontal deviation angle limit, vertical deviation angle limit and terrain undulation. This paper will calculate the SCSD value based on those three limits.

3.1 Horizontal Deviation Angle Limit

Projecting the 3D curve C on plane xoy, we can get the 2D curve C', which is shown in equation (2). Then the projection vector of the tangent that passes through point A to curve C on plane xoy shows as equation (3). Obviously, the vector of line AB could be defined in equation (4). Equation (5) determines the normal vector of the tangent plane that passes through point A.

$$\begin{cases} H(x,y)=0 \\ Z=0 \end{cases} \tag{2}$$

$$\boldsymbol{a}=\left\{-H_y(x_0,y_0), H_x(x_0,y_0), 0\right\} \tag{3}$$

$$\boldsymbol{n}_{AB}=\left\{0,0,H\right\} \tag{4}$$

$$\boldsymbol{n}_a=\boldsymbol{a}\times\boldsymbol{n}_{AB} \tag{5}$$

Because of the horizontal deviation angle limits, there must be two planes I_1 and I_1'. The angle between I_1 and the tangent plane passing through point A is α, and $-\alpha$ between I_1' and the same tangent plane. Figure 3 shows their ichnography. Equation (6) and (7) determine the normal vectors $\boldsymbol{n}_1$ and $\boldsymbol{n}_1$' of planes I_1 and I_1'.

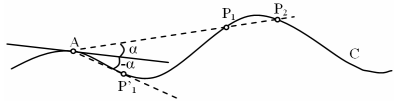


Figure 3 Ichnography of Horizontal Deviation Angle Limit

$$\begin{cases} (\boldsymbol{n}_1, \hat{\boldsymbol{n}}_a)=\alpha \\ \boldsymbol{n}_1\cdot\boldsymbol{n}_{AB}=0 \end{cases} \tag{6}$$

$$\begin{cases} (\boldsymbol{n}_1', \hat{\boldsymbol{n}}_a)=-\alpha \\ \boldsymbol{n}_1'\cdot\boldsymbol{n}_{AB}=0 \end{cases} \tag{7}$$

Given the normal vectors $\boldsymbol{n}_1$ and $\boldsymbol{n}_1$', we could construct the equations of planes I_1

and I_1', which are shown as equation (8) and (9). Afterward, through the simultaneous equations (10) and (11), we can get two groups of points of plane-curve intersection between plane I_1 or I_1' and curve C, which is shown in equation (12) and (13). Finally, the longest distance limited by horizontal deviation angle that could be seen by drivers is calculated by equation (14).

$$I_1(x,y,z) = \boldsymbol{n}_1 \cdot \boldsymbol{n}_{AM} = 0 \tag{8}$$

where M is a random point in 3D space.

$$I_1'(x,y,z) = \boldsymbol{n}_1' \cdot \boldsymbol{n}_{AM} = 0 \tag{9}$$

$$\begin{cases} F(x,y,z) = 0 \\ G(x,y,z) = 0 \\ I_1(x,y,z) = 0 \end{cases} \tag{10}$$

$$\begin{cases} F(x,y,z) = 0 \\ G(x,y,z) = 0 \\ I_1'(x,y,z) = 0 \end{cases} \tag{11}$$

$$P = \{P_1, P_2, \mathrm{K}\ , P_n\} \tag{12}$$

$$P' = \{P_1', P_2', \mathrm{K}\ , P_m'\} \tag{13}$$

$$Z_1 = \min(\min(|\overrightarrow{AP_1}|, |\overrightarrow{AP_2}|, \mathrm{L}\ , |\overrightarrow{AP_n}|), \min(|\overrightarrow{AP_1'}|, |\overrightarrow{AP_2'}|, \mathrm{L}\ , |\overrightarrow{AP_m'}|)) \tag{14}$$

3.2 Vertical Deviation Angle Limit

Given the equation of curve C and a point A:(x_0, y_0, z_0) on C, obviously, equation (15) is the tangent vector passing through point A to curve C. Then, the normal vector of road strip on point A could be determined in equation (18).

$$\boldsymbol{n}_T = \{1, \varphi'(x_0), \phi'(x_0)\} \tag{15}$$

with

$$\varphi'(x_0) = \frac{\begin{vmatrix} F_z & F_x \\ G_z & G_x \end{vmatrix}_0}{\begin{vmatrix} F_y & F_z \\ G_y & G_z \end{vmatrix}_0} \tag{16}$$

$$\phi'(x_0) = \frac{\begin{vmatrix} F_x & F_y \\ G_x & G_y \end{vmatrix}_0}{\begin{vmatrix} F_y & F_z \\ G_y & G_z \end{vmatrix}_0} \tag{17}$$

$$\boldsymbol{n}_2 = \boldsymbol{n}_T \times \boldsymbol{n}_a \tag{18}$$

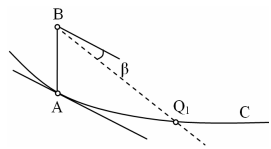


Figure 4 Solid Chart of Vertical Deviation Angle Limit

As Figure 4 shows, the normal vector $\boldsymbol{n}_b$ of sight plane I_3 could be calculated through equation (19). And then the equation of sight plane I_3 is established in equation (20).

$$\begin{cases} (\boldsymbol{n}_b, \hat{\boldsymbol{n}}_2) = \beta \\ \boldsymbol{n}_b \cdot \boldsymbol{n}_a = 0 \end{cases} \tag{19}$$

$$I_3(x, y, z) = \boldsymbol{n}_b \cdot \boldsymbol{n}_{BM} = 0 \tag{20}$$

Based on the following simultaneous equation (21), a group of points of plane-curve intersection between plane I_3 and curve C could be determined and shown in equation (22). Then the longest distance limited by vertical deviation angle seen by the driver could be calculated by equation (23).

$$\begin{cases} F(x, y, z) = 0 \\ G(x, y, z) = 0 \\ I_3(x, y, z) = 0 \end{cases} \tag{21}$$

$$Q = \{Q_1, Q_2, \mathrm{K}, Q_i\} \tag{22}$$

$$Z_2 = \min(|AQ_1|, |AQ_2|, \mathrm{L}, |AQ_i|) \tag{23}$$

3.3 Terrain Undulation

As a result of terrain undulation, some road segments will be hidden from the sight of driver, which usually be called "the lack of sight-distance". Suppose there is a random point N:$(x_0, y_0, z_0) \in C$ as shown in Figure 5, then the vector of sight would be $\boldsymbol{n}_{BN}$. Knowing the vector of line AB that is shown in equation (24), the angle between vector $\boldsymbol{n}_{BN}$ and $\boldsymbol{n}_{AB}$ could be defined in equation (25), which is a function with single degree of freedom along curve C. Let the partial derivative of θ equal to zero (see equation (26)), then the solution set of point N could be obtained and shown in equation (27). Finally, the longest distance limited by terrain undulation that could be seen by driver is calculated in equation (28).

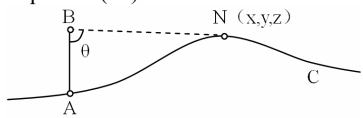


Figure 5 Solid Chart of Terrain Undulation

$$\boldsymbol{n}_{AB} = \{0, 0, H\} \tag{24}$$

$$\theta = (\boldsymbol{n}_{BN}, \hat{\boldsymbol{n}}_{AB}) \tag{25}$$

$$\frac{\partial \theta}{\partial s} = 0 \tag{26}$$

where s is the arc length.

$$R = \{R_1, R_2, \mathrm{K}, R_j\} \tag{27}$$

$$Z_3 = \min(\left|\frac{1}{AR_1}\right|, \left|\frac{1}{AR_2}\right|, \mathrm{L}\ , \left|\frac{1}{AR_j}\right|) \tag{28}$$

Integrating the three results Z_1, Z_2 and Z_3, which refer to the largest distance that could be seen by the driver under different situations respectively, we can obtain the Spatial Sight-Distance(SCSD) value on location A:(x_0, y_0, z_0) through equation (29).

$$SCSD_A = \min(Z_1, Z_2, Z_3) \tag{29}$$

4 Speed Prediction Model

4.1 The Correlation between SCSD and Speed

In order to find the correlation between SCSD and speed, many experiments have been performed. But the correlation between SCSD and speed can hardly be found. The reason is factors that can affect the speed of vehicle include not only SCSD, but also other factors such as weather, performance of vehicle, pavement, signs, driving skills and emotion of driver. Random tests, therefore, can not be used to analyze the correlation between SCSD and speed.

A selection must be made to minimize the influence of other factors like signs, volume, slope etc. in order to ensure the correlation between SCSD and speed can be found. Another test obeying the following 4 assumptions on passenger car traffic has been performed. Firstly, there are no ramps or signs in test road section; secondly, the traffic is free flow; thirdly, SCSD should be kept with the same value in a long distance, and fourthly, the road sections are level. These 4 assumptions are used to exclude the effect of other factors. The test results are shown in Figure 6, in which Y coordinate refers to 85th percentile speed, and X coordinate refers to SCSD value.

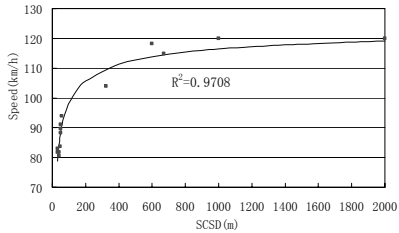


Figure 6 Tests of Passenger Car with 4 Assumptions: Speed versus SCSD

According to the distribution of data points shown in Figure 6, a nonlinear correlation and the property of convergence between SCSD and speed can be found easily. Thus the regression equation of passenger car is shown in equation (30).

$$V_d = 210.984\left(1 - \frac{1}{e^{\log SCSD}}\right) - 84.072 \tag{30}$$

where V_d is the driver desired speed (km/h).

The V_d on each location of road can be obtained from Equation (30).

A simulation shown in Figure 7 is carried out to validate the Equation (30).We see that almost all data points scatter in the right side of the curve, especially those

in the range of 40-80 km/h, which corresponds to equation (30) quite well. It illustrates that regression Equation (30) is the limit situation of driving, and will be reasonable only when the driver is not influenced by other factors and keeps the vehicle at the fastest and most comfortable speed the vehicle can reach synchronously. This simulation result matches our former assumptions well.

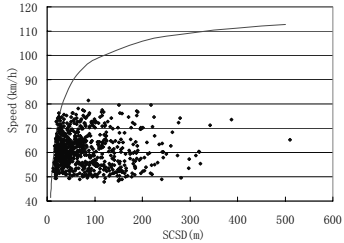


Figure 7 Microscopic Simulation: SCSD versus Speed

4.2 Grade Calibration

The V_d obtained from Equation (30) is not the prediction speed of vehicle, because the grade effects can not be ignored, and the V_d on the neighboring location should also be taken into account. Drivers attempt to maintain their desired speed, but may be limited from doing so by the performance capabilities of their vehicle on a grade. In order to estimate the real vehicle speed, the TWOPAS model is used for calibrating, this is a microscopic simulation model of traffic on two-lane rural highway. More details about TWOPAS can be found in literature (Federal Highway Administration, Turner-Fairbank Highway Reserch Center, 2000).

5. Application and Verification

According to the operating speed prediction model presented in this paper, the SCSD value on each location could be calculated firstly from the given alignment indices of road, and then the desired speed of driver on each location could be obtained, which is followed by grade calibration one by one. Finally, we can acquire the distribution of operating speed along road.

5.1 The Comparison between Different Operating Speed Prediction Models

Many operating speed prediction models have been presented in the research community. In order to validate the speed prediction model proposed in this paper, we chose the IHSDM model for comparison, which is very popular in the USA.

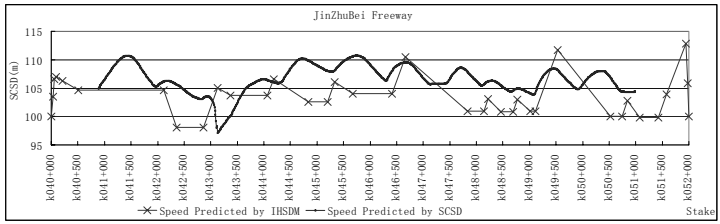


Figure 8 JingZhuBei Freeway: Comparison of Predicted Speed by Different Models

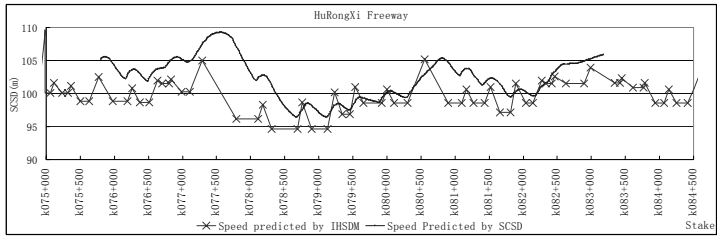


Figure 9 HuRongXi Freeway: Comparison of Predicted Speed by Different Models

From Figure 8 and Figure 9, We see that speed curves generated from SCSD and IHSDM respectively are very close. There are little differences for the most part of the curves, and the biggest one is no more than 10km/h. Another interesting that thing should be noticed is that the trends of two curves are coincident basically, even in the sections with biggest difference.

The operating speed prediction model based on visual information is developed on driving behavior theory and relies on the relationship between speed and SCSD. While IHSDM model is constructed on statistics, it uses statistic methods to establish a relationship between speed and alignment indices, and then predict operating speed. Although there are tiny differences between the results of two models, we can still conclude that the speed predicted by two models on HuRongXi Freeway, which is in HuBei province of China, and JingZhuBei Freeway, which is in GuangDong province of China, are coincident, and SCSD theory accords with the statistic rule. Therefore, it could be considered objective and scientific, and could be used in the field of road safety.

5.2 The Comparison between Predicted Speed and Realistic Speed

According to the flow, the operating speeds of YunSan freeway and TaiJiu freeway in ShanXi province of China are predicted by the model presented in this paper, and the comparisons between realistic speed and predicted speed are shown in Figure 10 and Figure 11, to validate and verify the model.

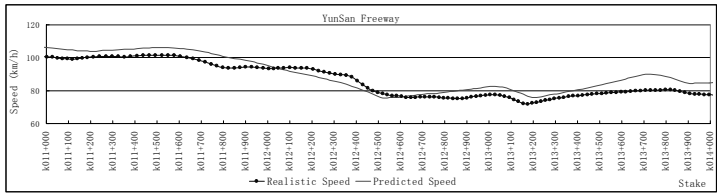


Figure 10 YunSan Freeway: Realistic Speed versus Predicted Speed

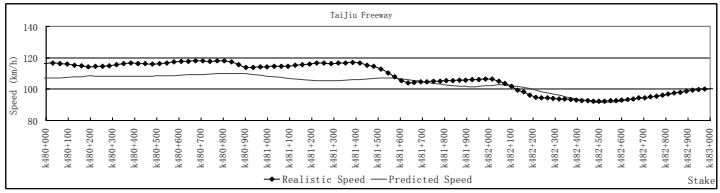


Figure 11 TaiJiu Freeway: Realistic Speed versus Predicted Speed

From Figure 10 and Figure 11, it is easy to see that the predicted speeds match realistic speeds fairly well on the above two freeways, and could represent the real driving process with relative high reliability. But, it should be mentioned that there are noticeable differences between predicted speed and realistic speed in none-free flow, and there is a need for further research in none-free flow.

6. Conclusion

The Driver is the dominant factor of the dynamic driving system, and plays an important role during driving. This paper, starts from a new point of view --the visual information of driver, analyzes the characteristics of visual information in order to research road. A novel concept, "Spatial Sight-Distance (SCSD)", different with the former concept Sight-Distance is presented to quantify visual information. It not only reflects the fluctuation of alignment, but also has significant effect on speed of vehicle. The model based on visual information could predict operating speed in free flow, and match reality fairly well.

This operating speed prediction model based on visual information will be one of the effective research tools in the field of road safety research, road safety audit, alignment assessment, accident black-pot discrimination and accident cause analysis etc. Hopefully, the method presented in this paper might bring some inspiration to other researchers.

References

Du, Boying (2003). "Operating Speed and Road Alignment." *Master thesis*, School of Transportation Engineering, Tongji University, China.

Federal Highway Administration, Turner-Fairbank Highway Reserch Center (1999). "Speed prediction for two-lane rural highways." 99-171, USA.

Federal Highway Administration, Turner-Fairbank Highway Reserch Center (2000). "The Capability and Enhancement of VDANL and TWOPAS for Analyzing Vehicle Performance on Upgrades and Downgrades within IHSDM." 00-078, USA.

Wang, Wuhong (2000). "Theory and Methods of Driving Behavior in Road and Traffic System." Beijing.

OPTIMIZING PASSENGER TRAIN SERVICES: A MODEL AND A GENETIC ALGORITHM SOLUTION

Jing TENG[1] , Baoming HAN[2] and Winggun WONG[3]

[1]Department of Transport Management Engineering, Tongji University, Shanghai 200092, P.R.China; PH (86) 021-51030222; FAX 021-65988372; email: tengjing@mail.tongji.edu.cn
[2]Department of Transport Management Engineering, Beijing Jiaotong University, Beijing 100044, P.R.China; PH (86) 010-51688518; FAX (010) 010-51688518; email: bmhan@center.njtu.edu.cn
[3]Department of Civil & Structural Engineering, The Hong Kong Polytechnic University, Hong Kong, P.R.China; PH (852) 27666067; FAX (852) 23346389; email:cewgwong@polyu.edu.hk

Abstract: Planning for passenger train services plays a vitally important strategic role in railway transportation. The objective of such planning is to achieve maximum net revenue for the rail operator whilst maximizing some specified functions of overall community benefit. However, there usually exists a trade-off between attracting passenger traffic and reducing operating costs. This paper puts forward a non-linear integer-programming model to optimize that trade-off. A Genetic Algorithm (GA) is introduced to solve the optimization problem efficiently. The model is described in the paper, together with the solution method. The paper also gives an example application of the proposed model by designing a passenger train service plan for China's GuangShen quasi-high-speed railway. The solution process is outlined in detail. A GA was found to be an effective technique to solve the stated transportation optimization problem. The procedure adopted here allows for practical considerations to be taken into account by providing a range of sub-optimal solutions in addition to the global optimal solution.
Keywords: *passenger train service plan, genetic algorithms, railway operations, railway optimization*

1 Introduction

Plans for passenger train services include train trip details, train stop-schemes and service frequency specification. Such planning mainly considers two aspects, namely, the demand for passenger trips and the operating cost to the operator or service provider. For the transport operator, the aim is to adequately take advantage of the transportation plant and equipment and attract passenger traffic, thereby maximizing net revenue (gross revenue minus costs) or profit. Therefore, passenger train service plans represent a critical link between transportation demand and supply. As such, those plans can be seen as making up the strategic level arm of the train scheduling decision.

Chang (2000) defined train services, and his study mainly focused on establishing a double-objective model for the best compromise between the minimization of passengers' total travel time and the minimization of operating costs. Chang (2000) used fuzzy mathematical programming to solve his model. Brännlund (1998) established a very large integer-programming model to schedule trains to obtain a profit-maximizing timetable, with track capacity constraints. He used a Langrangian relaxation solution approach to solve the optimization problem. Gorman (1998) offered an operating-plan model (OPM) to improve Santa Fe Railway's service level. The objective of OPM is to minimize the schedule related costs of services subject to rail-operating capabilities while meeting customers' expectations for freight services. The method used to solve OPM is a combination of genetic and tabu searches. It was claimed that applying OPM successfully improved Santa Fe Railway service by four percent while reducing costs by six percent over the existing operating plan.

The above optimization models are solved using various solution methods, whilst the objectives are all similar with a focus on profit maximization. Referring to the decision levels first proposed by Assad (1980a), the model in Chang (2000) can be considered a tactical planning tool. The models proposed by Brännlund (1998) and Gorman (1998) can be seen more as operational tools.

The profit maximization problem needs to be considered in all its facets. Firstly, it is necessary to look at gross revenues and operating costs; secondly, strategic decisions about what markets are to be serviced and the characteristics of those markets need to be taken into account; thirdly, the relationship between levels of service and likely passenger demand needs to be included in the analysis. The current paper focuses on the optimization of a passenger train service plan from a marketing perspective. Usually, passenger transportation service levels are evaluated by using the concept of service utility, which can be transformed into either travel time or a monetary value. Such passenger utility often includes some quality indices, such as comfort, safety and convenience and some quantifiable indices, such as travel time, service frequency, reliability of arrivals and delay times.

Passenger train services need to be planned so as to resolve the conflicting objectives of increasing passenger demand whilst controlling operating costs. The optimal resolution of such conflicts is particularly relevant when the rail operator is facing strong modal competition from private car trips and/or other public transportation modes. This paper deals with a means for the optimization of train service plans under operating cost constraints. Figure 1 shows diagrammatically the influencing factors that need to be taken into account to maximize profitability.

The twin role played by service plans in influencing mode share and operator revenue is evident from Figure 1. In that sense, service plans affect the operator directly and the community indirectly, through their impact on demand for other modes and hence on the community in general.

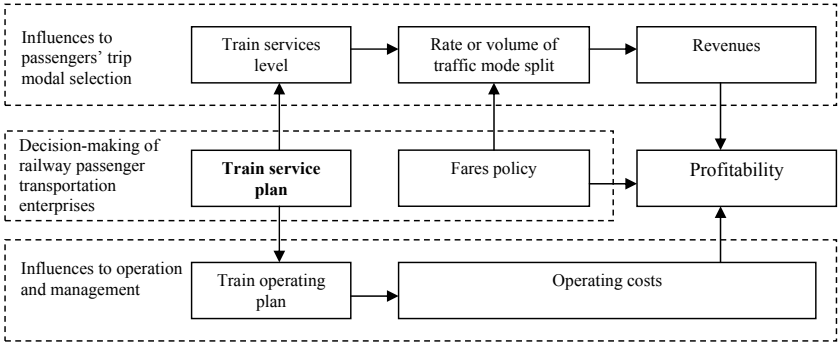


Figure 1. Passenger train services planning influences.

The paper is organized as follows: Following a brief introduction to mode share modeling in the next section, section 3 deals with the formulation of the services planning optimization model. Section 4 describes the solution algorithmand section 5 gives an example application in China. Finally the main conclusions are drawn out in section 6.

2 Modal-split modeling and service plans

As a first step in formulating the optimization model for service plans, it is necessary to establish the relationship between level of service and passenger demand. The main factors influencing mode choice decisions are shown in Figure 2. These factors are used in the form of scaling indices to define levels of service.

A standard logit modal-split model is used, namely:

$$P_i = \frac{\exp(U_i)}{\sum_k \exp(U_k)}, \qquad 0 \le P_i \le 1 \tag{1}$$

Where

P_i is the proportion of trips using mode; U_i is the utility function of mode; and k is the set of all the competing modes.

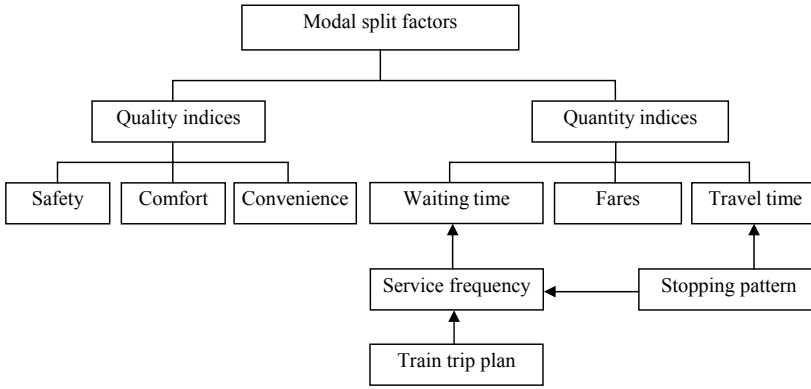


Figure 2. Factors influencing modal-split.

The utility function (U_i) can be expressed as formular (2).

$$U_i^{rs} = a_t T_{ti}^{rs} + a_w T_{wi}^{rs} + a_e T_{ei}^{rs} + a_p T_{pi}^{rs} + a_o T_{oi}^{rs}, \quad (2)$$

Where

$$T_{ti}^{rs} = T_{Ri}^{rs} + T_{Si}^{rs}, \quad (3)$$

$$T_{wi}^{rs} = \frac{t}{2F_i^{rs}}, \quad (4)$$

$$T_{pi}^{rs} = \frac{C_i^{rs}}{V^{rs}}, \quad (5)$$

Where

The travel time (T_{ti}^{rs}) for mode i from station r to station s, is mainly decided by the running time (T_{Ri}^{rs}) and the stop time (T_{Si}^{rs}), shown as equation (3);

T_{wi}^{rs} is the average waiting time of mode i in starting station (r);

F_i^{rs} is the service frequency from r to s in time period t;

Equation (4) gives the waiting time assuming a uniform distribution of arrival;

T_{ei}^{rs} is the total time that passengers spend on the section from origin to r station and from s station to final destination; (here we refer to T_{ei}^{rs} as the extension travel time) T_{pi}^{rs}, equation (5), represents the generalized fare in time units, from station r to station s; (C_i^{rs} is the fare from station r to station s. V^{rs} is the generalized time-price for passengers.)

T_{oi}^{rs} is the generalized cost in time units, which reflects the degree of safety, comfort and convenience of mode i; a_t, a_w, a_e, a_p, a_o are the calibrating coefficients.

3. Model formulation

The proposed model relates to a single line rail network without branches, using the GuangShen network in China as an example. We assume that there are m nodes or stations with each station named i in one direction and $i \in (1\mathrm{L}\ m)$. In time period t, we assume that there are n passenger train trips in the given direction with $n = 2^m - m - 1$, because the trip with only one station or zero station does not exist. Train trip j can be described with $y_j(y_{1j}, y_{2j}\mathrm{L}\ y_{ij}\mathrm{L}\ y_{mj})$, $j \in (1\mathrm{L}\ n)$. If the j train trip stops at station i, then y_{ij}=1; otherwise, y_{ij}=0. Now we let x_j be the number of the trains running on the j train trip and the set of the combination of each trip running train number is defined as an independent variable vector $X(x_1, x_2\mathrm{L}\ x_j\mathrm{L}\ x_n)$. Then the model can be established as follows:

Objective (one direction service plan):

$$\textbf{Maximize } Z^t(X) = \sum_r \sum_s C^{rs} Q^{rs} P^{rs} - (\sum_j E_o^j + \sum_f E_v^{f(f+1)}), \tag{6}$$

$1 \le r < s \le m$, $f \in (1\Lambda\ m-1)$, $j \in (1\Lambda\ n)$, $n = 2^m - m - 1$.

Sub-Function

$$T_S^{rs} = \frac{\sum_j x_j y_j^r y_j^s (\sum_{r<i<s} y_j^i \cdot t_{Sj}^i)}{\sum_j x_j y_j^r y_j^s}, \tag{7}$$

$$F^{rs} = \sum_j \beta_j \cdot x_j \cdot \mathrm{int}(\frac{y_j^r + y_j^s}{2}), \tag{8}$$

$\{if \quad \min(g \mid y_{gj} = 1, g = 1\Lambda\ m) = r \quad then \quad \beta_j = 1 \quad ; \quad else \quad 0 < \beta_j < 1\}$

$$\sum_j E_o^j = E_o \cdot \sum_j x_j, \tag{9}$$

$$E_v^{f(f+1)} = E_v^{f(f+1)}(\sum_j x_j \cdot A_j^{f(f+1)}, L^{f(f+1)}, W), \tag{10}$$

$\{if \quad \min(g \mid y_{gj} = 1, g = 1\Lambda\ m) \le f < (f+1) \le \max(g \mid y_{gj} = 1, g = 1\Lambda\ m) \quad then \quad A_j^{f(f+1)} = 1;$
$else \quad A_j^{f(f+1)} = 0\}$

Subject to

$$Q^{rs} P^{rs} \le \sum_j x_j \cdot \mathrm{int}(\frac{y_{rj} + y_{sj}}{2}) \cdot R, \tag{11}$$

$$\sum_r \sum_s Q^{rs} P^{rs} L^{rs} \le \sum_j (x_j \cdot L_j^{pq} \cdot R), \tag{12}$$

$p = \min\{g \mid \ y_{gj} = 1, g \in (1\Lambda\ m)\}$; $\quad q = \max\{g \mid \ y_{gj} = 1, g \in (1\Lambda\ m)\}$

$$\sum_{j=1} x_j \cdot y_{kj} \le M_k, \quad k \in (1\Lambda\ m), \tag{13}$$

$$\sum_{j=1} x_j \cdot A_j^{f(f+1)} \le G^{f(f+1)}. \tag{14}$$

The objective function (6) can be established as a form of profit maximization function. The fare revenue from the plan can be calculated by multiplying fares (C^{rs}) from trains from r station to station s, with the corresponding passenger traffic volume (Q^{rs}) and the rail mode share rate (P^{rs}).

Here P^{rs} is defined as in equation (2) and is a function of travel time T_t^{rs}, waiting time T_w^{rs}, extension travel time T_e^{rs}, generalized fare time T_p^{rs}, and generalized costs time T_o^{rs}. T_S^{rs} represents the average stop time for all trains traveling from r to s, shown in equation (7). t_{Sj}^i is the general stop time in stop station i and it has two components, namely: the actual stop time and the time lost during deceleration and acceleration. T_w^{rs} is a function of F^{rs}, as in equation (4) . When a passenger train trip j goes from from r to s, the train must stop at both stations. In this case, both y_{rj} and y_{sj} are 1 and the integer part of $\frac{y_{rj} + y_{sj}}{2}$ is 1; otherwise, it is 0. We can set $\sum_j x_j \cdot \text{int}(\frac{y_j^r + y_j^s}{2})$ as the train number (F^{rs}) servicing from r to s. For the passengers who will travel from r to s, the service of trains starting at r and the service of trains starting before r do not represent the same utility. Therefore, in formula (8), β_j is set as a discount coefficient in equation (8). T_{pi}^{rs} is decided by C_i^{rs} and V^{rs}, as given in eqaution (5).

The cost of the passenger train service plan includes two parts, namely, fixed cost (E_o^{rs}) and variable cost (E_v^{rs}). Fixed cost (E_o^{rs}) includes the train asset depreciation, locomotive and vehicle maintenance, passenger transportation management, and other overhead costs. It can be defined by equation (9), as the number of passenger trains ($\sum_j x_j$) and the average fixed cost of each train (E_o). E_o is given as a known input. Variable cost (E_v^{rs}) mainly varies with trip fuel consumption. In addition, the train operator needs to pay a railway-leasing fee, which is also an item of variable cost and is based on the proportion of train running distance. The variable cost of each trip (E_v^{rs}) can be seen as the function of the distance (L^{rs}) from r to s and the corresponding train weight (W). The sum of the variable cost can be expressed as equation (10). If in the trip j, the code for the starting station $\{\min(g \mid y_{gj} = 1, g = 1\Lambda\ m)\}$ is smaller than or the same as f and the code for the end station $\{\max(g \mid y_{gj} = 1, g = 1\Lambda\ m)\}$ is larger than or the same as $f+1$, then *No.j* trip must pass through $f \to (f+1)$ section and

$A_j^{rs} = 1$, otherwise $A_j^{rs} = 0$. The total number passing through f in time period t, is given by: $\sum_j x_j \cdot A_j^{f(f+1)}$.

Constraint (11) refers to train seat capacity for each origion-destination trip $(\sum_{j=1}^{n} \text{int}(\frac{y_{rj} + y_{sj}}{2}) \cdot R)$, where R is the number of the seats on each train.

Constraint (12) refers to seat-kilometers and requires the sum seat-kilometer capacity offered by train service plan can meet total person-kilometer demand in the given direction.

Constraint (13) imposes that during time t, the number of stopped trains $(\sum_{j=1} x_j \cdot y_{kj})$ on station k must be less than the train stopping capacity of station k (M_k).

Constraint (14) referes to the number of trains passing through $f \to (f+1)$ section in the given direction during time t $(\sum_{j=1} x_j \cdot A_j^{f(f+1)})$, being less than the passing capacity for this section $(G^{f(f+1)})$.

It should be stressed that the objective of the model is to maximize profits through optimizing the trade-off between passenger demand, revenues and operating costs.

4. Model solution method

The proposed model is a complex nonlinear integer-programming problem, which is NPC (NP-Complete). The complexity increases exponentially with the increased numbers of station on the line. A gradient-based search is not suitable to solve this problem, in order to avoid falling into many local optimal solutions. In addition, we hope to acquire a set of sub-optimal solutions or satisfied solutions besides the global optimal one. Thus GAs (Genetic algorithms), a heuristic algorithm can be applied to solve this model. GAs is considered suitable for application to complex search spaces not easily amenable to traditional techniques and has become a common and valid tool to solve combinatorial optimization problems, especially NP-hard problems.

Several points about GAs application need be illustrated. Firstly, we use binary digits to encode. Elements of the independent variable vector need be encoded to form a series of groups of genes and these gene groups then can be ranked and combined to form a chromosome. Secondly, when dealing with constraints, we use the penalty function method. The individuals that cannot satisfy the constraints will have penalties added to accelerate the evolution speed. Thirdly, a juxtaposition-crossover operator is introduced to accelerate the evolution speed. Fourthly, we preserve the optimal chromosome of former generations into the offspring generation to make the evolution process convergent.

4.1 Genetic encoding

Each element of the variable vector $X(x_1, x_2 \Lambda\ x_j \Lambda\ x_n)$ is x_j, which is a rational number. According to experience and the constraints, x_j has its maximum bound b and its minimal bound 0. For notational convenience, we express element x_j with gene segment having fixed binary digits, such as $\lambda_j = [\xi_{j1}, \xi_{j2} \Lambda\ \xi_{ju}]$ and $\xi_{j\cdot}$=1 or 0, then $X(x_1, x_2 \Lambda\ x_j \Lambda\ x_n)$ can be encoded as :

$$\Lambda = [\lambda_1, \Lambda\ \lambda_j \Lambda\ \lambda_n] = [\xi_{11}, \xi_{12}, \Lambda\ \xi_{1u}; \Lambda\ ; \xi_{j1}, \xi_{j2}, \Lambda\ \xi_{ju}; \Lambda\ ; \xi_{n1}, \xi_{n2}, \Lambda\ \xi_{nu}]. \tag{15}$$

Therefore Λ is the chromosome of individual $X(x_1, x_2 \Lambda\ x_j \Lambda\ x_n)$.

4.2 Population initialization

Without loss of generality, the population size we select is two times that of the chromosome's dimension number (the population size is $2nu$). In order to avoid falling into local optimal solution too early or avoid premature phenomenon, most individuals of the initial population are selected without prior knowledge and rationality of initial population need not be satisfied.

4.3 Setting record chromosome

If mutation happens with probability p_m $(0 < p_m < 1)$ and crossover happens with probability $p_c (0 \le p_c \le 1)$, then a simple GA may not always converge on the optimal solution. Here we set a special dimension to record the chromosome whose fitness is the best of all the former generations and which need not attend evolution operation. Then the solutions can converge into the optimal solution after the evolution cycle [13]. Thus the population size will be changed as $2nu + 1$.

4.4 Fitness and penalties

The objective function could be selected as the fitness function. The larger the objective value is, the better the chromosome is. Then the chromosomes with bigger fitness will be selected more often. However, this non-linear integer-programming model contains some constraints, and we must set a penalty function attached to fitness. In order to avoid premature convergence or converging on unfeasible solutions, we select an adaptable dynamic penalty function as given by Gen and Cheng (1996). The penalty function is constructed as the following.

$$\Delta b_1(X) = \max\{0, Q^{rs} P^{rs} - \sum_{j=1}^{n} x_j \cdot \text{int}(\frac{y_{rj} + y_{sj}}{2}) \cdot R\}, \tag{16}$$

$$\Delta b_2(X) = \max\{0, \sum_r \sum_s Q^{rs} P^{rs} L^{rs} - \sum_j x_j \cdot L_j^{pq} \cdot R\}, \tag{17}$$

$$\Delta b_3(X) = \max\{0, \sum_{j=1} x_j \cdot y_{kj} - M_k\}, \tag{18}$$

$$\Delta b_4(X) = \max\{0, \sum_{j=1} x_j \cdot A_j^{f(f+1)} - G^{f(f+1)}\}, \tag{19}$$

$$p(X) = 1 - \frac{1}{4}\sum_{i=1}^{4}(\frac{\Delta b_i(X)}{\Delta b_i^{\max}}), \tag{20}$$

$$\Delta b_i^{\max}(X) = \max\{\varepsilon, \Delta b_i(X)\}, \quad i = 1\Lambda\ 4. \tag{21}$$

In equation (20), $\Delta b_i(X)$ is the violation volume of constraint i and $\Delta b_i^{\max}(X)$ is the corresponding maximum violation volume of the present population. ε is very small and positive to avoid the denominator becoming zero. The fitness evaluation function can be written as:

$$eval(X) = Z^t(X)p(X). \tag{22}$$

4.5 Crossover and mutation

Crossover and mutation are the key operators in GAs. They will influence the evolution speed. Here the chromosome contains n gene segments, each of which represents a kind of train trip. In order to avoid destroying the evolution information of gene segments, we adopt the juxtaposition-crossover method. Except the record chromosome, the population size is $2nu$. We assume that $\varphi = 2nu$ and $\forall h \in (1\Lambda\ \varphi)$.

Above GA arithmetic is programmed based on the Genetic Algorithm Optimization Toolbox in MatLab 6.0 version.

5. Example Application

5.1 Background

Guang Shen QHSR (GS-QHSR) is the first quasi-high-speed railway of China. It is located in Guangdong Province and starts from GuangZhou (Capital of Guangdong Province) to ShenZhen, where GS-QHSR is joined with Kowloon-Canton Railway (KCR) of Hong Kong. The daily average passenger traffic volume starting and ending on GS-QHSR itself in 2001 was 43000 passengers per day, of which 26000 passengers used quasi-high-speed trains, (average speeds of 160 km/h). There are sixty-two pairs of quasi-high-speed trains running on GS-QHSR every day with average headway of 15 minutes.

In the transportation market between GuangZhou and ShenZhen, two highways parallel the GS-QHSR, namely, the GuangShen Freeway and 107 National Highway,

as shown in Figure 3. Passenger traffic in the GuangZhou region surpassed 100 million passengers in 2000, of which 80 percent use road based transport modes.

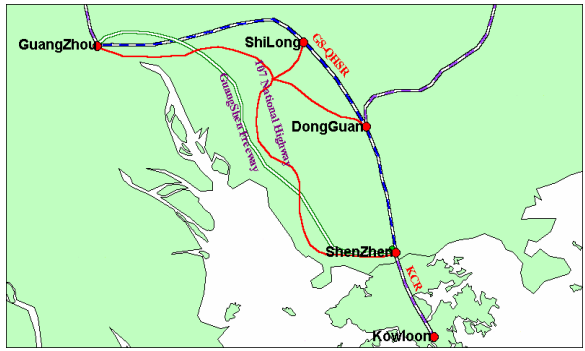


Figure 3. The map of GS-QHSR and its competition environment

Passenger traffic is still increasing at around 15 percent p.a. As a result, both railway and road transport are actively competing in this market.

5.2 Preliminary data analysis

On the GS-QHSR line, local passenger traffic accounts for more than 65 percent of the total passenger traffic. Among the 101 pairs of trains serving local passenger traffic, 62 pairs are express QHSR trains, which mainly serve the four large stations (GuangZhou, ShiLong, DongGuan and ShenZhen) and make up almost 90 percent of inter-city railway passenger traffic. The main competition for inter-city traffic is between rail and road based passenger vehicle transport.

We randomly selected passenger traffic volumes for one two-hour period in the morning. The passenger traffic volumes are shown in Table 1. Tables 2, 3 and 4 show the data used to define the characteristics of express train services. Table 5 lists the parameters associated with the road based alternative mode.

Table 1. The passenger traffic volumes between OD nodes from ShenZhen to GuangZhou. (In two hours)

	ShenZhen	DongGuan	ShiLong	GuangZhou
ShenZhen	0	3280	1839	5410
DongGuan		0	1625	3059
ShiLong			0	2074
GuangZhou				0

Unit: passengers

Table 2. Express train parameters (1).

	ShenZhen	DongGuan	ShiLong	GuangZhou
ShenZhen	0	57, 29, 30,30,24	78, 38, 20,50,24	147, 60,15,70,24
DongGuan		0	21, 14, 30,12,24	90, 35,25,45,24
ShiLong			0	69, 27,18,30,24
GuangZhou				0

From left to right: mileage [km], running time [min.], extension travel time [min.], ticket price [RMB Yuan] and section-servicing capacity [train]

Table 3. Express train parameters (2).

	ShenZhen	DongGuan	ShiLong	GuangZhou
Station-servicing capacity (train)	22	20	20	24
Stop-time (min.)	-	5	5	-

Table 4. Express train parameters (3).

Train weight	Carrying capability	Variable Cost		Fixed Cost	Tax
		Engine cost	Line-leasing cost		
(ton)	(Person/one train)	(RMB Yuan/104.ton-km)	(RMB Yuan/km)	(RMB Yuan/train-trip)	(%)
670	940	600	100	1382	18

Table 5. Commuter road vehicles parameters.

	ShenZhen	DongGuan	ShiLong	GuangZhou
ShenZhen	0	15,5,45,25	15,20,65,40	15,5,110,60
DongGuan		0	15,10,20,10	15,10,60,35
ShiLong			0	15,15,75,40
GuangZhou				0

From left to right: average extension time [min.], waiting time [min.], running time [min.] and ticket price [RMB Yuan].

In addition to the above data, we also need to input T_{oi}, which relates to service quality and reflects the degree of safety comfort, and convenience of traffic mode i.Here convenience mainly evaluates the procedures of buying tickets, boarding and alighting from trains, consigning and picking-up luggage. We use subscript 1 to symbolize commuter road vehicles and subscript 2 to symbolize express trains. According to equations (1) and (2), the express rail modal-split P_2 can be rewritten as equation (23), which can be seen the function (H) of $T_{o2} - T_{o1}$ on the assumption that other parameters are known. In the short term, we can assume that $T_{o2} - T_{o1}$ is constant so that the modal-split could be used to calibrate $T_{o2} - T_{o1}$ in the reverse direction.

$$P_2 = \frac{1}{1+\exp(U_2 - U_1)} = H(T_{o2} - T_{o1}) \ . \tag{23}$$

Based on historical data for the past year, $T_{o1} - T_{o2} = 0.0661$hrs., which means the difference between the two modes' qualitative servicing level is very small and the express train is appreciably better than commuter road vehicles. In addition, because all the utility elements are converted into generalized time, the values of the coefficients, a_t, a_w, a_p and a_o are all 1. When converting money into time, we still need use the time-price coefficient. Table 6 lists the statistical result.

Table 6. Values of time used between OD pairs.

	ShenZhen	DongGuan	ShiLong	GuangZhou
ShenZhen	0	13.56	13.51	14.83
DongGuan	-	0	12.69	13.51
ShiLong	-	-	0	13.45
GuangZhou	-	-	-	0

Unit: RMB Yuan/hour

5.3 Model process

The parameters of the GA are set as Table 7, and the evolution process is shown in Figure 4. The latter records the best fitness of every generation (upper curve) and the average fitness of synchronous generation (lower curve). With an increase of evolution generations, the curve gradually approaches a global optimum.

Table 7 The parameters of GAs.

Binary digits of chromosome	Population Size (except record individual)	Crossover probability	Mutation Probability	Maximal generation size
44	88	0.85	0.1	200

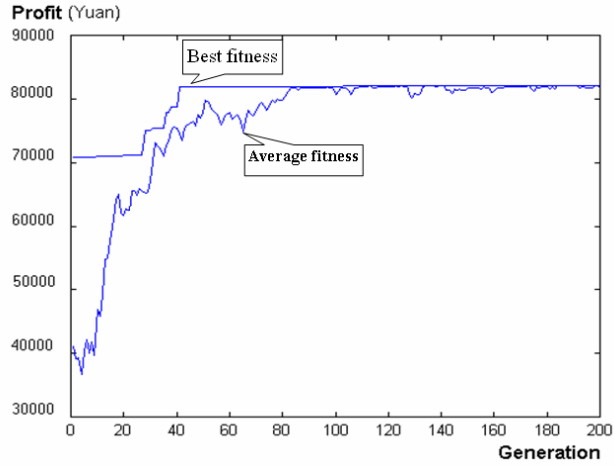


Figure 4. The evolution curve.

Now the global optimal solution x_o and a local optimal solution x_l, are given for comparison in Figure 5.

One vector x, symbolizing an entire train service plan, contains a combination of the number of each possible train trip y. The benefits comparison of the two train service plans is listed in Table 8. From the benefits comparison, the revenue generated by global optimal plan x_o is close to and even less than that by local optimal plan x_l. However, plan x_o successfully reduces the gross number of running trains from fourteen to ten. The evolution trace of GAs from local solution x_l to global solution x_o can also be seen here. The train trips y_3 and y_4 in plan x_l can be seen being combined into y_7 in plan x_o (See the sign "○" in Figure 5); and the train trips y_9 and y_{10} of plan x_l can be seen partly being combined into y_{11} in plan x_o (See the sign "!" in Figure 5). The process of the global optimal solution being evolved from local optimal solution accords with the evolution theory. Certainly, the external evolution process comes from complex calculation and precise comparison for continuously searching the trade-off between revenue and costs.

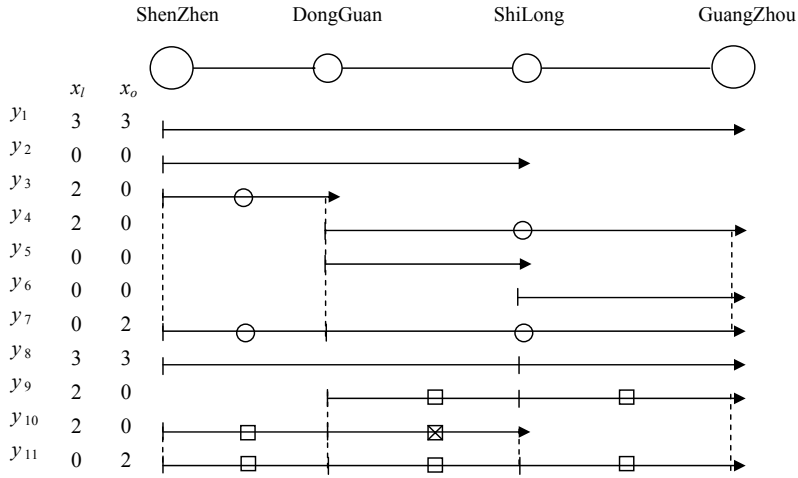


Figure 5. Comparison between a local optimal solution and the global optimal solution

Table 8. Benefits comparison between the two train service plans.

Service plans	Revenue	Tax	Cost			Profit
			Fixed Cost	Engine Cost	Leasing-line Cost	
x_o	390590	70306	13280	62712	156000	88292
x_l	390852	70353	18592	64401	160200	77306

Unit: RMB Yuan

Table 9. Modal-split for the global optimal solution

OD trip	Trains Number	Rail passengers (proportion)	Rail travel times (hrs.)	Road travel times (hrs.)
ShenZhen-DongGuan	4	0.425	3.295	2.993
DongGuan-ShiLong	2	0.332	2.303	1.605
ShiLong-GuangZhou	5	0.826	3.230	4.790
ShenZhen-ShiLong	5	0.436	4.951	4.693
DongGuan-GuangZhou	4	0.342	4.727	4.073
ShenZhen-GuangZhou	10	0.538	6.128	6.280

5.4 Solution analysis

As shown in Table 9, the global optimal solution gives maximum rail mode share on ShiLong to GuangZhou OD pair (0.826 to rail). The lowest rail share is on the

DongGuan to ShiLong OD pair, (0.332 to rail) The results show that the best strategies are those which provide for a rail/road mix rather than those which favor rail demand without regard to operating cost implications.

Conclusions

A GA was found to be an effective technique to solve the stated transportation optimization problem. The procedure adopted here allows for practical considerations to be taken into account by providing a range of sub-optimal solutions in addition to the global optimal solution. In this way, practical train scheduling and train operating plan aspects can be brought into play. The application of the proposed model to a real-world example of a transport corridor in China has shown that the GA solution provides realistic results, when compared with expected outcomes.

Acknowledgements

The research was supported by National Natural Science Foundation of China (Project No. 70601022) and the Program for Young Excellent Talents in Tongji University (Project No. 2006KJ023).

References

Assad, A.A. (1980). "Modelling of rail networks: toward a route/makeup model." *Transportation Research Part B*, 14(A), 101-114.

Assad, A.A. (1980). "Models for rail transportation." *Transportation Research Part A*, 14(B), 205-220.

Darryn, J.R.(1997). "Enchanced genetic operators for the resolution of discrete constrained optimization problems." *Computers & Operations Research*, 24(5), 399-411.

Kraay, D.R., and Harker, P. T. (1995). "Real-time scheduling of freight railroads." *Transportation Research Part B*, 29(3), 213-229.

Gen, M., and Cheng, R.(1996). "A survey of penalty techniques in genetic algorithms." *Proc. of the Third IEEE Conference on Evolutionary Computation*, IEEE Press, Nagoya, Japan, 804-809.

Holland, J.H. (1975). *Adaptation in Natural and Artificial Systems*, MIT Press, Cambridge.

Gorman, M.F. (1998). "Santa Fe Railway uses an operating-plan model to improve its service design." *Transportation science*, 28(4), 1-12.

Michael, R., Bussieck, P.K., and Uwe, T.Z.(1996). "Optimal lines for railway systems." *European Journal of Operational Research*, 96, 54-63.

Robert, B., Brent, H. and Paul, C. (1995). "A multi-objective optimization approach to urban school bus routing: Formulation and solution method" *Transportation Research Part A*, 29(A), 107-123.

Sang, N., Stefano, P. and Federico, M. (2001). "A modeling framework for passenger assignment on a transport network with timetables." *Transportation Science*, 35(3) ,238-249.

Brännlund, U., lindberg, P.O., Nou, A. and Nilsson, J.E. (1998). "Railway timetabling using Lagrangian relaxation." *Transportation science*, 32(4), 358-369.

Chang, Y.H., Yeh, C.H. and Shen, C.C. (2000). "A multi-objective model for passenger train services planning: application to Taiwan's high-speed rail line." *Transportation Research Part B*, 34(2), 91-106.

Wenxun, X., and Jinxing, X. (1999). *Modern optimization algorithms*, Tsinghua Press, Beijing.

Zbigniew, M. and Naguib, F.A.(1994). "Evolutionary optimization of constrained problems." *Proc. of the Third Annual Conference on Evolutionary Programming*, World Scientific Publishing, River Edge, NJ, 98-108.

Two Proposals for Solution to Transport Problems in Wuhan

Qi Cheng, Akio Kondo, Akiko Kondo, Hideaki Shimada

Ecosystem Engineering Graduate School of Engineering, The University of Tokushima
2-1, Minamijosanjima, Tokushima, ST 770-8506, JAPAN; PH(0081)88-656-7339;
FAX(0081)88-656-7341; email:{cheng, kondo, aco, hideaki}@eco.tokushima-u.ac.jp

Abstract: As the biggest water, land and air transport hub in inland China, Wuhan plays an important role in connecting east and west. In fact, Wuhan is situated as a geo-economic center in China. However, transport problems obstruct the development of Wuhan seriously. In this paper, introductions of water bus and staggered commuting hours which have been studied and implemented in Japan are proposed to cope with traffic problems in Wuhan. By the analysis of these two policies, considering geography and traffic in Wuhan, whether these policies can resolve traffic problems in Wuhan is discussed.
Keywords: *Wuhan, traffic, water bus, staggered commuting hours*

1 Introduction

Traffic jams in urban areas are serious problems which many countries face and struggle to resolve them. In Japan, various policies have been applied to cope with the traffic jams. Some of them such as park and ride, staggered commuting hours and collective pickup and deliver service contribute to decrease traffic jams and reduce exhaust gas from vehicles. These experiences are worth enough to refer when we study transport problems in Wuhan.

Wuhan which is the metropolitan in the middle of China, faces with the imbalance between development of transport facilities and demand of urban traffic. Many researchers have paid attention to urban traffic in Wuhan. In consideration of the situation of Wuhan and experiences of Japan, water bus and staggered commuting hours are thought to be effective policies to overcome transport problems in Wuhan.

2 Water Bus

Water bus is a ship which carries passengers on the river or the sea. In Japan, the water bus is seldom used for the public transportation. However, the researchers in Japan start to pay attention to water bus as one of public transportations in order to cope with traffic jams and serious environment problems which are brought by heavy traffic. Especially, there is a previous study in which the necessity and feasibility of introduction of the using water bus are analyzed in Tokushima city in Japan.

2.1 Water Bus in Japan

Tokushima city, the capital of Tokushima prefecture (equal to the province of

China) is one of middle size cities in Japan. Amount of vehicles in Tokushima is quite a few. It is reported in 2005 that the population of Tokushima is 267,906 and amount of vehicles is 181,635. During rush hours, traffic jams are very serious around the public facilities and business districts. Heavy traffic jams are occurred in fringes of urbanized areas as well. As a characteristics of urban area in Tokushima city, many rivers flow in the area and form water networks. Moreover, public facilities and business offices are mostly concentrated along the rivers. Around many bridges over the rivers, serious traffic jams are seen everyday. Therefore, it can be said that not only the density of river networks, but also traffic bottlenecks around bridges are characteristics in Tokushima city. In this situation, Tokushima adopts many policies to alleviate this problem. Among these policies, the water bus is one of the most effective projects for Tokushima.

A plan of introduction of river bus in Tokushima city which is proposed in an academic paper by Kondo etc. is briefly explained below. In the paper, the effects of introduction of the river bus are also analyzed. Concretely, in order to improve a convenience for passengers, 24 sites close to schools and public facilities are chosen as water bus stations. The river bus routes between these sites are proposed and it takes 5-15 minutes in each route. A single fare is 200 yen and commutation ticket is 7,800 yen in a mouth. As a result of the study, it is shown that the effects for reducing traffic jams and air pollution by implementing the water bus. Figure 1 shows a total travel time by cars at each time slice. The total travel time has decreased after implementing the water bus. Especially, reduction rate of 9:00a.m. is most effective. Thus, the introduction of water bus can improve the traffic condition remarkably. Figure 2 shows the change in exhaust volume of CO_2 due to the water bus. It has also decreased greatly, the reduction of exhaust volume is about 10,350 kg-c/day and reduction volume of CO_2 is about 11.4 $\times 10^6$kg-co_2, and reduction rate of total CO_2 volume is 7% per day.

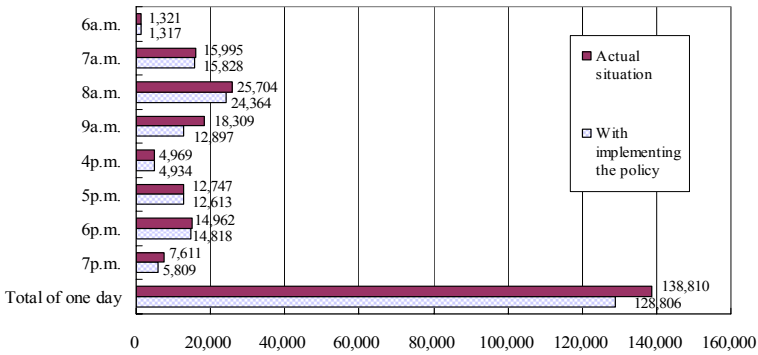


Figure 1. Total travel time (hours)[1]

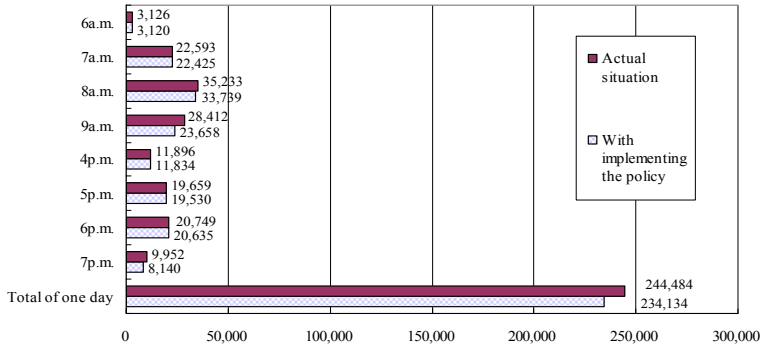


Figure 2. Exhaust volume of CO_2 (kg-c)[1]

2.2 Urban Traffic Related to Water Bus in Wuhan

Although there are many differences between Wuhan and Tokushma in dimensions of city, two cities have similar geographical environment and river networks. In addition, the density of river networks is the cause of traffic bottlenecks of road traffic in Wuhan. Therefore, an implementation of water bus project in Wuhan can be said to be available policy, and also expected to bring prominent effects.

Wuhan is a central city in the middle of China with an area of 8,494 km^2. There are many rivers, hundreds of hills and nearly 200 lakes of various sizes in Wuhan. Water area accounts for 25.8% of its entire territory, and Wuhan is ranked as the first in China for water resources. The Yangtze River, third longest river in the world and its largest tributary Hanjiang River meet in Wuhan cut the city into three parts of Hankou, Hanyang and Wuchang. Traffic crossing over the rivers is the most serious problem of transportation in Wuhan. There are three bridges across the Yangtze River between Wuhan city to other cities now. Total traffic volume of these three bridges is about 230,000 (vehicles per day) and total amount of trips in Wuhan city is 650,000 (trips per day). Wuhan Changjiang Bridge having 4 lanes was built in 1956. Average of traffic volume of this bridge is about 94,000 (vehicles per day) and the volume during rush hours is about 8,400 (vehicles per hour). Wuhan Changjiang Bridge No.2 having 6 lanes was built in 1995. Average of traffic volume of this bridge is about 110,000 (vehicles per day) and the volume during rush hours is 9,000 (vehicles per hour). Wuhan Changjiang Bridge No.3 having 6 lanes was built in 2000. Average of traffic volume of this bridge is about 30,000 (vehicles per day).

As Wuhan Changjiang Bridge No.3 is on the edges of the main city, traffic volume is fewer than the other two bridges. However, Wuhan Changjiang Bridge and Wuhan Changjiang Bridge No.2 are always saturated. Moreover, there are 5 bridges, among them, Jianghan Bridge which has a big traffic volume (about 98,000 vehicles per day) is on the key location of inner loop road. Total traffic volume of the other

[1] Kondo, A., Hirose,Y., Hitoshi, K., Kishimoto, H. (1999). "Effect of Introduction of Transport Policies Reflecting Characteristics of Tokushima City." *Papers on City Planning.,* No.34, 913-918.

four bridges is 200,000 (vehicles per day). With the increase in cars crossing rivers, the traffic pressure of roads connected with bridges is becoming heavier and heavier. At present, intersections where traffic volume is over 10,000 (vehicles per hour) are almost connected with bridges. Meanwhile, since 2000, average of increase rate of vehicles ownership in Wuhan is 17%, amount of vehicles at the end of 2005 is 653,000 and it still keeps on a growth from now on. The problem of traffic crossing river affects the development of Wuhan.

In order to alleviate traffic pressure, Wuhan funded for and adopted various policies such as renovations of Changjiang Bridge, Wuhan Changjiang Bridge No.2 and Jianghan bridge, and construction of Yuemachang underground tunnel. Although these policies can partly alleviate traffic pressure, they don't utilize the characteristics in Wuhan. That is the river networks which will be used as the advantage. When we consider the present traffic condition, it can be said that Wuhan needs another operable policy which can be completed in a short time. Therefore, a water bus is a good method.

Although Wuhan still maintains the ferry crosses the river, it is not convenient. Because the travel course by ferry is limited and its capacity is only 0.7% of total transportation crossing the river.

New development service, namely speed boat is introduced only for tour and can't alleviate traffic problem. Natural and little polluted water resources are not applied well. The water bus needs less investment and has a short project time, therefore it can utilize the advantage of Wuhan well.

2.3 Implementation of Water Bus in Wuhan

We plan to implement the project of water bus both in the Yangtze River and the Hanjiang River. In the project 10 sites are selected as ship stations from Yangtze River to Hanjiang River. Existing ship stations can be used in 9 sites among them, and only one is newly-bulit. Speed boat is introduced in the plan, in order to improve a convenience of passengers who travel from places to places in a short time. Passengers are allowed to take the water bus with their bicycles. Travel time between ship stations is about 5-10 minutes. The interval of departure in each station in the morning and evening rush hours which are 7:00-8：30a.m. and 5:00-6:30p.m. is 10 minutes, and that in another time is 15 or 20 minutes. In consideration of consumption ability, single fare is 1 yuan a person, bicycle is 0.5 yuan (per vehicle), commutation ticket is 30 yuan. Figure 3 shows the map of water bus.

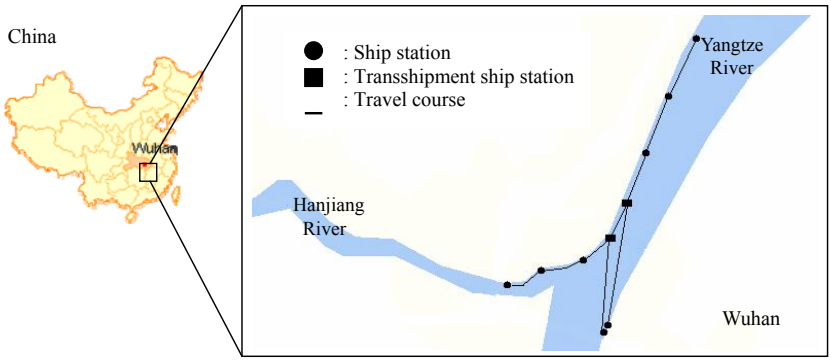


Figure 3. Map of water bus stations in Wuhan

Water bus is much more convenient than the ferry and doesn't cause traffic jams. Therefore, it will be easily accepted by the people who are familiar with water in Wuhan. We can make sure that, after the water bus is widely used in Wuhan, it can not only alleviate the pressure of river-crossing traffic but also reduce the vehicle utilization and exhaust gas. As a result, the living environment in Wuhan is improved considerably.

3 Staggered Commuting Hours

Staggered commuting hours means that the beginning of office hours is staggered to another time to avoid rush hours in commuting. Traffic which is concentrated in the morning and evening rush hours disperses to another time by staggering commuting hours. Therefore, the alleviation of traffic jam is promoted.

3.1 Staggered Commuting Hours in Japan

In many big and medium-sized cities in Japan, the traffic jam is very heavy during rush hours. Therefore, the staggered commuting hours has been introduced very early in Japan. The staggered commuting hours has been implemented in Tokyo in 1961. The introduction of this measure was followed by Osaka in 1962, Nagoya in 1974, Fukuoka in 1976 and Sendai in 1983. The detailed project is as follows:

(1) For the governmental department, staggered commuting hours is implemented forcibly;
(2) School, local government and office are encouraged to implement staggered commuting hours;
(3) Japanese government also draws up detailed schemes and goal for the above-mentioned cities.

It is reported in 1999 that the number of people participating in staggered commuting hours is 1,633,000 in Tokyo, 707,000 in Osaka, 119,000 in Nagoya, 68,000 in Fukuoka and 43,000 in Sendai. The other cities in Japan are also implementing staggered commuting hours with different scale.

In the long implementary course, the policy brought great results. The congestion rate[2] of above-mentioned 5 cities has been continuously decreased. Table 1 shows the detailed results.

Table 1. Congestion rate[3]

City	Year		
	1989	1994	1999
Tokyo	202%	194%	180%
Osaka	171%	*166%	144%
Nagoya	175%	163%	153%
Fukuoka	160%	146%	133%
Sendai	155%	151%	137%

* Data investigated in 1993

To take Sendai city as an example, the length of traffic jam from north to south is about 1,130 (m). In Fustuka city, the traffic jam is very heavy around every intersection. If the staggered commuting hours is not implemented in Sendai, the length of traffic jam is estimated to be about 1,800 (m). The congestion rate of subway between North-Sendai Station and North-Yonbancho in North-South Line is 180% where the road traffic jam is the heaviest in Sendai city. If the staggered commuting hours is not implemented, the congestion rate is estimated to be 220%.

In addition, from the questionnaire survey about merits of staggered commuting hours taken in Sendai, it is clarified that this policy brings many benefits which are the reduction of fatigue for commuting, improvement in the ability to work, decrease of tardy for work, cost reduction and decrease of traffic accident during the commuting time. Figure 4 shows the survey result.

[2] Congestion rate; 220%: people have no space to move easily
Congestion rate; 180%: people can read newspaper with bodily contact
Congestion rate; 150%: people can read newspaper easily only touching shoulder each other
Congestion rate; 100%: just reach a passenger capacity

[3] Directorate General for Policies on Cohesive Society, Cabinet Office, Government of Japan. (2001). "Promotion Planning of Staggered Commuting Hours." <http://www8.cao.go.jp/koutu/juten/tepu1.htm> (Jan, 2007).

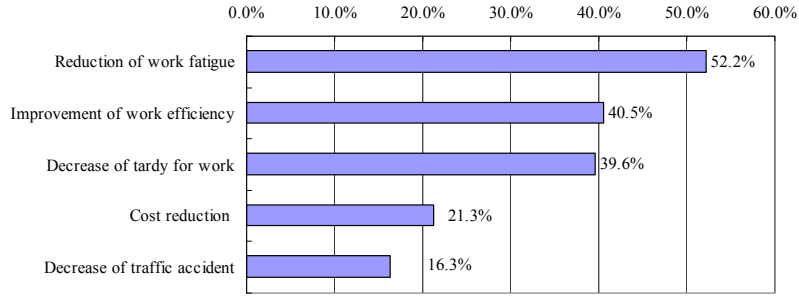


Figure 4. Survey in Sendai city (2003)[4]

3.2 Urban Traffic Related to Staggered Commuting Hours in Wuhan

With a population of 8.58 million, Wuhan is the biggest metropolis in hinterland China. There were 1,697,836 workers in Wuhan in 2005, and the number of workers in Chinese National Corporations is 1,043,587. There were 34,000 public servants in 2004. Wuhan is also known as "Education City" with great numbers of universities. There were 1,913,445 students in Wuhan in 2005. Therefore, Wuhan has been facing very serious concerns of heavy commuting traffic.

At the end of 2005, there were 55 intersections where the traffic volume is over 5,000 (vehicles per hour), among them, 27 intersections where the traffic volume is 5,000-7,000, 23 intersections where the traffic volume is 7,000-10,000, 5 intersections where the traffic volume is over 10,000. At the same time, there were also about 10 intersections where the traffic volume is over 5,000 (vehicles per hour) on the outskirts of the main city. It shows that the traffic jam is proliferating to the outskirts of the city. On the contrary, the traffic volume tends to decrease in the exception of rush hours.

3.3 Implementing Staggered commuting hours in Wuhan

As it is considered that introduction of staggered commuting hours into private enterprises is difficult, we propose to implement this policy only in Chinese National Corporations. We consider that Chinese National Corporations shift their commuting time. For example, they can choose 9:00am to 6:00pm or 9:30am to 6:30pm as their office hours. To some especial Chinese National Corporations, the staggered commuting hours will not be implemented forcibly. After being embraced this policy of staggered commuting hours by citizen in Wuhan, the policy will be implemented to the students. In Japan, staggered commuting hours is implemented to citizen and students at once.

It is clear that after the project is implemented successfully the congestion rate will be continuously decreased as well as commuting time will be shortened.

[4] Sendai City Official. (2003). "Survey about staggered commuting hours (Sendai)." <http://www.city.sendai.jp/toshi/koutsukikaku/jisa/jisa2.html> (Jan, 2007).

Moreover, it can be expected that the study efficiency of students and work efficiency are improved. As the commuting time of many citizens is changed along with the staggered commuting hours of public servants, it is more convenient for citizens to go to office. Staggered commuting hours can also reduce travel time and exhaust gas from vehicles, and contribute to improve the living environment in Wuhan. Certainly, there are also some disadvantages in implementing the project, for example, it is difficult for different companies or departments to shift and harmonize their time. However it can be confirmed as long as the project is implemented expediently. We believed that staggered commuting hours is accepted by citizen, and bring great benefits to Wuhan's development.

4 Conclusion

In this paper, two policies of the water bus and staggered commuting hours are proposed in order to resolve traffic problems in Wuhan. The policy of water bus has been investigated in Tokushima in Japan. The water environment in Tokushima is similar to Wuhan, therefore it can be said that the water bus is good policy to introduce in Wuhan. Staggered commuting hours have been implemented in Japan since 1961. It is reported that this policy contributes to reduce traffic jams in many cities in Japan. In conclusion, these two policies are good projects for Wuhan. In addition, they are also suitable to the other cities with the similar environment to Wuhan.

References

Kondo, A., Hirose,Y., Hitoshi, K., Kishimoto, H. (1999). "Effect of Introduction of Transport Policies Reflecting Characteristics of Tokushima City." *Papers on City Planning.,* No.34, 913-918.

Wuhan City Comprehensive Transport Planning & Design Institute. (2006). "Report of developent of Urban Traffic in Wuhan".

Sun, X. L. (2006). "The Application of TDM to Traffic Crossing the River in Wuhan." *Urban Transport of China*., Vo14, No.1, 48-51.

Directorate General for Policies on Cohesive Society, Cabinet Office, Government of Japan. (2001). "Promotion Planning of Staggered Commuting Hours." <http://www8.cao.go.jp/koutu/juten/tepu1.htm> (Jan, 2007).

Sendai City Official. (2003). "Survey about staggered commuting hours (Sendai)." <http://www.city.sendai.jp/toshi/koutsukikaku/jisa/jisa2.html> (Jan, 2007).

Ministry of Land,Infrastructure and Transport Japan, <http://www.mlit.go.jp/> (Jan, 2007).

Statistical Information of Wuhan, <http://www.whtj.gov.cn/> (Jan, 2007).

OPTIMAL COORDINATION OF ARTERY SYSTEM BASED ON MODIFIED CELL TRANSMISSION MODEL AND PARTICLE SWARM ALGORITHM

Zhengwu WANG[1, 2], Dayong LUO[1] and Zhongxiang HUANG[2]

[1]School of Information Science and Engineering, Central South University, 410075, ChangSha City; E-mail: zhengwu.wang@126.com
[2]School of Traffic and Transportation Engineering, Changsha University of Science & Technology, 410076, ChangSha City; E-mail: mehzx@126.com

Abstract: In order to relieve traffic congestion, the traffic signals of an artery system should be coordinately controlled. Therefore, firstly, modifying the cell transmission model and considering the delay time of the platoon dispersion of the exit lane, traffic flow models on the ordinary, merging and diverging links are developed respectively. Secondly, if common the cycle of artery system cannot be optimized, it is ineffective to optimize other parameters. So based on the relationship the of signal cycle, offsets, splits and the connection of offsets in each direction, a new optimal control system with bi-level hierarchical structure is presented. At the first level, the cycle and two-way offsets are determined simultaneously in accordance with traffic flows on the main road and secondary roads, which can cause delays of the artery system.. At the second level, a split of each intersection is determined to cause less delays. To solve the new model, a modified particle swarm algorithm is advised by the changeable inertia factor at a different stage. Finally, the new model is used for the artery system and good results are obtained.

Key words: *traffic engineering, artery system, common cycle, cell transmission model, particle swarm algorithm*

1 Introduction

The artery coordination control is a coordinated control measure to make vehicles encounter green lights as frequently as possible by choosing proper signal cycles, splits and offsets of artery system, which involves two kinds of control methods: timing control and responsive control. In timing control, green time is predefined according to the statistic data, and in responsive control, it is determined by the real-time traffic flow. Although the device of timing control is usually simpler, its control effect is no better than that of responsive control.

Many efforts have been made on the artery control system and some optimization algorithms have been put forward, such as maximal through-band method (Yang, 1995), SIGOP offsets optimization algorithm (Zhu, 2003), bi-level hierarchical structure optimization algorithm (Chen, et al, 1998), proportion signal offsets design method (Wei, et al, 2001), two-way offsets optimization regulation algorithms (Wan, et al, 2001), optimization offsets design methods based on genetic algorithms (Chang,

et al, 2003), variable through-band methods (Gartner, et al, 2002) and integral linear programming methods (Wey, 2000). Among them, maximal through-band method is the most commonly used method, where signal parameters of each subsystem are separately obtained, and then these parameters are coordinately designed in order to make green wave bandwidth the widest.

Current efforts on the artery control system, however, primarily focus on the optimal design of offsets and splits, while the optimal choice of a common cycle is usually neglected. A common cycle is usually chosen as either the maximal cycle (e.g. maximal through-band method) or the one appearing at most times. Such a choice can't ensure that the optimal cycle is obtained. Therefore, in this paper, based on cell transmission model (CTM) and delay time of the platoon dispersion on exit lane, traffic flow models of merging and diverging link in artery system are built. Then according to the relationship of the signal cycle, offsets, splits and the connection of offsets in each direction, the bi-level hierarchical model is built to coordinate and optimize the artery control system.

2 Network traffic flow models

In CTM, first sponsored by Daganzo (Daganzo, 1995), a freeway was regarded as a system constituted by many cells of the same length whose state was described by occupancy ratio. And the model was confirmed to be effective to simulate LWR macroscopic traffic flow. To make CTM adapted to the signal intersection, the following modification and hypothesis are considered:

- Links are composed of cells of different length, which can not only divide cells in terms of link geometry characteristics, but reduce the number of cells and consequently decrease state dimensions.
- The state of different length cell is described by density.
- If vehicles pass the parking line, they are thought to have passed the intersection.
- Take no account of the travel time and delays on exit lane vehicle flow.
- The dispersion of the exit platoon is described by the Robertson model, and it is thought that all delays will occur in the first cell behind the parking line.
- The first cell in front of the parking line should be long enough to accommodate all vehicles which are stopped by red lights.

The artery system composed of two intersections is shown as Figure 1. According to Daganzo's proposition, the maximum number of cells which a cell can connect should not be more than 3. So the cells of the artery system can be divided as shown in figure2.

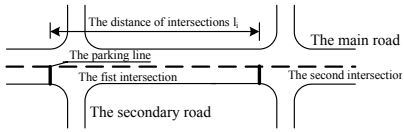


Figure 1. An artery system

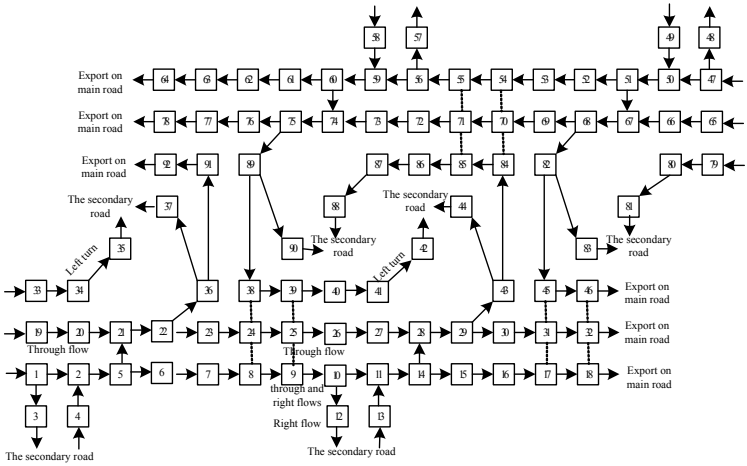


Figure 2. The cell partition of artery system

There are 92 cells totally. The merging cells are 2, 21, 11, 28, 50, 67, 59 and 74, and the diverging cells are 1, 5, 22, 36, 10, 14, 29, 38, 47, 51, 68, 82, 59, 60, 75 and 89. The left turn vehicle flows onto the main road are 33-34-35, 40-41-42, 79-80-81 and 86-87-88 when the right turn ones are 1-3, 9-10-12, 47-48 or 55-56-57. The right turn vehicle flows on secondary road are 4-2-5, 13-11-14, 49-50-51 and 58-59-60 while the through and left ones are 4-2-5-21-22-36, 13-11-14-28-29-48, 49-50-51-67-69-82 and 58-59-60-74-75 -89. Among them, 20 cells, such as cell 2, 5, 21, 22, 36, 11, 14, 28, 29, 48, 50, 51, 67, 68, 82, 59, 60, 74, 75 and 89, are fictitiously set up for the diverging movement. Assume that the fictitious cells can accommodate enough vehicles whose delay and travel time can be neglected. At the same time, for cell i, define k_i as the density, L_i as the length, C_i as the capacity, $k_{i,jam}$ as the jam density, $Q_i(w)$ as the maximum number of vehicles that can flow into it and $q_i(w)$ as the number of vehicles flowing into it within time $[w\tau, (w+1)\tau]$, where τ is time step length and w is integer. Define $\delta=V_{free}/\upsilon$, where υ is the spread speed of congestion wave and V_{free} is free flow velocity. In order to ensure the validity of the model, L_i and τ are subjected to $\tau \le L_i / V_{free}$.

2.1 Traffic flow model on ordinary link

According to the volume conservation theorem, within time $[w\tau, (w+1)\tau]$, the relationship between cell i and the closest downstream cell j is:

$$k_i(w+1)=k_i(w)+\frac{q_i(w)-q_j(w)}{L_i} \tag{1}$$

$$q_j(w)=\min\{L_i\cdot k_i(w), Q_j(w), \delta\cdot L_j\cdot[k_{j,jam}(w)-k_j(w)]\} \tag{2}$$

$$Q_j(w)=C_j\cdot\tau \tag{3}$$

2.2 Traffic flow model on merging link

The merging movement between the through vehicle on the main road and the left or right one on the secondary road has been controlled by the signal and separated from time, it can also be described by the traffic flow model on the ordinary link. However, the dispersion of the exit platoon must be considered. So within the time $[w\tau, (w+1)\tau]$, the relationship between the parking line cell i and the closest downstream cell j is expressed in the following way:

$$k_i(w+1)=k_i(w)+\frac{q_i(w)-q_j(w)}{L_i} \tag{4}$$

$$q_j(w)=\min\{L_i\cdot k_i(w), Q_j(w), \delta\cdot L_j\cdot[k_{j,jam}(w)-k_j(w)]\} \tag{5}$$

$$q_j(w)=\sum_{i=1}^{w-\zeta}F(1-F)^{i-1}q_j(w-i) \tag{6}$$

$$Q_j(w)=\begin{cases}0 & if\ \ w\ \ is\ \ red\ \ light\\ C_j\cdot\beta & if\ \ \beta\ \ is\ \ green\ \ light\ \ in\ \ \tau\end{cases} \tag{7}$$

where $q_j'(w)$ is the vehicle number of the downstream cell j which arrives within the time $[w\tau, (w+1)\tau]$ when the dispersion of the exit platoon is not considered. According to the Robertson model, $F=1/(1+\alpha\zeta)$, $\zeta=0.8\times tt/\tau$, tt is the average travel time on ordinary link, $\alpha\in[0.2,0.5]$ determined by the platoon length. Ψ is infinite, but generally assuming $\Psi\in[10,15]$.

2.3 Traffic flow model on diverging link

As the exclusive right turn lane and phase are not set up, the through and right vehicle flow on the main road and the secondary road will be diverged in front of the parking line. Define β_l, β_m as the distributing rates of the closest downstream right cell l and through cell m to the cell i mixed through and right flow. Because of the vehicle flow of right and through mixed, the capacity of the downstream through cell will limit the right turn cells, and vice versa, that is :

$$q_l \le \beta_l \cdot \frac{q_m}{\beta_m}, \quad q_m \le \beta_m \cdot \frac{q_l}{\beta_l}.$$

The dispersion of the exit platoon considered, within the time $[w\tau, (w+1)\tau]$, the relationship between the diverging cell i and the closest downstream through cell l, right turn cell m is :

$$k_i(w+1) = k_i(w) + \frac{q_i(w) - q_l(w) - q_m(w)}{L_i} \tag{8}$$

$$q_l'(w) = \min\left\{\beta_l \cdot L_l \cdot k_l(w), Q_l(w), \delta \cdot L_l \cdot [k_{l,jam}(w) - k_l(w)], \beta_l \cdot \frac{q_m'}{\beta_m}\right\} \tag{9}$$

$$q_m'(w) = \min\left\{\beta_m \cdot L_l \cdot k_l(w), Q_m(w), \delta \cdot L_m \cdot [k_{m,jam}(w) - k_m(w)], \beta_m \cdot \frac{q_l'}{\beta_l}\right\} \tag{10}$$

$$q_I(w) = \sum_{i=1}^{w-\tau} F(1-F)^{i-1} q_I'(w-i), \quad I = l, m \tag{11}$$

$$Q_I(w) = \begin{cases} 0 & \text{if } w \text{ is red} \\ C_1 \cdot \beta & \text{if green is } \beta \text{ in } \tau \end{cases}, \quad I = l, m \tag{12}$$

The traffic flow model of ordinary link, diverging and merging cell on the secondary road are basically the same as that of the main line.

3. Artery coordination control model

The coordination optimal problem of signal cycle, splits and offsets are so complex to solve. In this paper, in terms of the decentralization-coordination thought of a large system, the bi-level hierarchical structure is used, as shown in figure 3.

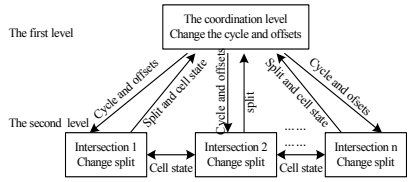


Figure 3 The bi-level hierarchical structure

The first level is to coordinate, which is mainly used to acquire common cycle and offsets of intersections n in terms of traffic conditions on the main and secondary roads. The second level is to control, which is composed of n intersection controllers with similar structure, and its output is the corresponding split respectively. The coordination level and the control level interact. The coordination level delivers common cycle and offsets to the control level, while the control level optimizes signal splits according to the traffic conditions of the main and secondary roads and then feeds back them to the coordination level in time.

3.1 The goal of coordination level

To prevent drivers from getting confused and eye strain caused by the excessively frequent change of signal cycle and offsets, the change of signal cycle and offsets is limited by using the small step length or limiting their change frequency. In this paper we adopt the latter, i.e., signal cycle and offsets are adjusted about every 6 cycles.

The control goal of an artery system is to obtain the widest through-band or the smallest delays etc. When the main line has acquired the widest through-band to make the main road traffic flow unblocked as far as possible, an excessive increase of the delays in the secondary road may occur and consequently a larger delay of the whole system will be caused. So this paper considers the purpose of the artery system management, the smallest delays are taken as the control goal of the artery system. Namely:

$$X(T(w),\varphi(w),\lambda(w)) = \min_{T,\varphi}\left\{\sum_{w=1}^{Z}\sum_{i} d_i(w,\lambda)\cdot\tau\right\} = \min_{T,\varphi}\left\{\sum_{w=1}^{Z}\left[\sum_{i}(k_{westi}(w,\lambda)\cdot L_i - q_{westi}(w,\lambda)) + \sum_{m}(k_{eastm}(w,\lambda)\cdot L_m - q_{eastm}(w,\lambda))\right]\right\} \quad (13)$$

where $k_{westi}(w,\lambda)$ is the density of upstream cell i of the west-east direction, $k_{eastm}(w,\lambda)$ is alike. $q_{westi}(w,\lambda)$is the flow volume of the closest downstream cell i of the west-east direction, $q_{eastm}(w,\lambda)$ is alike. $\lambda(w)$, determined by the control level, is the vector of signal splits. $\varphi_{i,j}(w,\lambda)$ is the offset of the closest intersection i, j; $T(w,\lambda)$ is common cycle of artery system, Z is about 6 times of the cycle.

3.2 The restriction of coordination level

The offsets relationship between west-east and east-west direction is:

$$\varphi_{i,j}(w) + \varphi_{j,i}(w) = T(w) \quad (14)$$

Define V as the limited speed on link and l_i as the distance between the intersections. Because of offsets, intersection delays are changed. The delays from the upstream intersection to downstream intersection fall into 2 categories: first, the head vehicle of the platoon is stopped by a red light when arriving at the downstream intersection; second, the tail of the platoon is stopped by the red light when arriving at the downstream intersection which is just changing from green light.

(1)The head vehicle is stopped at the next intersection

Define $\tau_{j,down}$ as the time interval from the arrival of the head vehicle to the red end in cell j on the downstream intersection. Then:

$$\tau_{j,down} = \varphi_{i,j} - \left[\frac{l_i}{V}\right] (\mathrm{mod} T)$$

where *modT* is the residue divided by T.

(2)The tail vehicle is stopped at the next intersection

Define $\tau'_{j,down}$ as the time interval from the first vehicle of the platoon encountering red to the last one arriving in cell j on the downstream intersection. Then:

$$\tau'_{j,down} = (\varphi_{i,j} - \left[\frac{l_i}{V}\right] (\mathrm{mod}\ T))$$

The relationship between β in Eq. 7 and 12 and $\tau_{j,down}$, $'\tau_{j,down}$ is :

$$\beta = \tau - \tau_{j,down} (\mathrm{mod} r) \quad \text{or} \quad \beta = \tau - \tau'_{j,down} (\mathrm{mod} r) \tag{15}$$

The control problem of the coordination level is to look for the combination of cycles and offsets which satisfy the restriction Eq. 1-12 and 14-15 and make Eq. 13 minimal.

3.3 The goal of control level

To each intersection, the control goal should be to optimize and choose signal splits under a given cycle, and to make the intersection delays fewer. That is, to the cell set I of the intersection k, $i \in I$, then:

$$X(T(w),\varphi(w),\lambda(w)) = \min_{\lambda}\left\{\sum_{w=1}^{Z'}\left[\sum_{i}(k_{westi}(w,\lambda)\cdot L_i - q_{westi}(w,\lambda)) + \sum_{m}(k_{eastm}(w,\lambda)\cdot L_m - q_{eastm}(w,\lambda))\right]\right\} \tag{16}$$

where Z' is about 6 times of cycle, $\varphi(w), T(w)$ is determined by the optimization of coordination level.

3.4 The restriction of control level

If λ_{min}, λ_{max} are the top and bottom boundary of the signal split respectively, to the intersection i, there exists

$$\lambda_{\min} \le \lambda_i(w) \le \lambda_{\max} \tag{17}$$

The control problem of the control level is to look for the signal split $\lambda_i(w)$ which satisfies the restriction Eq. 4-12 and 14-15 and make Eq. 16 less.

4 Modified particle swarm algorithms

Particle Swarm Optimization (PSO) is a kind of new global optimal algorithm presented by Doctor Eberhart and Kennedy (Kennedy, 1995), which is particularly applicable to the optimization problem of nonlinear functions. First, the algorithm will randomly produce a particle swarm from the solution space. Then it will seek the optimal solution through an iterative method. In iteration, Particles will update themselves through following two "extremums": the optimal solution P_{best} competed by Particles themselves and the current optimal solution G_{best} found in the whole Particle Swarm. The basic PSO algorithm of the particle velocity and position updating equation is:

$$V_{step} = w_now \times V_{step} + c_1 \times rand \times (P_{best} - Swarm) + \ c_2 \times Rand \times (G_{best} - Swarm) \tag{18}$$

$$w_now = w_now - Iteration \times \frac{w_start - w_end}{IterationMax} \tag{19}$$

where V_{step} is Particle velocity, *Swarm* is current Particle position, *rand* and *Rand* are random numbers from 0 to 1 respectively, c_1 and c_2 are learning parameters and usually takes $c_1=c_2=2$, *w_start* is the maximal inertial parameter and generally takes 0.9, *w_end* is the minimum inertial parameter and generally takes 0.1, *Iteration* is the current iterative times, and *IterationMax* is the maximal iterative times.

Because the basic PSO algorithm may easily fall into the local minimum point at the early searching stage, Shi et al proposed the adaptive inertial parameter algorithm (Shi, et al, 1995) which adds the inertial parameter to the basic PSO algorithm to improve its optimal performance. In this paper, the real time of the algorithm and the accuracy of the solution are comprehensively considered by following the thought of the adaptive inertial parameter algorithm and choose the inertial parameter according to the following rules: at the prior searching stage, the bigger inertial parameter is adopted to favor quick convergence and avoid sinking into the local minimum point. While at the later stage, the smaller inertial parameter is adopted to facilitate the local search and ensure the convergence accuracy of the algorithm. During the whole search, the parameter can adjust the particle balance between the global and local search, at the same time, the searching performance is more flexibly. The mathematic expression is:

$$w_now = \begin{cases} 1 - 1/m, & \text{if } 1 \le \text{Iteration} < (\text{IterMax} - \text{int}(e^{\ln(\text{IterMax})-1})); \\ 1 - 2/m, & \text{if } (\text{IterMax} - \text{int}(e^{\ln(\text{IterMax})-1})) < \text{Iteration} < (\text{IterMax} - \text{int}(e^{\ln(\text{IterMax})-2})); \\ \Lambda & \\ 1 - (m-1)/m, & \text{if } (\text{IterMax} - \text{int}(e^{\ln(\text{IterMax})-m+1})) < \text{Iteration} \le \text{IterMax} \end{cases} \tag{20}$$

where *int* means taking the integer, $m \in Z$ is the number of subsection, and should satisfy $ln(IterMax) > m$.

5 Simulation research

In figure 2, the cell length L of the link and the intersection are 100m and 180m respectively. τ is 9s. The inside lane of the main road is for exclusive left turn flow, the central one for through flow, and the outside one is the lane mixed through and right turn flow. The secondary road is two-way and double-lane. The capacity of the main road and the secondary road are 5400pcu/h and 1800pcu/h respectively.

The distance between two intersections is 480m, and the link between them has no vehicles in or out, and the limited speed V is 40km/h. The maximal cycle is 220s. The minimum green time of each phase is 10s. The lost time of each phase is 3s. The phase project is four-phase. The main road has the left turn phase and the through and right turn phase, and the secondary roads have the through and right and left turn phase. Simulation research lasts for about 1h.

The vehicle arrival obeys POISSON distribution, the main road average arrival rates of intersection 1 from west to east are 800, 850, …, 1100pcu/h respectively, intersection 2 from east to west are 700, 790, …, 1240pcu/h respectively. Each flow volume lasts for 9min. The starting time is 8:00. The secondary road average arrival rate takes the constant (400pcu/h). The proportion of left turn, through, right turn flow of the main street on the intersection 1 and 2 are 35%, 50%, 15% and 30%, 52%, 18% respectively. Hence, to intersection 1 and 2, β_l, β_m is 0.54, 0.46 and 0.49, 0.51 respectively. The proportions of left turn, through, and right turn flow on the secondary road are 30%, 30%, 40% respectively. The dispersion coefficient F is 0.35.

In the coordination optimal model of the artery system, the coordination level optimizes common cycle and offsets with Particle Swarm Optimization, whose adaptive function is Eq. 13, and the control level optimizes signal splits with it, whose adaptive function is Eq. 16. The initial state of cells is 0, the initial common cycle is 80s and the offset from west to east is 40s. For the intersection 1 and 2, the signal split of each phase are 21.6%, 30.9%, 11.6%, 11.6% and17.1%, 29.6%, 14.0%, 14.0% respectively. The relative control result between the maximal through-band method and the coordination optimal method introduced in this article is shown in table 1 and table 2.

Table 1 The control result using the maximal through-band method

Time	Common cycle (s)	Split of intersection 1				Split of intersection 2				Offset in west-east Direction (s)	Two-way through-band Width (s)	total delay (pcu.h)
		Phase 1	Phase 2	Phase 3	Phase 4	Phase 1	Phase 2	Phase 3	Phase 4			
8:00:00～8:08:00	80	21.64%	30.91%	11.59%	11.59%	17.09%	29.63%	14.01%	14.01%	40	41.4	41.19
8:08:00～8:16:00	80	21.64%	30.91%	11.59%	11.59%	17.09%	29.63%	14.01%	14.01%	40	41.4	
8:16:00～8:25:30	95	22.93%	32.76%	10.92%	10.92%	18.02%	31.23%	13.65%	13.65%	The distance between		
8:25:30～8:37:18	118	24.19%	34.55%	10.37%	10.37%	18.94%	32.84%	13.34%	13.34%	intersections and split		
8:37:18～8:59:18	220	26.47%	37.82%	9.45%	9.45%	22.61%	39.20%	12.78%	12.78%	of phase 2 are too small		
8:59:18～9:03:00	220	27.25%	38.93%	8.34%	8.34%	23.42%	40.59%	11.72%	11.72%	to realized the coordination Control.		

Table 2 The control result using the coordination optimal method

Time	Common cycle (s)	Split of intersection 1				Split of intersection 2				Offset in west-east Direction (s)	Two-way through-band Width (s)	The total delay (pcu.h)
		Phase 1	Phase 2	Phase 3	Phase 4	Phase 1	Phase 2	Phase 3	Phase 4			
8:00:00～8:08:36	86	19.59%	40.41%	12.50%	12.50%	14.75%	45.24%	12.50%	12.50%	43.0	73.1	
8:08:36～8:21:30	129	20.94%	54.25%	7.75%	7.75%	19.06%	40.94%	7.75%	7.75%	61.2	80.8	
8:21:30～8:34:26	129	19.84%	55.35%	7.75%	7.75%	18.53%	41.47%	7.75%	7.75%	63.3	82.9	37.23
8:34:26～8:51:38	172	20.50%	60.90%	5.81%	5.81%	20.49%	60.91%	5.81%	5.81%	86.0	122.0	
8:51:38～9:03:00	172	20.71%	60.70%	5.81%	5.81%	20.71%	60.70%	5.81%	5.81%	86.0	122.0	

According to table 1 and table 2, compared with the maximal through-band method, the coordination optimization can lower total delay time by 10.64%. In the maximal through-band method, the optimal cycle length of each intersection is determined by Webster method. At the same time, as the common cycle is the biggest period, when the distance between intersections and the signal split of the through and right-turn phase are both small, the coordination control of the artery system can't be realized. In the coordination optimization, the coordination control has been implemented. The delays of through and right-turn flow on the main road and the system delays decrease. However, the delays of the left turn flow on the main road and the delays on the secondary increase. The software, developed in the environment of Matlab, is of high executive efficiency and the solving time is 50s (the iterative times are 1000, the number of seed swarm is 10, the number of variable is 6, CPU is Pentium IV). It is relatively suitable for the control system whose control project is selected.

6 Conclusions

In this paper, based on the modified CTM, the dispersion delay of the exit platoon is considered, the intersection traffic flow model is built. The bi-level hierarchical structure is built to synchronously design control parameters of artery system. At the coordinated level, the common cycle and two-way offsets are optimized simultaneously in accordance with traffic flows on the main road and secondary roads to minimize the delays. At the control level, the split of each intersection is optimized to minimize the delays. Applying the modified PSO algorithm to the bi-level control system, the optimal problem of the coordination and control level is solved respectively. The coordination optimization is applied to the simulated artery system. The control result is better than maximal through-band method. Further work should focus on the multi-objective optimization combined with

the minimal parking time or the coordination with other management measures (e.g. route guidance) to improve traffic management for a larger-scale road networks.

Acknowledgements

This work is sponsored by National Natural Science Foundation of China (No.50348027), the Ministry of Communications of China Application Foundation (No. 2006319815080, 200431982515) and HuNan Province Natural Science Foundation (No. 06JJ50104).

References

Chang, Y. T., et al.(2003). "Urban arterial road coordinate control based on genetic algorithm." Journal of Traffic and Transportation Engineering, 3(2):107-112.

Chen, S. F., and Xu, J. Q., and Chen, H., et al (1998). "The Examination of Fuzzy Control for Traffic Lights in an Urban Arterial Road." Operations Research and Management Science, 7(1):35-41.

Daganzo, C. F. (1995)."The cell transmission model part II：network traffic.". Transportation Research Part B, 29(2)79-93.

Gartner, N. H., and Stamatiadis, C.(2002). "Arterial-Based Control of Traffic Flow in Urban Grid Networks." Mathematical and Computer Modelling, (35):657-671.

Kennedy, J., et al. (1995). "Particle swarm optimization." Perth，Aust: Proceedings of the 1995 IEEE international conference on Neural Networks, (4):1942-1948.

Shi, Y., et al (1999). "Empirical study of particle swarm optimization." Washington,USA: Proceedings of the 1999 congress on Evolutionary Computation, (3):1927-1930.

Wan, X. J., and Lu, H. P. (2001). "An Optimal offset Model for artery traffic signal control system." China Journal of Highway and Transport, 14(2):99-102.

Wei, L., Chen, Y. S., and Yasunoyi, L. (1994). "The Optimal Offsets Pattern and Development of Simulation in Signals Control System on Urban Main Road." Journal of Traffic and Transportation Engineering, 2001, 1(1):72-76.

Wey, WannMing.(2000). "Model formulation and solution algorithm of traffic signal control in an urban network." Computers, Environment and Urban Systems, (24):355-377.

Yang, P. K. (1995). *Traffic Management and Control*. Beijing: China Communications Press,.

Zhu, Y. (2003). *Research on Theory and Methods of Urban Traffic Signal Control*. Beijing: Northern Jiaotong University.

A Modified Heuristic Search Algorithm for Pedestrian Simulation

GAO, Peng[1], XU, Ruihua[1] and ZOU, Xiaolei[1]

[1] Department of Transportation Management Engineering, School of Transportation Engineering, TongJi University, 4800 CaoAn Road, JiaDing campus, Shanghai 201804; email: TJ_JT_GP @yahoo.com.cn

Abstract: A* algorithm is a typical heuristic search algorithm, which is widely used in Game Programming, Artificial Intelligence and robotics. However, AI programmers have found again and again that the standard A* algorithm can be woefully inadequate for achieving realistic movements.

A modified A* algorithm is proposed in this paper to solve the pathfinding problem for an Agent-based pedestrian simulation system. First, several heuristic functions are introduced and compared. Second, a “Trimming Algorithm” has been developed to achieve smooth straight-line movements. Third, the problem of the output path “clinging to obstacle” of the standard A* algorithm has been solved by imposing a penalty on the nodes near obstacles in map preprocessing. With these modifications, we can get a more realistic path like a human trail, the feasibility and efficiency having been proved in the simulation tests.

Keywords: pedestrian simulation, path finding, path planning, A* algorithm

1 Introduction

Pathfinding is a core component of many intelligent agent applications, ranging in diversity from commercial computer games to robotics. Characters should move in some goal-directed manner, and the program must be able to identify a good path from an origin to a goal, which both avoids obstacles and is the most efficient way of getting to the destination.

Despite the fact that the industry as a whole is quite aware of the "magic A* algorithm", it remains a perennial problem because the implementation is so game specific. For example, AI programmers have found again and again that the standard A* algorithm can be woefully inadequate for achieving the kind of realistic movements required in games.

This paper focuses on several techniques for achieving more realistic results from path finding. All the techniques discussed here were used in the development of an Agent-based pedestrian simulation system, which was made for startlingly more realistic and human-like movement for larger numbers of pedestrians in the simulation.

The focal topics presented here include:

(1) A*'s ability to vary its behavior based on the heuristic functions can be very useful. We can simulate the variety of pedestrian route choices by designing diverse heuristics. In chapter 2, several heuristic functions are introduced and compared.

(2) A "Trimming Algorithm" is presented in chapter 3, which could reduce the "zigzag effect" and achieve smooth straight-line movements.

(3) The problem of the output path "clinging to obstacle" of the standard A* algorithm has been solved in chapter 4 by imposing penalties on nodes near obstacles in map preprocessing.

2 Heuristic

2.1 The Meaning of heuristics

Our application of interest is grid-based pathfinding. An agent has to traverse through a two-dimensional map. The searched area is simplified and the map is divided into a square grid; each square can be identified by a unique ID.

Consider Figure 1: Suppose an agent wants to travel from location A (green square) to location B (red square), and there is a wall (blue squares) between them.

1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	A	17	18	19	B	21
22	23	24	25	26	27	28
29	30	31	30	33	34	35

Figure 1. Grid map

Within one step, it is possible for the agent to move to one of the 8 adjacent squares, which is walkable. The cost of straight movement can be defined as *"Cost_Straight"* (take 100 in this paper), and the cost of diagonal movement can be defined as "*Cost_Diagonal"* (take 141 in this paper). The reason of taking these values is that the "Cost_Diagonal" is $\sqrt{2}$ times more than "Cost_Straight", approximately 1.41 times.

Heuristic searching means calculating the cost of every location in the search space and choosing the lowest one for each step. Therefore, tremendous unnecessary searching work can be avoided and high efficiency can be obtained.

The key to determining which squares to use when figuring out the path is the following equation:

$$F(n) = G(n) + H(n) \quad (1)$$

Where:

G = the cost to move from the starting point A to a given square on the grid, following the path generated to get there.

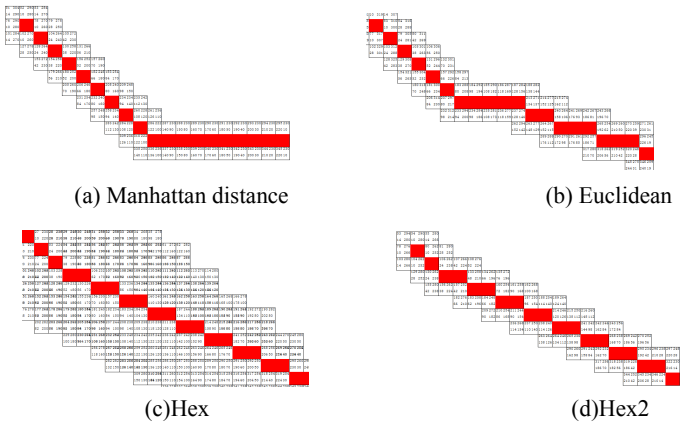
H = the estimated cost to move from that given square on the grid to the final destination, point B.

H often refers to the heuristic, because we really do not know the actual distance until we find the path; all kinds of barriers can be in the way (walls, water, etc.).

2.2 Heuristic functions

The absolute difference between the row of square n and the row of destination can be defined as "*diffRow*", and the corresponding column difference can be defined as "*diffCol*". So the standard heuristic "Manhattan distance" can be formulated as:

$$H(n) = (diffCol + diffRow) * Cost_Straight \quad (2)$$



(a) Manhattan distance (b) Euclidean

(c)Hex (d)Hex2

Figure 2. Various output paths by adopting different heuristics

With this simple and low CUP consumption heuristic, we often get a bent path on a map which has no obstacles (Figure 2 a). One of the factors contributing to the problem is that the estimated costs are too high, if H(n) is always lower than (or equal to) the cost of moving from square n to the goal, then A* is guaranteed to find the shortest path. The lower the H(n) is, the more node A* expands, making it slower. At one extreme, if H(n) is 0, then only G(n) plays a role, and A* turns into Dijkstra's algorithm, which is guaranteed to find the shortest path; at the other extreme, if H(n) is

highly related to G(n), then only H(n) plays a role, and A* turns into BFS.

Euclidean distance is the "straight line distance", which can get a better path (Figure 2 b) with a higher calculation price:

$$H(n)=Cost_Straight \times \sqrt{diffRow^2 + diffCol^2} \quad (3)$$

If you want to get rid of the "polyline" in the output path of Manhattan distance, some modifications are necessary. One of the best solutions is "Hex Distance":

$$H(n) = (Max\{0, diffRow - diffCol / 2 - Correction\} + diffCol) * Cost_Straight \quad (4)$$

Where:

$$Correction = \begin{cases} colTarget \% 2, & \text{when rowTarget} < \text{rowStart and diffCol is odd} \\ colStart \% 2, & \text{when rowTarget} > \text{rowStart and diffCol is odd} \\ 0, & \text{when diffCol is even} \end{cases}$$

"*rowStart*" stands for the row index of square n, and "*colStart*" stands for its column index.

"*rowTarget*" stands for the row index of destination, and "*colTarget*" stands for its column index.

The output path by adopting "Hex distance" is more realistic than that of "Euclidean distance" with a lower calculation price, but it makes A* expend more nodes. We can obtain "Hex2 distance", which is able to make A* expend fewer nodes and find a similar path, by raising the estimated cost:

$$H(n) = (Max\{0, diffRow - diffCol / 2 - Correction\} + diffCol) * Cost_Diagonal \quad (5)$$

The meanings of variables are consistent with Eq. 4.

It is worthwhile to note that although we can find a similar path by Hex2, we could not simply regard Hex2 as a better heuristic than Hex in any case. If the map size is comparatively small and the map structure is somewhat simple, A* might find similar paths, or might not.

Finally, there are 7 heuristics, respectively named "Hex", "Hex2", "Hex3", "Manhattan", "Euclidean", "Octile" and "Octile_UniCost", which can be used to develop different kinds of path planning algorithms. So we can simulate various pedestrians' preferences for route choices.

Figure 3 shows 7 paths by 7 search algorithms (respectively in red, orange, yellow, green, cyan, blue and violet, possibly overlapped).

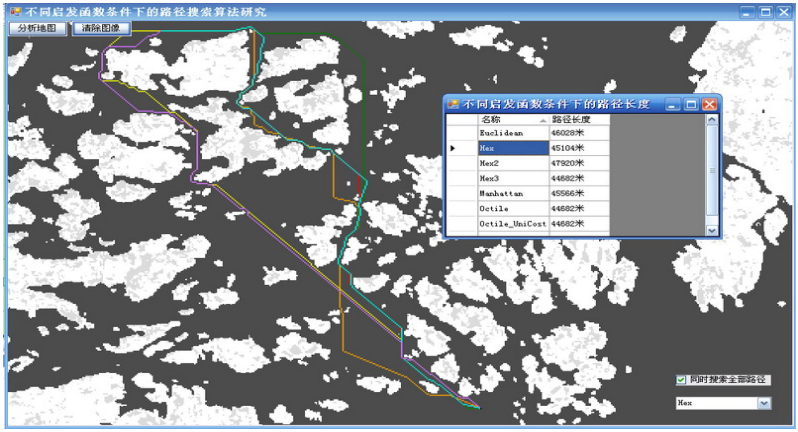


Figure 3. Seven paths by seven search algorithms

3 Achieving smooth straight-line movement

As mentioned above, although we can simulate various pedestrians' preferences for route choices by adopting diverse heuristics, some heuristics may lead to an unfortunate "zigzag" effect in Pathfinding (as shown in Figure 3). One of the reasons is the output paths may contain redundant nodes, which must be found and removed. This paper presents post processing solutions for smoothing the path, whose core component is a recursive algorithm named "Walkable". The fundamental principle is to check every 3 waypoints along the path; if there are no obstacles between the head and the rear, the middle waypoint is redundant. The pseudo code for the "trimming algorithm" as follows:

To judge whether there are any obstacles between the point "P1" and the point "P2":

Step 1: Calculate the straight-line distance "d" between P1and P2;

Step 2: If d < tile width, then return and assert "there is no obstacle";
else turn to Step 3;

Step 3: If the middle point of P1 and P2 （named "Center"） is a blocked tile,
then return and assert "there are some obstacles";
else turn to Step 4;

Step 4: Judge recursively whether there are any obstacles between P1 and Center, Center and P2, if any, then return and assert "there are some obstacles";
else return and assert "there is no obstacle".

After the post processing trimming, the output path is very short and includes no redundant nodes (as shown in Figure 4 and Figure 5), so it can serve as Macro-navigation for virtual pedestrians in the simulation.

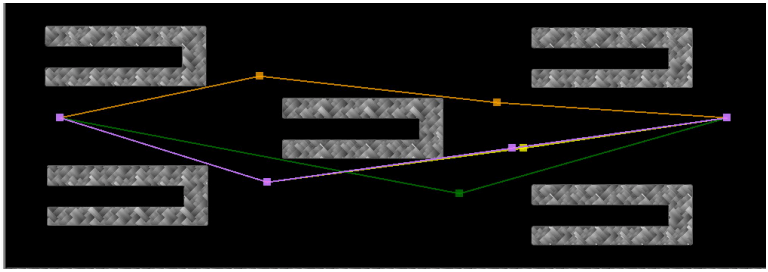


Figure 4. Paths after Trimming

4 Toward more realistic Pathfinding

Most pedestrians are inclined to walk on "the shortest paths" in real life and seldom love getting too close to obstacles. Unfortunately, the output paths of standard A* algorithm seem to be "clinging to obstacles" too closely, and nobody would move in this way, as Figure 5 and Figure 6(a) show..

With map preprocessing, the problem has been solved by imposing a penalty on nodes near obstacles. The map preprocessing algorithm calculated the distance between every node of the map to its nearest obstacle. Then we can add a penalty function to equation (1):

$$F(n) = G(n) + H(n) + P(d) \quad (6)$$

With d= the distance between node n to its nearest obstacle.

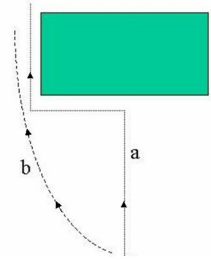
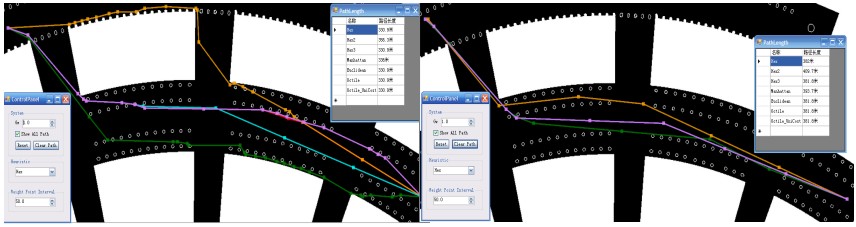


Figure 5. Unrealistic path



(a) Output paths by standard A* algorithm (b) Output paths by modified algorithm

Figure 6. Paths before and after disposing

Figure 6 shows a part of the Inner Concourse of the Athens Olympics 2004. By

comparing (a) and (b) of Figure 6, it is obvious that the modified algorithm, which achieved "line of sight" movements and kept a certain distance away from obstacles, has made the difference.

The fundamental principle of map preprocessing is to calculate the distance between every node and its nearest obstacle; the pseudo code is as follows:

Step 1: Let R=1, R is the index of "square layer", which is composed of 8*R nodes, around the given node;
Step 2: If "R< map width or map height" then turn to Step 3;
else turn to Step 4;
Step 3: Analyze every node of the "square layer":
If you come across a blocked node, turn to Step 4;
else let R=R+1, turn to Step 2;
Step 4: Obtain the distance between the given node and its nearest obstacle:R,
return.

5 Conclusions

A* is a well-known algorithm, widely used in commercial games and AI. To some extent, we should not merely regard A* as a concrete algorithm, but a thought or a framework instead. Based on the framework, implementation can be various. So we can develop many new algorithms, which could meet some specific needs.

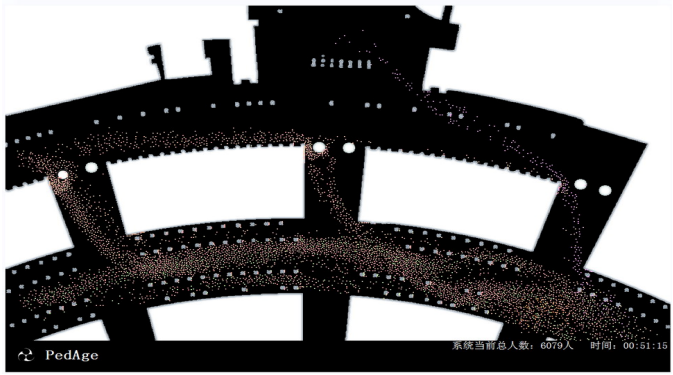


Figure 7. A scenario of the Athens Olympics 2004

A modified A* algorithm is proposed in this paper to solve a pathfinding problem for an Agent-based pedestrian simulation system. Figure 7 depicts a scenario of the Athens Olympics 2004. The effect implies that we can achieve more realistic movements with the modifications proposed in this paper. The point to emphasize here is that the output paths by search algorithm do not serve directly as the factual trail, but as a kind of macro-navigation. The movement of every pedestrian depends on both

macro-navigation and crowd dynamics.

References

Anthony Stentz. (1995). “The Focused D * Algorithm for Real-Time Re- planning” [A]. In：Proceedings of the international Joint conference on A artificial Intelligence[C].

Dirk Helbing, Martin Treiber. (1998). “Jams, Waves and Clusters” [J]. Science , 282, 2001 - 2003

Dirk Helbing . (1997). “Modeling the evolution of human trail systems” [J]. Nature, 388, 47-50

G. Keith Still . (2004). “Crowd Dynamics” [D]. London: University of Warwick, BSc Physical Sciences (Robert Gordon's Institute of Technology, Aberdeen 1981)

S.P. Hoogendoorn & W.Daamen. (2004). “Design assessment of Lisbon transfer stations using microscopic pedestrian simulation” [J].Computers in Rilways, IX ,135-147

Winnie Daamen. (2004). “Modeling Passenger Flows in Public Transport Facilities” [D]. Delft: Delft University of Technology

SIGNALIZED INTERSECTION DESIGN AND OPERATIONS: EXPERIENCES IN GERMANY AND JAPAN

Keshuang TANG[1] and Hideki NAKAMURA[2]

[1]Department of Civil Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan; PH (81) 52-7895175; FAX (81) 52-7893837; email: tang@genv.nagoya-u.ac.jp
[2]Department of Civil Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan; PH (81) 52-7892771; FAX (81) 52-7893837; email: nakamura@genv.nagoya-u.ac.jp

Abstract: An official nationwide manual on intersection planning and design remains absent in China. Presently, a research group led by Tongji University is addressing its first version. Toward gaining more profound knowledge on state of the art in intersection design and operations in overseas while developing Chinese manual, this paper thus presents German and Japanese experiences. Conclusions, future studies and efforts essential for China are discussed finally.
Key Words: *Intersection design, signal operations*

1 Background

In the field of traffic engineering, a nationwide guideline for intersection planning and design remains absent in China, where engineers and operators currently apply regional manuals developed on the basis of local traffic characteristics or documents combining Chinese situations with international experiences, published by research institutions. In order to establish a Chinese national standard, Ministry of Construction, China, has organized a research group led by Tongji University to address the first version of *Manual on At-grade Urban Intersection Planning and Design*. It will be based upon the comprehensive research efforts in the field of intersection done in China and practical experiences so far. Meanwhile, existing guidelines and current practice related to intersection design and operations in typical developed countries, such as Japan and Germany, will be significant references.

It should be stated that the fundamental philosophies for intersection design and operations in Germany and Japan vary considerably. They create distinguished features both in intersection geometry and signal control, which likewise lead to the variations in intersection operational and safety performance and might have effects on user behavior too. It could easily be a consensus that each philosophy possesses benefits as well as drawbacks in terms of reliability, efficiency and applicability. Therefore, when developing a Chinese manual, it is of great significance to obtain more profound knowledge on them, which is the objective of this study.

Note that this study focuses on at-grade signalized intersections only. Both guidelines and practice will be concerned in the paper. Thereby, German Partial Revision 2003 of Guidelines for Signal Control (RILSA, FGSV, 2003), German

Guidelines for Intersection Design (RAS-K, FGSV, 1988), Japanese Edition 2006 of Manual on Traffic Signal Control (MTSC, JSTE, 2006) and Japanese Manual on At-grade Intersection Planning and Design (JSTE, 2004) act as the basis of this study since they partly reflect the latest research results and guide current practice. Additionally, attentions should be given throughout the paper that left-hand traffic is legal in Japan.

Structure of this paper is arranged as follows: fundamental philosophies for intersection design and operations in Germany and Japan will first be introduced in Section 2; Section 3 describes the basic features in geometric design and signal operations; finally, Section 4 will outline the conclusions, and discuss the future research and efforts necessary for improving current situations in China.

2 Fundamental Philosophies for Intersection Design and Operations in Germany and Japan

In Japan, current road planning and design methodologies are basically link-based. So far, a quality of service concept has not been introduced during road planning and design process in practice. Such facts result in that intersection design and operations are largely dependent upon the crossing road classes and section forms instead of essential performance of the intersections, determined by its specific functions in whole road network and OD properties of the prevailing traffic. It impedes efficient utilizations of intersections sometimes. For instance, in the case that very light demand of heavy vehicles exists at the intersections located in urban area due to road planning, wider lanes and larger turning radius become demanded in order to deal with those heavy vehicles. As a result, intersection scales have to expand, which not only requires more land use but also brings out unnecessary long cycle lengths because of longer minimum green times for pedestrians and clearance times. Consequently, capacities of the intersections may increase, but delays increase as well. In the aspect of signal control, it should be mentioned here that although delay evaluation methodologies have been provided in MTSC (JSTE, 2006), their utilizations in practice are quite inadequate (Nakamura et al. 2006). In addition, police is responsible for intersection design and operations in Japan and their primarily pursued objectives are safety and capacity. Therefore, stable phase sequences and typical two-, three- or four-phase plans, decreasing phase switch frequency, long cycle lengths, enabling higher capacities, and longer clearance times, effective to reduce the conflicts during phase change intervals according to police's opinions, tend to be used even at the intersections without very high traffic load in addition to those key intersections. Generally speaking, Japanese philosophy fosters unnecessarily large intersection sizes in some cases, uniform control patterns as well as relatively long cycle lengths as compared to other developed countries.

In Germany, road planning and design methodologies more emphasize bottleneck-based in practice. Such methodologies allow to explicitly define the suitable performance of intersections based on its unique utilization characteristics, including its positions in road network and OD properties of the prevailing traffic, during the road planning and design process. Hence, a so-called performance-oriented

philosophy somewhat underlies in German guidelines for intersection design and operations. It is no doubt that intersection planning and design in Germany also relies on crossing road classifications to a great extent. However, as service quality management procedures are well established in Germany (Boltze, 2006), elaborate and flexible treatments for achieving necessary performance, ensuring safety but not at the expense of service quality of traffic flow, have been taken into account in the process of intersection design and operation. Those treatments include compact intersection size, channelized right-turn lane, alarm signal display (red-amber before the green on-set), camera enforcement measure and movement-based adaptive control policy, etc. Usually, alternative plans correspondent to the utilization characteristics are projected in advance and then the one that enables to match the requirements best will be selected after comparison, accounting for mobility and safety (the basic procedure is consistent in Japan). In general, ensuring service quality through flexible treatments of intersection design and operations is primary objective of German philosophy.

Compared with German one, Japanese philosophy is convenient to be executed in practice and cause less costs. It is also able to achieve higher capacities, owing to longer cycle lengths and less number of phase change in each cycle (resulting in less start-up lost time), as well as lower phase switch frequency and longer clearance time, which may have safety benefits. However, Japanese experts have increasingly argued that densely spaced intersections and limited land use in Japan call on compact intersection sizes and short cycle lengths as users have to experience more delay in the case of high intersection density. On the other hand, longer cycle length will result in more waiting time for users, which probably induce red-light violations. In addition, longer clearance time may encourage more stop-line crossings during intergreen time interval, which increases vehicles' exposure probability to conflicting traffic during phase change interval, in reverse to police's opinions. Thus, a few Japanese professionals believe that Japanese philosophy is unable to fully dig potential capability of facilitating mobility and safety at intersections. They have put forward a bottleneck-based road planning and design concept and recommended that intersection design and operation policies currently applied in Japan should also be reexamined (Nakamura et al. 2006).

3 Basic Design and Operational Features at German and Japanese Signalized Intersections

In fact, intersection design and operations are iterative process in which geometric design and signal operation treatments link each other and behave in a manner of adjusting each other to reach the design targets. In the stage of geometric design, major concerns include horizontal and vertical alignment, lane configuration and width, setback of stop-line, turning radius and crosswalk width. Control strategy, signal phasing, green time and intergreen time, etc. are figured out during signal operation stage. This part will address basic features of the key elements in the two nations by way of guideline reviewing and practice introducing.

3.1 Geometric design

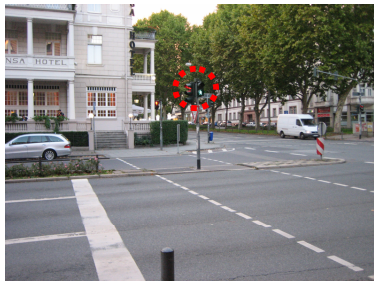
(1) Lane width

In Japan, the policy for determining lane widths at intersections is rather close to that for road sections, which aims to preserve enough lateral space for vehicle maneuver with the considerations of mobility and safety. Based on the target capacity, prevailing traffic demand and crossing road classes, lane widths are usually selected among several uniform values between 2.75 and 3.5 m, such as 3.0, 3.25 and 3.5, etc. If exclusive turning lanes required, standard lane width for them is 3 m. It could be reduced in the cases of low heavy vehicle percentages and limited land use, however no less than the minimum value of 2.5 m. In practice, for the sake of safety, lanes less than 2.5 m are allowable only in very exceptional cases that traffic demand is extremely over-loaded and road space expansion is impossible. In Germany, it is prior to ensure necessary number of approach lanes in order to maintain capacity when making decisions on lane widths. For this purpose, lane widths are flexibly determined in terms of required capacity and design vehicle. If necessary, e.g. when exclusive left turn lanes needed for accommodating increased left-turning traffic demand and separate lane usages demanded for implementing a movement-based multi-phase plan, lanes less than 2.5 m wide may also be applicable in spite of they probably influence discharge rate negatively and present sideswipe conflicts among the movements in the same direction. Even though such very narrow lanes are also not common in Germany, they can more often be seen than in Japan.

(2) Stop-line setback and turning radius

Stop-line setback and turning radius are handled simultaneously in the procedure of geometric design as they correlate each other. In Japanese guidelines, it is recommended that certain turning radius and setback must be maintained to guarantee smooth turnings and avoid disturbances among vehicle trajectories. Provided that the previous requirements have been satisfied, stop-lines will be set close to the downstream exits as much as possible. The basic policy is consistent with German guidelines in which right-turning radius is recommended to be as small as possible to reduce vehicle speeds so that enough perception and reaction time become available for drivers to detect surrounding non-motorized traffic. In addition, left-turning radius is chiefly based on vehicle paths inside intersection and only crosswalk setbacks below 5-6 m are allowed. It is also stated in German guidelines that, to provide safe space for vehicle turnings, stop-lines at left-turning lanes in certain cases need to be shifted backward. However, in practice, curb radius and setback tend to be too much larger correspondent to design vehicle in Japan, which sometimes is due to provide enough space for left-turning vehicles at shared through-left turn lanes when they are waiting for acceptable gaps of pedestrians to traverse. It somewhat avoids impeding discharge rate of behind vehicles queued at the shared through-left turn lanes. However, it enlarges intersection size and increases the speeds of left-turning vehicles, which threatens pedestrian safety. Moreover, illegal parking at the curb corners also often appeared to be a problem in last years in Japan since spare space is available at the corners.

(3) Crosswalk



Median refuge islands (with signals) at crosswalks for pedestrian twice crossings (Germany)

Median barrier (without signals) not utilized as refuge islands for pedestrian two-step crossings (Japan)

Figure 1. Common crosswalk forms and median barrier utilizations at medium- and large-size signalized intersections in Germany and Japan

Both Japanese and German guidelines recommend a default value of 4 m and a minimum value of 3 m as the reference crosswalk widths. In the case of heavy pedestrian demand, greater values will be adopted. In addition, 1-2 m ahead of stop-line is proposed to be a suitable position for crosswalks in Japan, 1 m in Germany. In practice, at medium and large-size intersections in Germany, median refuge islands for pedestrians at crosswalks are usually present (Figure 1). Due to that, flexible phasing plans, releasing pedestrians separately not providing successive crossings, become possible. As minimum pedestrian green time is shorter in that case, cycle length can also be decreased. Because of the decreased cycle length, the pedestrian demand per cycle drops, which thus allows relatively narrow crosswalk widths and less waiting space. However, pedestrians are probably hurt by running vehicles and feel uncomfortable when waiting at the refuge islands for the second crossings. Therefore, it is also recommended that refuge islands should be at least 2 m wide in German guidelines in order to protect pedestrians. However, the pedestrian demand per cycle in Japan is averagely higher than that in Germany because of the relatively long cycle lengths and total pedestrian demand. Thus, the crosswalk widths tend to be larger, e.g. exceeding 4 m. In addition, in spite of median barriers are usually present at the crosswalks of medium- and large-size intersections in Japan, very few of them have been utilized as pedestrian refuge islands for two-step crossings due to phasing plans at the intersections (Figure 1).

(4) Channelization

A significant feature at many German intersections, especially large-size intersections, is the existence of channelized right-turn lanes with raised islands (Figure 2). Less (refers to channelized left-turn lanes here) can be found in Japan however. Channelizing islands provide for the separation of conflicting traffic movements into defined paths of travel to facilitate the safe and orderly movement of vehicles, pedestrians and bicycles. Its benefits are also obvious in shortening the distance between the opposing stop-lines, which results in shorter clearance time. However, research (Fitzpatrick et al. 2005) has shown that comparably high speeds of

turning traffic and its merging maneuver with a limited sight for left-side traffic likely causes rear-end collisions at channelized right-turn lanes. In addition, it needs more land use, which partly explains why few intersections in Japan have implemented channelized left-turn lanes. In the case that land use is available, pavement makings are more often applied for separating left-turn movements in Japan (Figure 2). However, pedestrian crosswalks are not correspondingly set forward in order to allow the use of marking areas for pedestrian's crossings. In addition, it seems seldom efficient in defining the vehicle paths compared with raised islands according to the authors' observations as drivers usually neglect the pavement markings and utilize the marking areas for waiting or driving.

Channelization with raised islands (Germany)

Channelization with pavement markings (Japan)

Figure 2. Channelizations at signalized intersections in Germany and Japan

To a summary, German policy for geometric design is more flexible and able to compact intersection size, which allows smaller minimum pedestrian green time and less clearance time, contributing to shorten cycle length. Owing to channelized right-turn lanes and median refuge islands at crosswalks, it also enables flexible phasing plans, such as two-step crossings for pedestrians and right-turn on red. Japanese policy for geometric design is easier to create wider lanes and larger turning radiuses, which might have benefits on the achievement of higher discharge rate, facilitating capacity. However, it also tends to cause relatively larger distance between the opposite stop-lines, resulting in longer cycle lengths. In addition, it is hard to fulfill the various phasing plans as well.

3.2 Traffic Signal Operations

Various control strategies and approaches, generally classified into adaptive and fixed-time control, are illustrated in both Japanese and German guidelines. In Germany, movement-based adaptive signal control is widely adopted at the key intersections and those with transport public. Meanwhile, fix-timed signal plans run at most of the other intersections. Movement-based control refers to such control pattern that the controller is capable of varying the amount of time it allocates to each movement rather than stage (in this case, it is called stage-based signal control that is commonly applied in Japan) based on the traffic demand. Therefore, in theory, movement-based control is more efficient in obtaining operational performance. Such

control policy requires essential lane configurations and equipments, e.g. detectors (for adaptive control) and distinct signal displays for each movement as well as exclusive left-turn lanes in some cases.

Due to comparably high intersection density and heavy pedestrian demand, the implementation of such flexible control policy (movement-based adaptive control) at signalized intersections in Japan, is apparently constrained as minimum pedestrian green time is longer and traffic signal coordination becomes more pronounced as well provided those traffic conditions. The fact, together with the considerations of cost and available technologies, stage-based signal programs and time-dependent selections of signal programs dominate current signal control at intersections in Japan. In the latter one, several patterns defined by the traffic volume and occupancy are selected during one day or a certain period according to detector data in time intervals, 2.5, 5 or 15 min. The other remarkable features relevant to signal operations in the two countries are briefly presented as follows.

(1) Phasing plan

Phasing plans are commonly figured out through combinational applications of the following control strategies in Germany: green time adjustment, phase swapping, request of demand phase and signal program formation. In most cases, multi-phase plans run at medium and large-size intersections and are variable in number of phases and green time for each movement based on detector data and control logic. Behind them, gap and occupancy detections perform as core technologies. The basic idea of gap adaptive control method is to switch off the green and turn on the amber if no vehicle passed the detector's position in a previously given time interval between 2 and 5 s, usually 3 s. In addition, the detectors, adjacent to the stop-lines, help the controller to assign a phase by detecting the existence of traffic. However, traditional two-phase (permitted right-turn only), three-phase (protected and permitted right-turn) and four-phase (protected right-turn only) signal programs play the major parts in the signal control at typical four-leg Japanese intersections. As an increasing trend to improve current signal operations in Japan, Universal Traffic Management Society of Japan (Japan UTMS) has launched several projects recently to investigate other efficient signal operation policies including the movement-based control.

(2) Cycle length

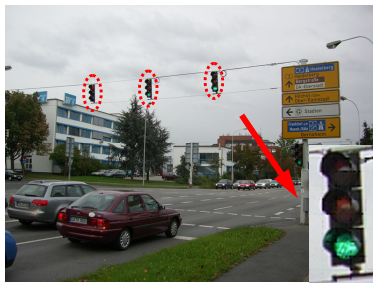
Webster's formula is referred for the calculations of optimal cycle lengths at isolated intersections in the two nations. In Japanese guidelines, 120 s is proposed to be appropriate maximum cycle length considering users patience to stand delay, and 180 s is the actual threshold value for applicable maximum cycle length. However, in practice, very long cycle lengths more than 120 s often exist at the intersections even without very high degrees of saturation in Japan. Those long cycle lengths mainly sourced from large intersection size, the pursue of capacities and preservation of minimum pedestrian green times. Another reason is that police hate phase switching due to safety reason because longer cycle lengths correspond to lower switch frequency given a certain time period. Theoretically, longer cycle lengths enable to improve intersection capacity, especially at saturated intersections, because of a lower

proportion of loss time in whole cycle. Therefore, it is useful at key intersections with high traffic loads. However, it will also likely bring superfluous delay to road users, which may induce red violations. In Germany, cycle lengths within 60-90 s are common candidates, not more than 120 s in any case. Shorter cycles raise phase switching frequency and consequently produce extra start-up lost time. In order to overcome the deficiency, red-amber signal display in one-second length before the green on-set has been adopted at German intersections to hint the drivers to start vehicles.

(3) Intergreen time

In German and Japanese guidelines, calculation of amber change interval bases upon the dilemma zone theory. However, it should be noted that the definition of clearance varied in the two countries before the publication of the revised edition of MTSC in July 2006. Stop-line was looked upon as a reference when judging if a vehicle clears or not before the end of all-red in older Japanese guideline. Clearance, however, is defined as the clearing vehicles completely pass the conflict points, not yet reach or pass stop-lines in German guideline. The different definitions result in a longer all-red clearance interval in Japan provided identical conditions. In German case, it is easily understood that, due to shorter all-red clearance time, it will be more dangerous if drivers enter into the intersections after the red on-set. Thus, for preventing red light violations, camera enforcement measure has been widely put into practice in Germany.

(4) Signals and its setting patterns



Germany

Japan

Figure 3 Signals and its setting patterns in Germany and Japan

In Germany, signal heads are basically 3-Vertical (Figure 3, Noyce et al. 2000) and set as lane-by-lane pattern, implying independent signal displays for each approach lane. Regarding positions, they are usually near-side, which eliminates visible signal changes for the drivers who are inside intersections if no assistant signals at the corners of intersection. It may lead to aggressive or conservative gap-acceptance behavior of left turners waiting inside intersections, as they cannot recognize when the green for opposite through traffic ends. Therefore, assistant signal heads exclusively for left-turners are sometimes set at the corners of intersection.

Signal heads in Japan, however, are traditionally 4-Cluster (Figure 3, Noyce et al. 2000) and set at both near-side and far-side neighboring stop-lines, giving indications for all the lanes at one approach. Accordingly, drivers are probably capable of keeping watch on the signal displays for all directions even those for the crossing directions when they intend to traverse intersections during intergreen time. It may induce red-light violations, principal cause of right-angle incidents, because drivers tend to be more confidently when choosing to pass instead of stop during all-red interval.

(5) Signal indications

The indications for drivers given by a traffic signal in Germany are normally displayed as green-amber-red-(red+amber)-green, where red and amber are displayed simultaneously in one-second length before the green on-set. It is for reminding the drivers that green will start very soon. Apparently, starting response time and start-up lost time can be saved in that case. Thereby, discharge flow rate can be improved. The sequence of green-amber-red-green is generally applied for vehicular traffic in Japan. Since amber or red allows are not defined in Japan, a right-turn green arrow is ordinarily displayed at the same time with a circular red indication (red ball) when releasing protected right-turn traffic in Japan.

As compared to Japanese policy, German policy for signal operations performs more flexible in accommodating the prevailing traffic, and possesses the ability to adjust signal program with the alterable traffic conditions. However, the conditions for implementing such policy, e.g. operation cost and the requirements on geometric design are more strict. It also produces relatively short cycle length and clearance time. In addition, obvious different characteristics in signals, its setting pattern and display sequence between the two countries exist as well.

4 Concluding Remarks and Future Works

This study presented German and Japanese experiences in signalized intersection design and operations. As discussed earlier, due to the differences in fundamental philosophies, local traffic conditions, available technologies and even cultural customs, the treatments of intersection design and operation elements in the two countries are distinguished, especially in practice. Those distinct treatments may influence traffic characteristics and user behavior. Accordingly, the intersections might perform differently in terms of mobility and safety. Toward deeper understanding on the impacts, the authors have as well done a series of analysis on traffic characteristics, driver behavior and intersection performance at several typical German and Japanese signalized intersections. Unfortunately, those analyses were not presented in this paper because of limited space. For that, readers can refer to Tang, K.S. and Nakamura, H. (2007).

The ongoing research project of developing Chinese Manual on At-grade Intersection Planning and Design provides a rare opportunity to improve intersection design and operations by absorbing and digesting international experiences as well as combining traffic conditions in China. Unlike Germany and Japan, extremely heavy pedestrian and bicycle demands usually prevail at Chinese intersections. Thus,

research on the impacts of intersection geometry and operation policy on pedestrian and cyclist behavior is particularly essential. On the other hand, discussions in the paper also concluded that developing a manual only even a perfect one could not resolve all the operational and safety issues at intersections because intersection design and operations chiefly relies on the road planning and design. It is scientific road classifications and function positions that allow intersection performance be explicitly and rationally oriented. Meanwhile, efforts on enhancing designers and operators' comprehension on the philosophy behind guidelines as well as cultivating their proper attitudes to intersection design and operations should also be emphasized in future because they directly connect practice with the guidelines or manuals.

References

Boltze, M. (2006). "Recent Developments in Germany related to Highway Capacity and Quality of Service." Proceedings of the 5th International Symposium on Highway Capacity and Quality of Service of the TRB.

FGSV – Forschungsgesellschaft fuer Strassen- und Verkehrswesen (1992). *Richtlinien fuer Lichtsignalanlagen (Guidelines for Traffic Signals)*, FGSV 321, FGSV-Verlag, Cologne.

FGSV – Forschungsgesellschaft fuer Strassen- und Verkehrswesen (2001). *HBS – Handbuch fuer die Bemessung von Strassenverkehrsanlagen (Highway Capacity Manual)*, FGSV 299, FGSV-Verlag, Cologne.

TRB – *Highway Capacity Manual*, National Research Council, Washington, D.C., 2000.

JSTE – Japan Society of Traffic Engineers (2006). *Revised Edition of Manual on Traffic Signal Control*, JSTE, Tokyo.

JSTE – Japan Society of Traffic Engineers (1994). *Manual on Traffic Signal Control*, JSTE, Tokyo.

JSTE – Japan Society of Traffic Engineers (2004). *Manual on At-grade Intersection Planning and Design*, JSTE, Tokyo.

JSTE – Japan Society of Traffic Engineers (2006). *Traffic Capacity Data Book*, JSTE, Tokyo.

Nakamura, H. and Oguchi, T. (2006). "An Overview on Capacity and Quality of Service Studies in Japan." Proceedings of the 5th International Symposium on Highway Capacity and Quality of Service of the TRB.

Noyce, D.A. (2000). "Traffic characteristics of protected/permitted left-turn signal displays."
Presented at 2000 Annual Meeting of the Transportation Research Board.

Tang, K.S. and Nakamura, H. (2007). "A Comparative Study on Traffic Characteristics and Driver Behavior at Signalized Intersections in Germany and Japan." To be published in the Journal of the 7th Eastern Asia Society for Transportation Studies, Vol.6, 2007.

ANALYSIS OF RAILWAY ORIENTATION IN THE FREIGHT EXPRESS SERVICE NETWORK

Shuai CHEN[1] Youwang SUN[2] Zhihong YIN[3]

[1] School of Traffic and Transportation Engineering, Tongji University, Shanghai, ST 201804; PH (86) 021-69589879; email: chenhandsome2008@yahoo.com.cn
[2] School of Traffic and Transportation Engineering, Tongji University, Shanghai, ST 201804; PH (86) 021-69589879; email: sunyouw@263.net
[3] School of Traffic and Transportation Engineering, Tongji University, Shanghai, ST 201804; PH (86) 021-69589879; email: realreal26@163.com

Abstract: Multiple types of transportation modes are required to cooperate through the network to fulfill the "door to door" freight express service. In the range of the entire freight express industry, highway, air and rail compete at diverse extents in the trunk corridor. Rail is capable of working better in the freight express industry with the sixth large-scale speed-up in the nationwide railway in China and the construction of the Beijing-Shanghai High-Speed Railway. This paper presents the railway orientation and its line operation mode in the Chinese freight express market through an analysis of network characteristics and Occident practices in the perspective of service network.
Key words: *service network; railway orientation; operation mode; freight express*

1 The Characteristics of Express Freight and Express Service

1.1 The characteristics of express freight

Besides letters, which are monopolized by the postal service, the main express freight includes office supplies and daily living necessities, fittings, electronic products, newspapers, fresh and live products, other high unit price commodities and so on. The key to express freight is promptness and reliability in delivery.

The development of the internet is changing the structure of the freight express industry. The transportation volume of pure information items, such as letters, not only personal letters, but also business letters, has been gradually declining. On the other hand, the rise of the e-commerce industry has opened up a new market for freight express. More and more consumer goods need long-distance transport support. The percentage of parcels and other small items has been growing rapidly. The average weight of single pieces in the worldwide freight express industry is 5 -6 kilograms, or 6-8 kilograms if the weight of the package is included.

1.2 The characteristic of express service

The full service process for freight express includes freight collection, transportation, delivery, service acknowledgment, information inquiries, and so on, and international clearance is needed in international freight express. Express enterprises finish the service in the commitment of time, usually including arrival on the same day, after 1 day, 2 days and 3-5 days. In general, the arrival after more than 2 days is called delay-service.

Express delivery service to be finished on the same day can only be carried out in a certain space area mainly by the highway transportation sector.. Express delivery service to be finished after 1 day is mainly done by air and highway transfer, and it is a global standard for transportation mode combinations of freight express. And a hub-spokes transfer transportation network is widely used. Express delivery service to be finished after more than 2 days (i.e delay-service) is done mainly by road transport.

In the perspective of transport mode, the railway will be an important force in the future development of the Chinese freight express industry, which is the focus of this paper .

Table 1 The service features of different types of express service

Type of Service	Express in the same day	Express after 1 day	Delay-delay service express	Postal services
Cost	High	High	Medium	Low
Distance	Within a certain space area	Unrestricted	Unrestricted	Unrestricted
Freight size	Unrestricted	Unrestricted	Unrestricted	Restricted
Service Time	In the same day	The next day	More than 2 days	Differences
Global standardization services	No	Yes	No	No
Delivery Method	Door to door	Door to door	Door to door	Differences

2 The Service Network Structure and Operation Mode of Freight Express

2.1 Significance of service network

The transportation network includes two levels: First, the infrastructure network, which is the material base for transport network. Second, the transport service management and organization, which is the performer of the transport business. As a part of the second level, the transportation service network is the network structure and operation mode of the transport service business management and organization.

There are three levels in the research of a transportation service network: (1) the service network of a transportation enterprise, the research objects of which usually include transportation enterprises transport network scope, node grade relations, the amount of conveyance, transport line, etc. (2) The service network of a single transportation mode industry, the research objects of which include the logical structure of the network, the hub operation pattern, the transport line operation pattern and the information platform of the industry. (3) The service network of integrated transportation, the research objects of which include the logical structure of the network, the hub operation pattern, the corridor operation pattern, integrated transportation information platform, the interest allocation model, and so on.

This paper aims to analyze the orientation of a railway in a freight express network, so it only covers part of the research objects in a transportation service network, mainly the logical structure of the service network in the industry, the corridor operation pattern, and transport line operation pattern which includes the service scope provided by the line, line rank and the way to achieve economies of scale and so on.

2.2 Service network structure of freight express

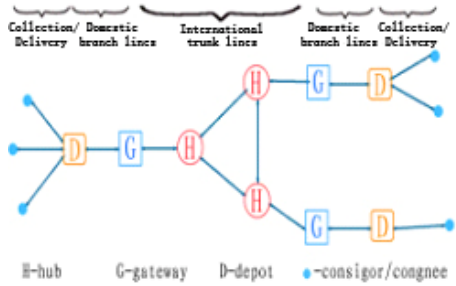


Figure 1 Service network structure of international freight express

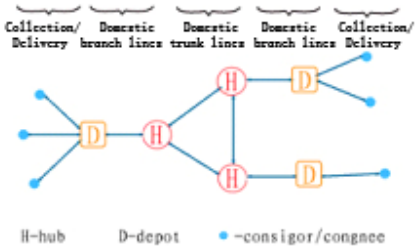


Figure 2 Service network structure of a domestic inter-city freight express

In the structure of a typical international freight express service network, there are 3 types of transportation nodes: D-Depot, G-Gateway, H-hub. Receiving from/Delivery for the terminal clients are completed in D nodes. G nodes, which collect express freight from D nodes, are national immigration and emigration nodes (such as the airport). H nodes are hub nodes of the highest level of big regions. The express freights from G nodes are collected and sorted in H nodes, then transferred to the other G nodes in the same big regions or H nodes in other big regions. For example, the H-hub nodes in Europe and North America of the FedEx Company are located in Paris and Memphis respectively. In the international freight express, transportation between G nodes and H nodes as well as between one H node and another H node are finished by air, while transportation between the inland area and airport is completed mostly by road.

The domestic freight express business doesn't need freight collection from immigration/emigration airports to hub nodes of a big region. So only D-Depot and H-hub nodes are included in a domestic freight express service network, and G-Gateway nodes are not included. The links among H nodes are trunk transport corridors of domestic freight express which are served by highway, air and rail.

3 Corridor Operation

Corridor Operation includes competition and collaboration (i.e. intermodal transportation) among different transportation modes, and competition is the main aspect.

3.1 Collaboration in corridor

National railway companies in Germany and France are considering running high-speed freight trains between Frankfurt and Paris, which is called the FEX plan. This plan pays more attention to coordination in the international freight express trunk

transportation corridor. It is prepared to carry out air-rail intermodal transportation in Koln Airport (the European hub of UPS Company) and Frankfurt Airport. The train will spend 6 hours running through Frankfurt, Koln, Brussels and Paris to target at the door-to-door one-day arrival service market. In the future, some other cities such as London, Zurich and Amsterdam will be included in this project. The air-rail intermodal transportation in this project will mainly serve the international freight express business, and the role of the railway is to connect G and H nodes in the international freight express business network.

Figure 3 FEX Plan of Germany and France

3.2 The competition of one-day arrival freight express business in corridor

Taking Beijing-Shanghai corridor as an example, the trunk link transportation of one-day- arrival freight express is mostly finished by air at present. After the completion of the construction of a high-speed railway between Beijing and Shanghai, the high-speed railway will have the capacity of providing the trunk link transportation service for one-day- arrival freight express. Then the market share will be competed by air and high-speed railway.

Table 2 The rail freight services categories in European Railway Express Study Report

Types of service	Maximum speed	Main types of conveyance
High-speed railway freight	>200km/h	Freight train transformed from the high-speed passenger train
Quasi-high-speed rail freight	140-200 km/h	Passenger train; Improved cargo train
Traditional rail freight	<120(140)km/h	Traditional cargo train

A high-speed rail freight train, which is represented by TGV Postal Train, is transformed from a passenger train. A quasi-high-speed rail freight train is developed from a traditional cargo train, which is represented by Swedish B-Postal Freight Trains and German container trains.

At 200km/h, which is the critical speed point between high-speed and quasi-high-speed rail, the train will spend 7 hours on the trunk line transportation between Beijing and Shanghai by rail. Therefore, it can meet the time requirement of the one-day (24 hours) arrival freight express business. What's more, the cost is much lower than that of airplane transport.

For long-distance freight express, safety/reliability and punctuality are becoming more and more important. The high-speed trains run according to the signals inside the driving cab, so they can complete the transport tasks in a timely manner without being influenced by bad weather such as big winds, rain, snow or fog.

A general trend is that the high-speed railway is entering the one-day arrival freight express market in the world. As the land areas of most countries in Europe are small, the length of each single high-speed railway in these countries is very short, less than 400KM. Furthermore, the administration system of a high-speed railway is different in each country. So it is difficult for the high-speed rail to run at its maximum speed. In order to make the high-speed rail enter the one-day arrival freight express market, cooperation has been carried out among European countries. Besides the FEX Plan of Germany and France mentioned above, the high-speed rail between Amsterdam and Milan will provide freight express service with mixed marshalling of freight trains and passenger trains, and the high-speed freight trains will run through Netherlands, Germany, Switzerland and Italy.

3.3 The competition of the delay- arrival freight express business in corridor

The trunk link transportation between Beijing and Shanghai for the delay- arrival freight express is competed between highway and rail.

There are 3 forms for freight express transportation by highway：whole-truck-load express; less-than-carload express; freight carried by buses. Generally, it is believed that the economic distance for a highway is 500KM, but the actual highway transportation distance in practice is much longer than this economic distance because the transportation capacity of rail is inadequate and unable to fully meet the needs of long-distance delay- arrival freight express.

The delay- arrival freight express by rail has three forms: a Baggage and Parcel Transportation Special Train; a Postal Items Transportation Special Train; Luggage

Compartments in a Passenger Train. The main problem facing rail freight express now is that the railway trunk lines are too busy, which results in the instability of service quality and uncertainty of arrival time. From the perspective of active participation in market competition, the railway freight express companies in China have already participated in market competition acting as business operators, while in the European market the rail transportation companies only act as one of the carriers in the process of freight express transport and do not face the freight express market independently.

In order to develop the rail freight express industry, the most basic thing is to construct new railway lines to enhance the transportation capacity of rail, so that the service quality of the rail freight express will be improved. In the near future, the primary task for rail delay- arrival freight express is to control the cost. Take the price of 10KG freight "door to door" express service between Beijing and Shanghai as an example. the freight is delivered in one day by air and the price is about 100RMB; delivered in 2-3 days by highway, the price is about 50RMB; while China Rail Express Co. Ltd can only promise to deliver within 3 days by railwith the price of 35RMB.

Compared with highway freight express, rail freight express has a certain advantage in price, but the price advantage is not so remarkable. What's more, the rail freight express can not confirm the exact number of service days as a highway can, such as the example mentioned above: that China Rail Express Co. Ltd can only promise to deliver within 3 days. In conclusion, the key to rail freight express competition in the market is to control cost while the quality of services is assured.

3.3 The orientation of railway freight express in the Beijing-Shanghai corridor

In recent years, in the analysis of the European railway freight express, the orientation of railway freight express has often been concluded as "More rapid than highway, cheaper than air". In our opinion, this orientation is not quite accurate. We deem that there should be different orientations according to different markets.

In the one-day(24 hours) arrival freight express market, the main competitor of the railway is the air transport, and the customer's primary concern is the time for service that the freight express operator spends. So compared with air transportation, the railway should put more emphasis on its speed to meet the customer's needs and on ensuring the reliability of the freight express service process, but not on service price.

In the two-day (or more than two-day) arrival freight express market, the major competitor of the railway is the highway. Speed is no longer the critical and primary concern. Customers are more likely to consider the price of freight express services.

So compared with the highway, the railway should focus on the low-cost advantage. For example, in Germany, the delay-arrival freight express by rail has emphasized speed much more than cost. The Inter Cargo Express (ICGE) trains began to run in 1991, and the speed of the ICGE trains running between Bremen and Stuttgart was 160 km/h at first, but the profit of these trains was poor. As a result, the speed was adjusted to 120km/h and the focus was shifted from speed to cost control.

In conclusion, the orientation of the railway in the freight express in the Beijing-Shanghai corridor is summarized as the following two points:

(1) The high-speed railway between Beijing-Shanghai joins the freight express service network to compete with air transportation for the one-day (24 hours) arrival freight express market;

(2) The conventional electrified railway should control cost to compete with highway for the two-day (or more than two-day) arrival freight express market.

4 The Railway Line Operation

In order to adapt the railway orientation to the freight express service network, the suitable line operation mode is needed, including the service content provided by the line, line rank and the means to achieve economies of scale.

4.1 The practice of railway line operation in Europe and the United States

4.1.1 The service content provided by the line

(1) The passenger special railway line opening to the freight express service business:

According to its service target, the high-speed railway line can be classified into "Passenger Special Railway Line" and "Railway Line Serving for both Passenger and Freight". In France, the passenger special railway line started to run postal trains in the1980s and the parcel express trains in 1997. The freight express train TGV Duplex has the merits of large capacity and high speed, and can carry 10 small containers. The French National Railway has benefited much from the freight express service. Therefore, in order to use high-speed railway freight express more extensively, in recent years French National Railway has started the "TGV Fret Plan" to study the economic and technical problems of the high-speed railway freight express.

The common characteristics of freight transportation by high-speed railway in Europe are as follows: Freight trains run on high-speed lines at night but not daytime to make full use of the transportation capacity of thecorridor; in consideration of the maintenance of the railway line,, the large-load freight trains are not permitted to run on the passenger special railway line, while the small-load express trains are gradually permitted.

(2) Diverse trains marshalling forms

In both European transnational railway freight express and American railway freight express, mixed marshalling of passenger and freight trains accounts for a high proportion. On some railway lines of the American national rail company—Amtrak (which has dropped out from the express market due to policy restrictions) - more than 40% of the total revenue comes from freight express. In many mixed passenger-freight trains, the number of freight trains is far more than passenger trains. For example, trains between New York and Chicago are partly formed by 24 freight trains and 5 or 6 passenger trains.

4.1.2 Railway line rank

The hub-spokes transfer transportation network is the typical network structure of air freight express. The nodes in the network and the different rank lines set in the network can support the operation of big hubs. But because of the speed limitation of the rail, the railway lines do not have the obvious diversity of different rank lines and the main transport of rail is point-to-point transportation, so the rail lines can not support the multi-rank-line network structure. In Europe, the rail lines can act as the branch lines of air freight express in the international freight express network.

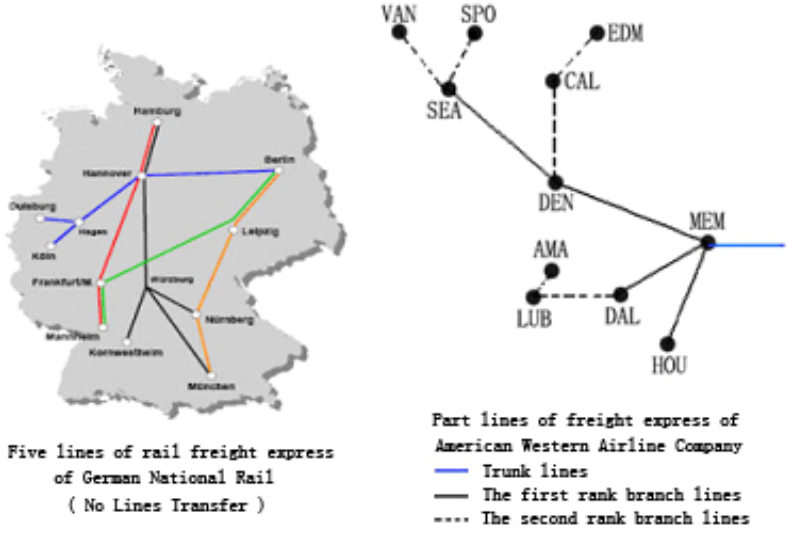


Figure 4 The feature differences between rail and air line transfer

4.1.3 The way to achieve economies of scale

The economies of scale that relate to rail include the following: the economy of region-size covered by transportation network, economy of transportation service

scope, economy of transport density on lines, economy of station and yard operation capacity, economy of transport capacity of vehicles, and economy of fleet size.

The economies of scale mentioned above that relate directly to the line operation are as follows: the economy of region-size covered by transportation network, economy of transportation service scope, and economy of transport density on lines.

The economy of transport density on lines is the phenomenon that the average transport cost is reduced gradually because of the transport density increase on a specific line. It can promote the realization of the economy of transport density on lines if the high-speed railway starts the freight express service business.

The economy of transportation service scope is that the average cost of providing several transportation services together is much lower than the average cost of providing only one transportation service. The mixed marshalling of freight and passenger trains that is widely used in Europe and United States can promote the freight express service to realize the economy of transportation service scope.

The economy of region-size covered by a transportation network is the phenomenon that if the transport density on the transportation network remains unchanged, the expansion of region size covered by the transportation network will result in the same proportion increase of total transport service output, so that the average cost of transport service will be reduced gradually. American railway passenger volume has been declining in recent years, but during the period when it operated the freight express service business, the railway company Amtrak set up the national freight express network so that the big freight express stations covered the whole country and the region size covered by its transport network was expanded. As a result, the freight express service operator can optimize the transport routing more reasonably and use the transport vehicles more efficiently. The economy of region-size covered by transportation network has promoted the low cost of rail freight express in America, so that the market share of rail freight express is much higher. In America, 29% of the freight transport volume of the UPS Company is transported by rail while the corresponding number is only 4% in Germany.

Figure 5 The freight express transport network of Amtrak company in 2000

4.2 The suitable railway line operation mode for the China freight express industry

As for the service content provided by the line, in order to fit theChina railway orientation into an integrated express service network, the passenger special railway line should first be opened to freight express service, which is the basic condition for the Beijing-Shanghai high-speed railway competing for the one-day arrival freight express market with air. In term of trains marshalling, the moderate mixed marshalling of passenger and freight trains of a high-speed railway can enhance the freight express frequency, and meanwhile increase the train departure density in non-peak hours, so that the win-win of passenger transport and freight transport will be gained.

As for the railway line rank, the point-to-point direct transportation is dominant in railway freight express in China, and the transfer function of railway lines is not significant. There does not exist the rank of trunk line and branch line in the railway transportation service network for freight express. However, in the integrated transportation freight express service network, with the gradual establishment of Shanghai Pudong airport and Guangzhou Baiyun airport as H hubs key positions for worldwide freight express, the railway line has the chance to act as domestic branch lines to link the international trunk air lines.

The key point for a conventional railway to compete against the highway in delay- arrival express market lies in cutting down costs through various scale-economy means. With regard to line operation, the most important is to give full

play to effects of region-size economy, widening the area that the railway freight express service network can cover by amending the national railway network, and consequently creating favorable conditions for optimization of transportation routing and efficient use of transport vehicles.

5 Conclusions

This paper has analyzed part of research objects in a service network research frame, logical structure of a freight express integrated transportation service network, a corridor operation mode as well as a line operation mode of a railway service network, and has explored the suitable orientation of railway freight express development in China. In order to bring the power of service network into full play, all steps in the transportation process should be integrated logically with the help of further comprehensive analysis on element characters, interrelationship, function performance of each level of service network and corresponding measures on controlling and regulating by government.

References

Rong, Chaohe and Dong, Yanhua (2004). "China Railway Express' Success and Its Revelation for Railway Reform" .*Communication and Transportation Systems Engineering and Information.,* Vol 4(1), 3-11.

Rong, Chaohe (2001). "On the Economies of Scale and Scope in Transport Network." *China Railway Science.,* Vol 22(4), 97-103.

Zhong, Shan and Wang, Shouyang (2003). "A Hub-and-Spoke Structure and Its Applications in the Air Logistics". *Communication and Transportation Systems Engineering and Information*., Vol 3(3), 25-36.

An Improved Ant Colony Algorithm for UE Traffic Assignment Problem

Ruijun QIAO[1], Xiaoning ZHU[2], Yanhu LIU[3]

[1,2,3]School of Traffic and Transportation, Beijing Jiaotong University，Beijing City, P.R. China; PH 86(10)5146-7610; Email:qiaoruijun2006@sohu.com

Abstract： As the last process of traditional "four processes" management method, traffic assignment plays an important role both in theory and practice. UE traffic assignment is a basic problem in the field of urban traffic equilibrium assignment and many studies focus on it. Ant colony algorithm is a new simulating evolution algorithm which has solved many typical combinatorial optimization problems. This paper tries to solve UE traffic assignment problem with ant colony algorithm and improve the goal function value and the formula of transition probability according to the character of UE traffic assignment. The algorithm may get brush-fire optimum result and stop earlier in large scale problems. To solve this problem, this paper optimizes the formula of trail pheromone increment and use max-min ant system to limit the concentration of path pheromone into a scope. Finally, a numerical example is given to prove the feasibility and validity of this algorithm; the result is also compared with the one gotten by traditional algorithm.

Key words: *Ant colony algorithm, UE traffic assignment, max-min ant system*

1 Introduction

User Equilibrium traffic assignment is a basic problem in the field of urban traffic equilibrium assignment. Scholars have done much research on this problem and its solutions. Among all the solutions, the method of successive averages is a typical one. However, if the algorithm cannot be convergent strictly, the result will not always be accurate (Liu and Pu, 2005); furthermore, the iteration times needed to find optimum result is excessive. Some modern intelligent algorithms have been introduced into this field. Liu and Pu (2005) used genetic algorithms to solve stochastic user equilibrium problem and Li and Zhong (2001) used genetic algorithms to solve combinatorial equilibrium problem. These algorithms, which have no requirement of continuity and differentiability of the problem to be solved, can receive good effect in practice.

Ant colony is a heuristic algorithm from nature. It is all-purpose and can discover the best result under very difficult condition (Hou and Jiang, 2001). Now, ant colony algorithm has been widely used to solve TSP problem, quadratic assignment problem and job-shop scheduling problem. Hou and Jiang (2001) demonstrated the rationality and feasibility of using ant colony algorithm to solve UE traffic assignment problem. Xu and Wang (2004) used ant colony algorithm to solve traffic assignment problem of fixed demand. On the foundation of consulting literatures above, and combining the character of UE traffic assignment, the paper improves ant colony algorithm in order

to get better result.

2 UE Traffic Assignment Model

Given a transport network $G(N,A)$,where A and N are the sets of links and nods respectively, and each directed link $a \in A$ is associated with a travel time $t_a(x_a)$ as an function of link flow x_a. W is the set of origin destination pairs, and for each pair $r\text{-}s \in W$ .There is a given traffic demand q_{rs}, and it is deterministic in this paper. The user equilibrium assignment problem is stated as follows (Huang, 1994).

$$\text{minimize } z(x) = \sum_a \int_0^{x_a} t_a(x)dx \tag{1}$$

$$\text{subject to } \sum_k f_k^{rs} = q_{rs} \qquad \forall r,s$$

$$f_k^{rs} \geq 0 \qquad \forall k,r,s$$

$$x_a = \sum_r \sum_s \sum_k f_k^{rs} \delta_{a,k}^{r,s} \quad \forall a$$

Where: x_a =flow on link a, $a \in A$; t_a=travel time on link a, $a \in A, t_a = t_a(x)$; K_{rs}= set of paths within OD pair r-s; f_k^{rs} =flow on path k within OD pair r-s, $k \in K_{rs}$; q_{rs} = origin-destination demand between r and s; $\delta_{a,k}^{r,s}$=1 if link a belongs to path k; otherwise, $\delta_{a,k}^{r,s} = 0$.

It has been proved by Huang (1994) that the result of model (1) is also the one of user equilibrium. When network is in equilibrium state, the impedance of every path is equal. The impedance of path k is expressed by T_k.

$$T_k = \sum_a \lambda_a t_a(x) \quad k \in K_{rs}, \quad a \in \mathrm{A} \tag{2}$$

Where: $\lambda_a = \begin{cases} 1 & \textit{if link } a \textit{ is used by path } k \\ 0 & \textit{otherwise} \end{cases}$

T_k is just the goal function when use ant colony algorithm to calculate UE traffic assignment problem.

3 Ant Colony Algorithm and Its Improvement

3.1 The main variable of ant colony algorithm

The basic variable and constant are as follows (Xu and Wang, 2004).

m = amount of ants; η_{ij} = visibility of link (i,j); τ_{ij} = trail intensity of link (i,j)

$\Delta\tau_{ij}^{s}$ = increment of pheromone on link (i,j) caused by ant s in the process of current iteration

$$\Delta\tau_{ij}^{s}=\begin{cases}\dfrac{Q}{T_S} & \text{if link (i, j)} \in \text{tour done by ant } s\\ 0 & \text{otherwise}\end{cases} \tag{3}$$

Where: Q=a constant that incarnate the pheromone amount ants leave ($1\le Q\le 10000$); T_s= goal function value of ant s.

p_{ij}^{s} = transition probability of ant s

$$p_{ij}^{s}=\frac{\tau_{ij}^{\alpha}\eta_{ij}^{\beta}}{\sum_{j\in J_s(i)}\tau_{ij}^{\alpha}\eta_{ij}^{\beta}} \tag{4}$$

Where: α =relative importance of trail intensity ($0\le\alpha\le 5$); β =relative importance of visibility ($1\le\beta\le 5$); $J_s(i)$ =set of nods that ant s can choose

The updating equation of trail intensity is as follows.

$$\tau_{ij}^{new}=\rho\cdot\tau_{ij}^{old}+\sum_{s}\Delta\tau_{ij}^{s} \tag{5}$$

Where: ρ =permanence coefficient of trail intensity ($0\le\rho\le 5$); 1- ρ is attenuation coefficient

3.2 Improvement of ant colony algorithm

Connecting with the character of UE traffic assignment, some improvement has been done.

The visibility of link (i,j) is always calculated by: $\eta_{ij}=1/d_{ij}$. And d_{ij} is the length of link(i，j). For UE traffic assignment, link impedance $t_a(x)$ is the only factor that influences driver's choice about path. The BPR function of time cost in link is as follows.

$$t_a(x_a)=t_a(0)\left(1+e\left[\frac{x_a}{c_a}\right]^{f}\right) \tag{6}$$

Where: c_a=traffic capability of link a; $t_a(0)$ =travel time on link a, when there is no traffic; e、f = coefficients of road conditions.

When $t_a(0)$ is given, only link flow x_a can influence $t_a(x_a)$. So the visibility of link η_{ij} is not taken into account in this paper. Transition probability of ant s becomes:

$$p_{ij}^{s}=\frac{\tau_{ij}}{\sum_{j\in J_s(i)}\tau_{ij}} \tag{7}$$

When calculating the increment of pheromone on link (i,j), Xu and Wang (2004) only calculated the one that used by the optimum path. For other links, they are considered to be zero. Along with the increase of iteration times, the trail intensity of

links not used by the optimum path will decrease gradually. If there is not a minimum limit of the intensity, it can even decrease to zero. Then, the algorithm may get brush-fire optimum result and stop earlier. In this paper, for every link where ant passed, the increment of pheromone must be calculated and should be worked out by the goal function value corresponding with the exact ant, not the optimum one.

Introduce max-min ant system to limit the concentration of path pheromone into a scope $[\tau_{\min}, \tau_{\max}]$. The concentration value that beyond the scope will be defined to $\tau_{\min}$ or $\tau_{\max}$. Then, it can be effetely avoided that arithmetic cannot continue because concentration of one path is far more than others' and all ants centralize on one path.

$$\tau_{\max} = x_a \cdot \frac{1}{1-\rho} \cdot \frac{1}{Z_{opt}} \tag{8}$$

$$\tau_{\min} = \frac{\tau_{\max}}{5} \tag{9}$$

Where: Z_{opt} =current optimum goal function value

$$\tau_{ij}^{new} = \begin{cases} \tau_{\max} & if\ \tau_{ij}^{new} \geq \tau_{\max} \\ \tau_{\min} & if\ \tau_{ij}^{new} \leq \tau_{\min} \end{cases} \tag{10}$$

Initial value of trail intensity should be ascertained suitably by $\tau_{\max}$ and $\tau_{\min}$. This paper first distributes traffic by the method of "All-or-Nothing Assignment Method"; then calculates the goal function value T_0 of one path randomly. Initial value of trail intensity is calculated by

$$\tau_0 = \frac{2\tau_{\max}}{5} = \frac{2}{5(1-\rho)T_0} \tag{11}$$

4 Procedures of ant colony algorithm

The procedures using ant colony algorithm to solve UE traffic assignment can be summarized as follows.

Step1 Set all parameters nc (iteration times), m, Q, ρ and initialize pheromone trails, $\tau_0 = 2/5(1-\rho)T_0$, $\Delta\tau_{ij}$=0.

Step2 From origin, calculate transition probability p_{ij}^s based on Eq. (4) one by one; when doing the fist iteration, the transition probabilities from the same nod to any one that ant can choose are equal. According to p_{ij}^s, move ant *s*(s=1, 2…m) to the next nod *j*, until all ants finish a tour.

Step3 Calculate goal function value T_s (s=1, 2…m) based on Eq. (2).

Step4 Update trail intensity of every link based on Eq. (3), Eq. (5), and Eq. (10).

Step5 for every link (i，j), do $\Delta\tau_{ij} \leftarrow 0$， nc←nc+1.

Step6 if nc reaches the times scheduled or the error has been in allowable scope, arithmetic stops; otherwise, turn to Step2.

Step7 print current optimum value.

5 Numerical Example

The network, shown in Figure 1, consists of 5 links and 4 nods. The standard BPR function is adopted. And correlative parameters of links are listed in Table 1.

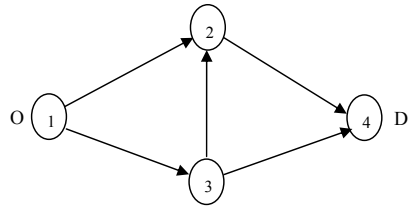


Figure 1. Network

Table 1. Correlative parameters of links

Link	$t_a(0)$	e	f	c_a
1—2	4.0	0.6	1.0	500
1—3	3.0	0.5	1.0	500
3—2	0.5	0.4	1.0	500
2—4	2.0	0.5	1.0	500
3—4	2.5	0.6	1.0	500

It is presumed that, the origin destination demand $q_{1,4}$=600, ρ=0.8, Q=1. Calculate with ant colony algorithm and method of successive averages separately. The flow of every link calculated in different iteration times is in Table 2.

Table 2. Link flow

Link	Iteration times									
	50		100		200		500		1000	
	MSA	ACA	MSA	ACA	MSA	ACA	MSA	ACA	MSA	ACA
1—2	180	191.5	174	178.2	177	175.6	175.2	175.4	175.2	175.4
1—3	420	408.5	426	421.8	423	424.4	424.8	424.6	424.8	424.6
2—4	348	360.3	348	348.3	348	346.2	346.8	346.3	346.2	346.3
3—2	168	168.8	174	170.1	171	170.7	171.6	170.9	171	170.9
3—4	252	239.7	252	251.7	252	253.8	253.2	253.7	253.8	253.7

The impedance of every path calculated by ant colony arithmetic in different iteration times is in table 3.

Table 3. Path impedance

Path	Iteration times				
	50	100	200	500	1000
1→3→4	7.4445	7.5205	7.5345	7.5347	7.5347
1→2→4	7.6398	7.5520	7.5353	7.5347	7.5347
1→3→2→4	7.5137	7.5300	7.5340	7.5347	7.5347

It can be seen from table 2 that it only needs 200 iteration times to get good value by ant colony arithmetic for this problem; but the times is 500 by the method of successive averages. Compared with MSA, ACA has higher efficiency. Table 3 shows that the impedance of every path tends to be equal along with the increase of iteration times, and it is content with the UE condition. So it is reasonable and feasible to use ant colony arithmetic to solve UE traffic assignment.

6 Conclusions

In this paper, traditional ant colony arithmetic is improved and used to solve UE traffic assignment. The improvement and main conclusions are as follows.

Take path impedance as the goal function value for ant colony arithmetic.

Simplify the formula of transition probability. For UE traffic assignment problem, it is unnecessary to take visibility of link into account.

Improve the formula of trail pheromone increment. For every link where ant passed, the increment of pheromone must be calculated and should be worked out by the goal function value corresponding with the exact ant, not the optimum one.

In order to avoid algorithm getting brush-fire optimum result and stopping earlier, introduce the max-min ant system to limit the concentration of path pheromone into a scope. Upper limit and lower limit of the scope are also given.

The rationality and validity of using ant colony arithmetic to solving UE traffic assignment are proved by a numerical example.

In ant colony arithmetic, the enactment of some parameters, such as ρ, should be studied further. In this paper, ρ is simply defined to be a constant. If ρ can be connected with η and change adaptively, better result can be gotten by the arithmetic.

References

Bullnheimer, B. Hartl R F, and Strauss C. (1997). "A New Rank Based Version of The Ant System-A Computational Study." Central European J Ope r Res Econom Jul., 25-38.

Dorigo, M, and Cambardella L M. (1996). "A Ant System: a Cooperative Learning Approach to The Traveling Salesman Problem." IEEE Trans on Evolutionary Computation. 1(1),53-66.

Dorigo, M. Maniezzo V, and Colorni A. (1996). "The Ant System: Optimization by a colony of cooperating agents." IEEE Trans on System, Man, and Cybernetics-Part B. 26(1), 29-41.

Hou, Liwen. (2001). "Method of Traffic Allocation Based on Ant Algorithm and Its Application." Journal of Shanghai Jiaotong University. Jun., 930-933.

Huang, Haijun. (1994). "Urban Transportation Network Equilibrium Analysis: Theory and Practice." People's Traffic Publishing Company of China,Beijing.

Huang, Xiyue. (2005). "Theory and Application of Modern Intelligent Algorithms." Science Publishing Company of China, Beijing.

Li, Zhichun, and Zhong Chun-ping (2001). "Integrated Equilibrium Assignment Model in Urban Traffic Network and Algorithm." Journal of Chongqing Jiaotong University. Mar., 65-67.

Liu, Haixu, and Pu Yun.(2005). "Genetic Algorithm for Stochastic User Equilibrium Assignment Model." World Science Research and Development. Feb., 68-70.

Xu, Xunqian, and Wang ya-ping. (2004). "Solving Traffic Equilibrium Assignment Problem of Fixed Traffic Demand with Ant Algorithm." Journal of Nantong Institute of Technology. Jun., 24-27.

Transit Priority Oriented Signalised Junctions—Design and Realisation in Geneva, Switzerland

Jaubert Jianan TAN

ITRA Ingénierie du trafic SA, Geneva, Switzerland, email: Jaubert.tan@itra.ch
Itsignals traffic science & technology Ltd, Xian, China

Abstract: Public transport offers alternative modes of transport to the private motor car. It can relieve traffic congestion and reduce air pollution from road transport. For these reasons, the idea of development of advanced public transport is well accepted in Switzerland. Each year from 2004, there is a new tramway built in Geneva. Extension of tramway will be continued in the coming years. To make public transport more reliable (less delay and timetable respecting), the transit priority strategy (such as transit lane concept and transit priority treatment measures) is always applied at signalised junctions. A signalised junction where the transit movement may have favourable green (for example, green will be given to a transit movement when it arrives at junction) is a transit priority oriented signalised junction. It is a type of intelligent signalised junction.

Based on more than ten year's experiences over 70 intelligent signalised junctions' design and realization in Geneva, the author presents in this paper the main steps of a transit priority oriented signalised junction's design and its realization. The following main steps are discussed: junction control strategy study; determination of control equipment; junction detail signal design; controller programming and test, analyse of recording data and evaluation of the effectiveness of the transit priority signal control. The development of useful tools for intelligent signalised junction design and a new signal group oriented traffic controller are also described. Some further developments and researches are suggested.

Key words: *intelligent signalised junction, signalised junction design, junction control strategy, transit priority signal control, signal group-oriented signal control*

1 Introduction and a new approach in signal control projects

The primary purpose of a traffic signal is to assign the right-of-way to vehicles at intersection to prevent traffic collision. Nowadays, signalised junction has become a very powerful instrument in the traffic management to realise control objects predefined by traffic managers. Transit priority oriented signal control is one excellent

example to use signal control to achieve the control object for promoting public transport.

Transit priority oriented signalised junction is a signalised junction where transit movements (with or without transit lane) have favourable green (green will be given to a transit movement when it arrives at junction).

To make a transit priority oriented signal control more efficiency, a new approach in signal control is used: signal group-oriented signal control method. It is a signal control using signals as its control elements. It is quite different from the traditional "the phase-oriented signal control method".

The "phase-oriented signal control method" is the signal control method using the "phases" as its control elements. A phase is consists of one or more compatible signals which have the green in the same time. The number of phases, the duration of each phase, sequence of phases and transition between two phases should be determined in signal designs. In this case, the variation of duration of phases is the main way to take account of real traffic situation in an actuated signal control. Therefore the extension or reduction of the green length of a phase means extension or reduction of the green length of all the signals defined in the phase. En consequence, if a part of signals need more green based on information from detectors and if the green length of the phase is increased, then the green lengths for the signals that do not need more green are also increased. Hence, some greens are given without traffic. If the signal timing plan in the signal group-oriented signal control is translated to the corresponding signal timing plan in the phase-oriented signal control, it can be observed that the green length of certain signals in the phase-oriented signal control is less than that in the signal group-oriented signal control (fig.1).

On the other hand, if we need even one second more for one signal in a phase for a special case, a new phase should be defined. En consequence, to realise relatively complex control logic, the number of phases should be very high, programming become very complicated and sometimes the limit of phase number of controller may be passed.

These are the main weakness of phase-oriented signal control method. Signalised junction design with this control method sometimes may give very bad image of traffic signal control. The signal group oriented signal control method has no these weakness. There is no limitation of the number of combination of compatible traffic signals. The transition between two conflict signals is determined by their inter-green that can be calculated precisely. As there are these advantages over the phase-oriented signal control method, the signal group oriented signal control method has been used in most signalised junction projects in Geneva, especially in transit priority oriented signalised junction projects.

A transit priority oriented signalised junction design and its realization consists of following main steps as shown in fig.2: (1) Junction control strategy study, (2) Determination of control equipment, (3) Junction detail signal design, (4) Controller

programming and test, (5) Real operation and verification and (6) Analysis of recording data and evaluation.

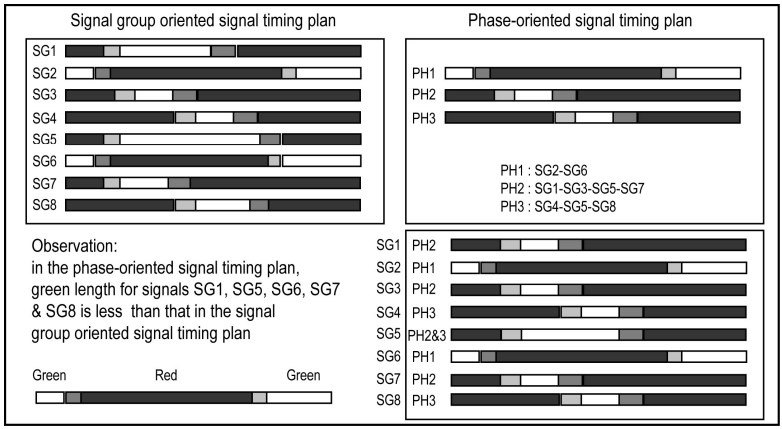


Figure 1. Green length of signals in the signal group-oriented signal control and in the phase-oriented signal control

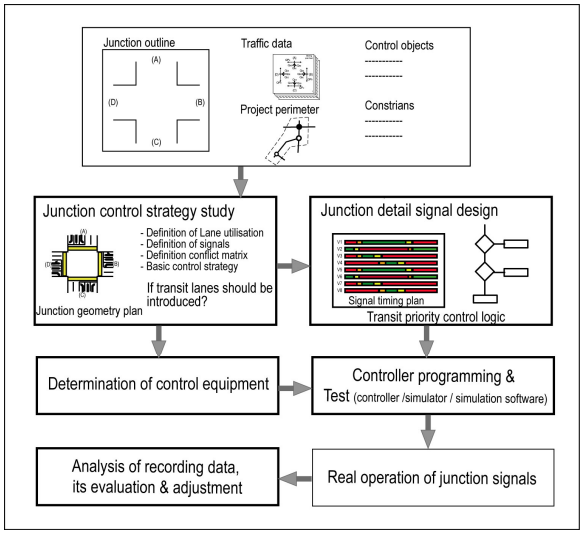


Figure 2. Realisation of a transit priority oriented signalised junction

2 Junction control strategy study

In the step of junction control strategy study, control objects and constrains should be defined first, where the main traffic control object is transit priority. Evidently, the higher level of transit priority, the lower is transit delay. But the transit priority has a cost (increase delays) to traffic movements who are conflict with transit movements. The higher level of transit priority, the higher cost (traffic queue and delays) is to these traffic movements. It is possible that a badly-dosed transit priority level in the signal control at a junction may lead to traffic congestion at neighbouring traffic network. Generally, the realistic control object is to maximise transit priority with acceptable cost to other traffic movements.

Based on control objects, constrains and traffic data for the design purpose, a junction geometry design alternative including following elements should be developed:

- Pedestrian crossings
- Entry lanes and exit lanes for each junction branch
- The use of these entry lanes (car, cyclists, bus, tram or mixture)
- The destination of entry lanes
- The definition of signals for each entry lanes, pedestrian crossings and conflict warning
- The compatibility of signals (conflict matrix)

Here if an entry lane only used by bus or tram is a transit lane.

Evidently, there exist a lot of alternatives. For example, an alternative with a bus lane may be developed and another alternative without bus lane may also be developed. This is the way to evaluate the effectiveness of a bus lane.

To evaluate the transit priority performance of each alternative, transit greenable length, transit performance ratio and junction capacity used in percentage1 should be estimated by applying designing traffic data.

Transit greenable length (GTmax in sec/cycle) is defined as the maximum green length may be distributed to a transit movement in one cycle without resulting traffic saturation at the junction. The transit performance ratio is defined as the ratio of maximum green length to the cycle length (for evaluation purpose only, usually 90") : GTp = GTmax / CL, where CL is cycle length (in sec). Junction capacity used is defined as the percentage of cycle time should be used to pass arrived designing traffic volume3,4. It is evident that a higher transit greenable length leads to a higher transit performance ratio. Mean transit delay is defined as TCd, in second.

It can be easily proved that a higher transit performance ratio leads to a less mean transit delay1. Therefore a higher transit greenable length leads to a less mean transit delay. So the transit greenable length is the most important index. The alternative with

the highest transit greenable length is normally a good alternative because it has the highest greenable length in one cycle. Junction capacity used in percentage is also very important index to evaluate the junction alternative. The alternative with the lowest junction capacity used is normally a good alternative, because the reserve of capacity can be attributed for transit movements.

If the junction is saturated and the transit movement is saturated too (usually the case of transit mixed with private traffic), then the transit delay increases rapidly. This is the worst thing to the transit and it should be avoided.

Suppose there is a signalised junction, and two alternatives of the junction geometry design are proposed: Alternative A is one without transit lane; Alternative B is one with a transit lane.

The bus using the transit lane usually always passes the junction within one cycle. So the maximum mean delay for transit is less than the half of signal cycle length. While in the case without transit lane, the maximum mean delay for transit may be more than a signal cycle length (especially at over-saturated situation). Therefore, if transit mixed with private traffic with high delay, the transit lane concept is possibly a good solution to protect transit from traffic queues, and hence to decrease maximum transit delay.

It can be proved that the transit lane may decrease junction capacity and increase the junction capacity used1.

To evaluate the introduction of a bus lane, transit greenable length (GTmax) and junction capacity used (CU) for two alternatives of the same junction are estimated. Suppose the transit greenable length and the junction capacity used for two alternatives are GTmaxA, CUA (without bus lane) and GTmaxB, CUB (with bus lane). It can be proved that CUA<=CUB.

If CUB <100% then introduction of bus lane will not leads to junction congestion. Furthermore, if GTmaxB >= GTmaxA then bus lane alternative is certainly better than other one. If GTmaxB << GTmaxA, in this case, the bus lane alternative have no advantage over the alternative without bus lane

If CUB >100% then introduction of bus lane will lead to junction over-saturation. In this case, some traffic measures (such as signal control with traffic queue treatment) should be taken so that the queue will not be formed at adjacent junctions. Similarly, if GTmaxB >= GTmaxA then bus lane alternative is better than other one. If GTmaxB << GTmaxA, the bus lane alternative have no advantage over the alternative without bus lane

In order to limit the impact of transit lane on junction capacity, following type of 'partial transit lane' may be used:

-transit lane disappeared at X meter before junction;

-transit lane starts at Y meter before junction;

-transit lane mixed with a traffic movement having a low traffic volume.

Transit lane is often used where a traffic entry control is needed to protect transit from traffic queues. As at this type of junction, decrease junction capacity is needed and make the entry control in a more nature way.

In this step, the following basic control strategies should be determined for the alternative selected:

- Coordination between signals?

  a) If signals should be coordinated with some other signals in this junction

  If the junction is consists of more than two sub-junctions (controlled by one controller), the signal coordination between this two sub-junctions should be studied. Usually, if the distance is small, the signals for the movements passing two sub-junctions should be coordinated.

  b) If signals should be coordinated with some signal in adjacent junctions

  If there is a signalised junction near-by, the signal coordination between these two junctions should be studied. Usually, if the distance is small, the signals for the movements passing these two junctions should be coordinated. As shown in fig4, for the project of new tramway Acacia, there are 4 signalised junctions GE208, GE76, GE204 and GE179 at sector "Etoile". The distances between adjacent junctions are very small (<200 m), and signal coordination for non transit movements are necessary in order to avoid traffic congestion. Furthermore signals of transit movements should be coordinated to minimise transit delay.

  If there is no a signalised junction near-by or signalised junction is far away, signal coordination has no sense.

- If signals should be equipped with detection system?

  For all traffic movements, the probability that there is no traffic arriving in one cycle can be calculated based their traffic volume during different periods. The calculated results may be used to determine if signals should be equipped with detection system.

  For the of detection of non transit movements, there are detectors with following functions, green requirement, green expansion, queue detection, traffic counting and speed detection.

  For signals of transit movements, the special detection system who can know bus arriving information should be used (see next section).

- How does signal control work if its detection system is out of work?

  Usually, several fixed-time signal timing plans for different period should be developed and will be used in case detection system is out of work. For example, fixed-time signal timing plans: "equilibre" (daily off peak period), "nuit" (night), "sortie" (evening peak hours) and "entrée" (morning peak hours) are usually developed.

-Actuated control logic such as transit priority basic control logic and queue reaction control logic will be proposed.

- Conflict justification between certain signals and using flashing signal to show the conflict

Usually, signals that share green are not conflicts. Only some small conflicts will be accepted. Fox example, traffic on an entry lane can go ahead or turn to right, if the percentage of traffic turning to right is low, the conflict between vehicle turning to right and crosswalk in parallel may be accepted. Generally, the conflict should be indicated by at least one flashing signal. For example, as shown in fig.4 (GE204) the conflict between V4-turning to right and P8 is accepted. The flashing signal CL10 shows the conflict and means vehicle (V4) turning to right should give ways to the pedestrian in the crosswalk P8.

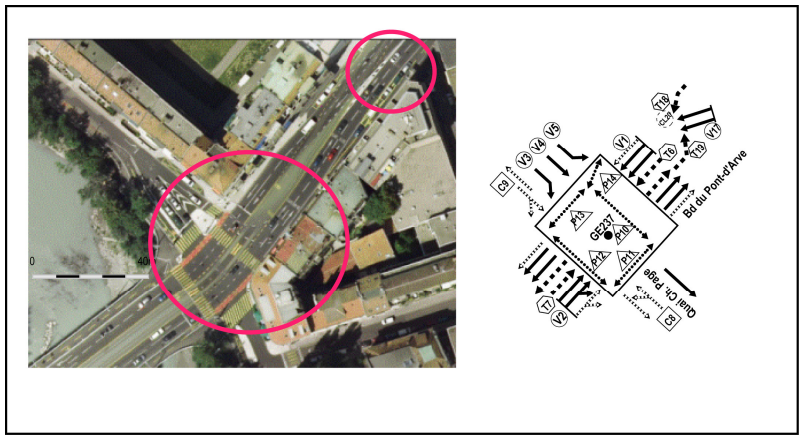


Figure 3.GE237: the transit signal T19 should be coordinated with the transit signal T7 and the transit signal T6 should be coordinated with the transit signal T18

3 Determination of control equipment

After the first step, the junction controller and equipment could be determined.

As transit priority oriented signal control is a type of intelligent signal control. For a transit lane, a special transit signal is used to provide more information for drivers of public transport. More detail information about this transit signal is shown in appendix.

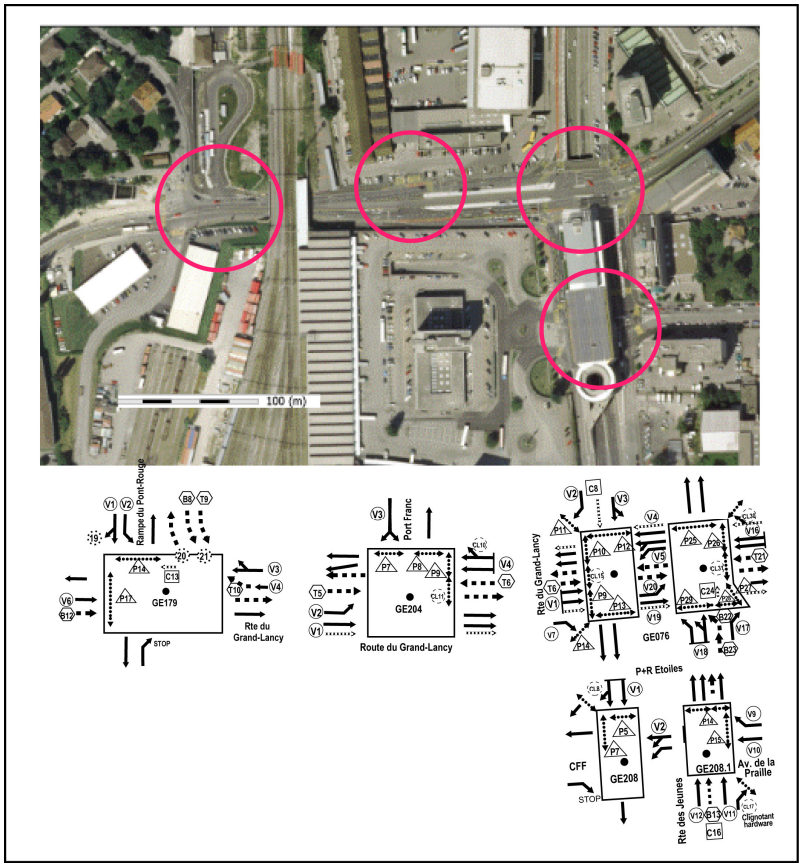


Figure 4. GE208⇔GE76⇔GE204⇔GE179: Signals of adjacent junctions should be coordinated for both transit movements and non transit movements

For transit movements, a special detection system (named as IFAS) that can provide transit arriving information is used. For a transit movement, maximum four points in his traffic lane, tab0, tab1, tab2, tab3 may de defined. A message (bus line No, bus course no, bus length, bus destination, time difference relating to timetable, etc.) will be send to junction controller each time bus passes a point. Point tab0 and tab1 are the points for sending bus arriving message. tab3 is located at bus station for sending bus departure message (when bus doors are closing). Tab2 is located at about 10m after stop line for sending bus leaving junction message. With this detection

system, the time lengths of all transit will arrive at stop line can be known. This transit arriving information is used in the transit priority control logic. The good operation of the transit detection system is vital to transit priority signal control. Unfortunately, the reliability of the detection system is not 100% (>95%). If a signal for a transit movement can have green only based on the bus arriving message, two special bus/tram loops (before and after the stop line) should be used in order provide green (without transit priority) in case that the special bus detection system does not work.

First of all, the controller selected should be compatible with the transit signal and the special transit detection system.

The type of controller and equipment to be used should satisfy all the requirements of the transit priority oriented signalised junction design. As signal group oriented signal control method is used in the design, then the controller determined should be signal group oriented controller. If signal coordination is needed, the controller should have the capability of signal coordination with adjacent signalised junction (by TX, controller cycle second). In order to realise transit priority control logic, the programming tool for the controller selected should be able to perform any complicated control logic proposed in the detail signal design. In order to make easy the testing process, error checking and maintenance in real time operation, the controller selected should have a recording system (can record signal state, control parameters, detector state and other useful real-time information about junction signal control in each second for 24 hours).

The list of control equipment is produced for the purchase purpose, it includes following main information:

-Type of controller
-Number of signals (car, cycle, pedestrian, bus/tram, flashing)
-Number of detector (car and cycle)
-Number of transit detector (type of frequency)
-Number of push button for crosswalks
-List of signals and their description
-List of detectors and their description about detector type and size (their positions are shown in the junction plan)
-Junction geometry plan

4 Junction detail signal design

Based on control strategy study, different junction detail signal design alternatives can then be developed. Each junction signal design alternative consists of (fixed time and/or actuated) a base signal timing plans or actuated signal operation logics (with or without taking account of signal coordination with adjacent signalised junctions). These alternatives can be verified by the evaluation of its signal timing

plans and by simulation for the transit movements. The optimum alternative is the alternative where the all transit movements have less delays and the impact of transit priority on other traffic movements is limited.

In this step, a matrix of inter-green used for all signal design alternatives should be calculated and special signal timing plans, for the case where transit detection system is out of work, should also be developed.

The main idea of transit priority control logic is to arrange green for transit movement based on the transit arriving information. According to the type of signal control determined in the control strategy study, detail signal design for transit priority can be done as follows.

Detail signal design for coordinated signal control:

Case 1: having a cycle time: coordinated signalised junctions always have the same cycle length and same cycle second (TX)

First, different alternatives of base signal timing plan for signal coordination may be developed. If transit movements are belongs to the main coordinated traffic movements, then find the alternative in which transit movements have less delays and the impact on other traffic movements is limited.

If a transit movement is not belongs to the main coordinated traffic movements (for example there is a bus stop between two coordinated junctions), the coordination of transit movement is should be taken account too. In this case, the signal timing plan with taking account of transit movements should be produced. Compare this signal timing plans with the base signal timing plan selected, the impact of transit priority on signal control can be quantified.

For the transit movements, if its greenable length (GTmax) is small, then it is prefer to try to use double openable zones in one cycle where only one open zone can be really used based on transit arriving information.

Case 2: having no cycle time but some signals should be coordinated

This case only can be realised at very simple junction (with few signals). First the different alternatives of signal coordination schema may be developed. If transit movements are belongs to the main coordinated traffic movements, then find the alternative in witch transit movements have less delays and the impact on other traffic movements is limited. If a transit movement is not belongs to the main coordinated traffic movements, the green for transit movement should be arranged based on the transit arriving information.

Detail signal design for no-coordinated signal control

Based on transit arriving information, current signals in green that conflict with the arriving transit movement should be closed in time and the transit signal should be green when transit arrives at junction. Transit delay in no-coordinated signal control should be near to zero.

5 Controller programming and test

In the step of controller programming and test, the base signal data should be input to controller and actuated signal operation logics should be translated to controller's language.

There are two type of programming method:

1. Detail programming from getting info from detectors to control signals such as Masmo, TL, VB.
2. Parameterisation of signal control model: control logic is realised by enter parameters, such as VSPLUS, ITSignal Programmer

The advantage of first type of programming is that it can realize very complicated control logic and it needs less time for doing the test. Its disadvantage is that the programming need more time and the programmer should know well the programming language.

The advantage of the second type of programming is that it needs less time and the programmer need not know well any programming language. Its disadvantage is that it can only realize control logics predefined and need more time for doing the test if the programmer does not understand entirely the signal control model.

An offline test with simulation of transit arriving should be done. There are two type of test: (1) test by a real controller or simulator; (2) test by computer simulation

Test by controller or simulator means to test programming by the real controller or a simulator. The simulation of detector by switches can be done. As there are a lot of possibilities, the test process needs a lot of time.

Test by computer simulation means to test controller programming by a traffic network simulation model. If it is exist, the computer simulation with real controller's program is the best way to test real controller's program with different traffic volumes and transit arriving conditions. It is the best way to verify traffic control strategy, especially the transit priority function. Furthermore, the transit priority effectiveness of several signalised junctions in a traffic network can be simulated and evaluated.

The signal control state and parameters during test process can be recorded, by using software named SigRecAnalyser5 (see next section); the test process may be shortened.

6 Analyse of recording data and evaluation of the effectiveness of the transit priority signal control

After the junction is in service with controller's new program, the observation is necessary to adjust some parameters based on real situation. To evaluate the whole signal control quality and transit priority performance, an analysis on real-time recording data (normally for 1 day) by SigRecAnalyser5 will be done.

The main results are:

-minimum, maximum and mean green interval length of each signal

-minimum, maximum and mean red interval length of each signal

-minimum, maximum and mean red length of the signal for which green request is detected,

-minimum, maximum and mean travel time and theoretical travel time of transit movement between two message-sending points (hence, minimum, maximum and mean transit delay can be determined).

The position of maximum red length of the signal for which green request is detected can be localised by the software and by analysing in detail the recording data, the reason of this red length can be easily understood. Similarly, the position of maximum delay of a transit movement can be localised and by analysing in detail the recording data, the reason of this delay can be easily understood. Modifications on controller's program are some times necessary to improve the signal control quality after analysis of recording data.

The effectiveness of transit priority signal control can be verified by analysing real recording data. Fig.5 shows recording data analysis results (T13) of a transit priority oriented signalised junction (GE88). This is a signalised junction with 3 sub-junctions, where signals for some transit movements should be coordinated but without cycle length in the signal control. Transit arriving announce time distribution in signal cycle shows that tram may send arriving message at any time of a signal cycle and the travel time distribution shows the travel time is nearly same as theoretical time. That means most trams pass without delay. Fig.6 shows recording data analysis results (T19) of a transit priority oriented signalised junction (GE233). This is a signalised junction with 2 sub-junctions; signals for transit movements and non-transit movements should be coordinated and having a cycle length. Arriving announce time distribution in signal cycle shows that tram send arriving message mainly in the time interval (45,75) and the travel time distribution shows the travel time is nearly same as theoretical time (most tram pass without delay).

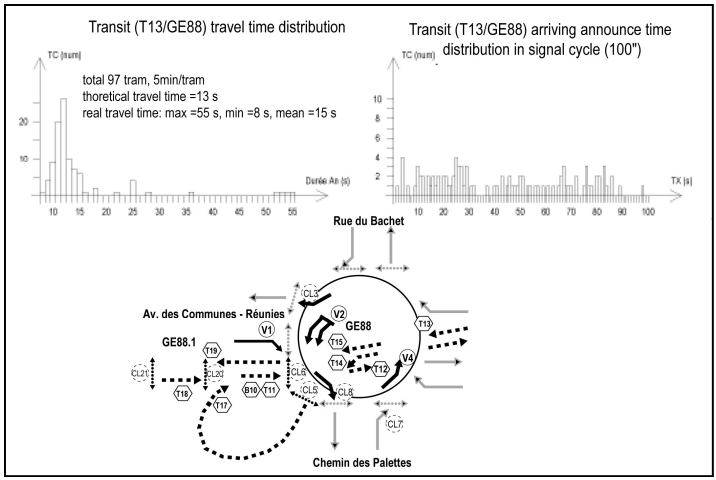


Figure 5. GE88: Travel time distribution and arriving announce time distribution of a transit movement (T13) based on real recording data

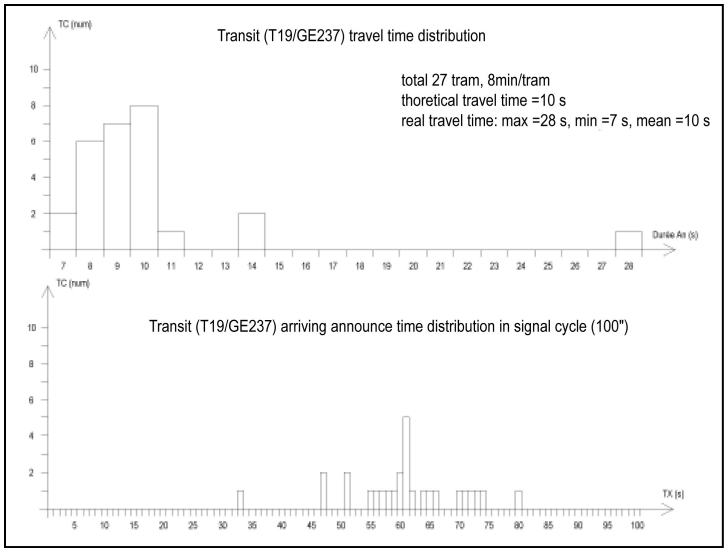


Figure 6. GE237: Travel time distribution and arriving announce time distribution of a transit movement (T19) based on real recording data

7 Development of useful tools for signalised junction design and a new signal oriented traffic controller

The design and realisation of an intelligent signalised junction, especially transit priority oriented signal control, is a very complicated process. In order to find the optimum signalised junction design quickly and accurately, to make easy for controller programming and test, computers tools are indispensable. On the other side, an optimum signalised junction design needs a high performance signal group oriented traffic controller that should have the possibility of perform complicated control logic programming. Based on more than ten years' experiences on signalised junction design, controller programming and test, Following useful computer tools and a new traffic controller are developed as shown in fig.7.

A useful computer tool named as SJDesigner2 (Signalised Junction Designer), consists of SJAnalyser (Signalised Junction Analyser) and SigDesigner (Signal Designer), is developed. It is used for intelligent signalised junction design, especially for transit priority oriented signalised junctions.

In the step of junction control strategy study, by the help of SJAnalyser, different junction geometry design alternatives can be easily developed. For each alternative, transit greenable length, transit performance ratio and junction capacity used in percentage can be calculated by applying different designing traffic data.

In the step of junction detail signal design, by the help of SigDesigner, different base signal timing plan alternatives and different signal coordination alternatives can be easily developed and evaluated. Furthermore, it can be used to develop intelligent traffic control logic, especially transit priority control logic. The final signal design data can be export to a file. The file can be used as an input file of controller ITSignals to generate its base signal control data.

A new traffic controller, named as ITSignals is developed. It is a signal group oriented traffic controller, very flexible and high performance. Signal coordination with adjacent signalised junction using ITSignals can be easily realized without cable. Controller programming is very easy by only filling the control parameters and very complicated control logic, especially transit priority control logic, may be easily realized. It has a recording system that can record useful signal control information in each second for programming test, for signal control evaluation or for the proof of signal operation in case of traffic collision at the junction. ITSignals needs two type data: base signal control data and intelligent signal control data

ITSignal Programmer is a programming tool to input directly the base signal control data and intelligent signal control data for controller ITSignals. The data file of ITSignal Programmer can be used by controller ITSignal and the traffic network simulation software ITSimu

ITSignal Simulator is a programming tool not only to input directly the base signal control data and intelligent signal control data for controller ITSignals, but also

for programming test. The programming test process can be recorded and analysed by the software SigRecAnalyser. The data file generated by ITSignal Simulator can be used by controller ITSignals and the software ITSimu.

ITSimu is traffic network simulation software for signalised junctions with fixed-time control or intelligent signal control with the controller ITSignal.

SigRecAnalyser is a programming tool to analyse the recording data generated by the controller ITSignal, by the software ITSignal simulator or by the software ITSimu.

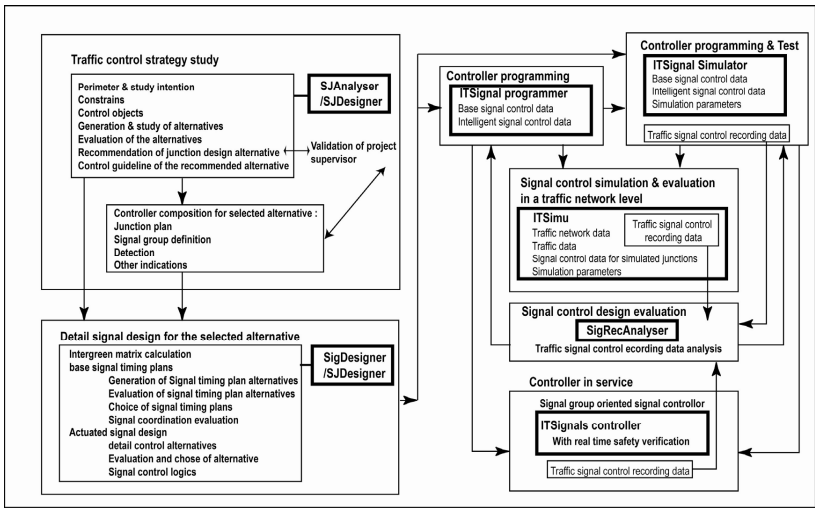


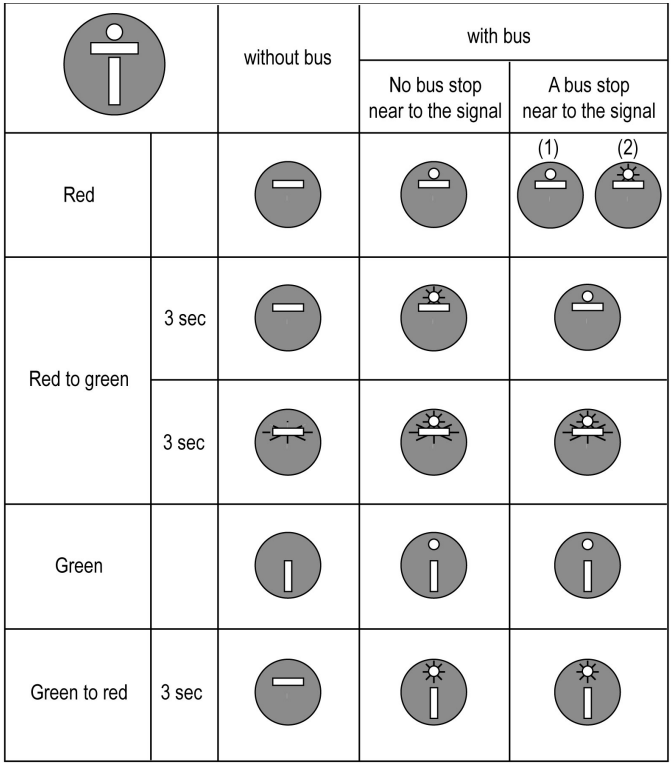
Figure 7. Use of computer tools for traffic control strategy study, detail signal design, controller programming, test and signal control evaluation

8 Summary and further researches and development

The realisation of a transit priority oriented signalised junction is a complicated process. The effectiveness of transit priority oriented signalised junction can be measured by the transit delays decreased and the level of impact of transit priority on non transit movements. Often, the best design alternative to be searched is the alternative in which transit delays should be minimised and the impact of transit priority on non transit movements is acceptable. Because it is the junction geometry design alternative who determines junction capacity and mean transit delays, so the most important thing is to find the junction geometry alternative in the junction control strategy study. The detail signal design, based on this geometry alternative, may have the possibility to give the best result for transit priority signal control. The signal control equipment is determined based on this junction geometry design alternative. A well designed transit priority oriented signalised junction is a good combination of a

junction geometry design, its signal design, and high performance control equipment with fully tested programme. Computers tools are very important in the design and realisation of transit priority oriented signalised junctions.

Continually beneficing experiences from real transit priority oriented signalised junction control projects and new knowledge on intelligent traffic control, computer tools and Itsignals controller will be improved constantly. Further effort should be done to find the method that can improve further transit priority for coordinated signalised junctions. Also further efforts should be down to find an evaluation method on signal coordination design in order to answer the following questions: if signal coordination of two adjacent junctions is necessary? How to measure quantitatively the necessity of signal coordination? What is the relation between signalised junction capacity, distance between adjacent signalised junctions and their signal coordination?



(1) it is not in the period to close bus door (2) it is in the favourable period to close bus door to have green

Figure 8. Special transit signal

References

J. Tan (2005) "Transit priority performance at signalised intersections", 5th Swiss Transport Research Conference, Ascona, Switzerland.

J. Tan (2006) “SJDesigner: A powerful tool for signalised junction design", 6th Swiss Transport Research Conference, Ascona, Switzerland.

J.Tan (2004) “Key movements and signalised junction’s capacity”.

R.Golay, Ph. Glayre and Ph. Bovy (2001) “La régulation des carrefours à feux”, Cahier no11, TEA-ITEP, Swiss Federal Institute of Technology Lausanne.

J.Tan (2005) “SigRecAnalyser, A program to evaluate an intelligent signalised junction’s operation by analysing its real signal recording data”

J.Tan (2006) “ITSignals Controller: An intelligent traffic signal controller”

Appendix: Transit signal

In order to provide more comfort to public transport, a special transit signal for public transport is introduced. As shown in fig.8, the horizontal bar equal to “red”, the vertical bar equal to “green”. If a bus / tram is coming and it is recognized by the junction controller, the “point” in the transit signal will be shown. When the point is starting to flash, the transit will be green soon in 5sec. During the green period, the point means the transit has not left the junction and the last 3 sec of green, the point will be flashing. During the red period, if the bus is at the bus stop, the point in flashing means it is best moment to close the bus doors and start up as green will come soon.

Integrative Architecture of ATMS for Urban Elevated Expressway Network in Shanghai

DU, Yuchuan [1] and YUAN, Wenping [2] and SUN, Lijun[3]

[1]Key Lab. of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 200092, China; PH:(86)21-65985194; FAX:(86)21-65982273; email: ycdu@mail.tongji.edu.cn
[2]Shanghai Municipal Engineering Administration Department, Shanghai 200030, China; PH:(86)21-62808822; FAX:(86)21-62804648; email:wenpingyuan@sina.com
[3] Key Lab. of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 200092, China; PH:(86)21-65980561; FAX:(86)21-65982273; email: ljsun@mail.tongji.edu.cn

Abstract: With the development of economy, Shanghai, the biggest city of China, suffers from severe traffic congestion. The elevated expressways in downtown Shanghai are highly clogged during peak hours as well as off-peak hours. To improve traffic situation of elevated expressway without adding more infrastructure and creating negative impacts on environment, an ATMS for elevated expressway has been established. Considering the complexity of urban elevated expressway network, an integrative architecture is proposed in this paper for system development of ATMS. This integrative architecture includes traffic management strategy, physical framework and main functions of ATMS. The field equipment part of this ATMS consists of real-time traffic data collection subsystem, traffic guidance and control subsystem, and real-time surveillance subsystem. This paper gives the detailed introduction of the features of these subsystems. In addition, it gives the evaluation results for ATMS demonstration project and gives recommendations for further research.

Keywords: *urban expressway, ATMS, traffic management strategy, physical framework, System function*

1 Introduction

The elevated expressway network in downtown Shanghai has been built from 1992. With the traffic openness of YanAn elevated expressway on September 15, 1999, a large-scale and extraordinarily complex expressway network (as illustrated in Fig. 1) has formed, which plays a more and more important role in urban economic and daily life activities. Due to continuous traffic and non-disturbance, this expressway network has attracted more and more intra-city traffic. In recent years, the traffic demand of this elevated expressway network has exceeded one million each day and the traffic volume of some sections has already topped 200,000 PCU. It has been reported that nearly half sections of network are highly clogged during peak hours as well as off-peak hours. Traffic congestion has decreased the overall level-of

service of the elevated expressway network and may introduce the potential of more traffic crashes.

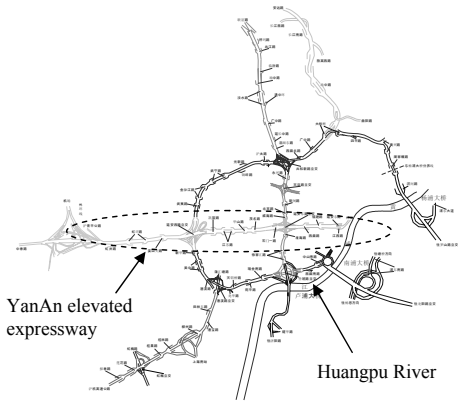


Fig.1 The expressway network in downtown Shanghai

Suffering from severe traffic congestion, the road authority in shanghai have made lots of efforts to alleviate congestion. To improve traffic conditions without adding more infrastructure and creating negative impacts on environment, ITS technologies have been considered a cost-effective option. With such considerations, Tongji University was responsible for transportation planning and development of traffic control system, and the Shanghai Municipal Engineering Administration Department (SHMEAD) organized the implementation of "the YanAn elevated expressway Advanced Traffic Management System (ATMS)" for a demonstration of deploying systematic ITS technologies. The demonstration project starts from HongQiao Airport to the Bund and is about 14 km long. The demonstration area is shown as Fig. 1. This project is also an important part of "the key technological development of ITS and demonstration project", which is the large-scale research project of the Ministry of Science and Technology of the people's republic of China.

An advanced transportation management system (ATMS) is the basic building block of intelligent transportation systems (ITS). The Strategic Plan for IVHS in the United States (1992, 1997) identified six functional areas as the basis for early IVHS planning and project implementation and the first functional area is ATMS. CommuterLink (2004), the ATMS of Utah Department of Transportation, has grown substantially since it was established five years ago. CommuterLink integrates more than 800 traffic signals, 1400 detector stations, 250 closed-circuit television (CCTV) cameras, and 70 Variable Message Signs, saving Utah travelers US$179 million and 9.8 million hours annually and preventing 948 traffic accidents and 3 traffic-related deaths each year. Only a few cities such as Tokyo and Osaka in Japan which have a large scale urban expressway network widely used ATMS to optimize network traffic flow (2001, 2005). Because of the difference of traffic management process, these application systems have various system frameworks.

After analyzing the traffic features of Shanghai elevated expressway network and the requirement of modern urban traffic administration, an integrative architecture was proposed and used in the YanAn elevated expressway ATMS to accommodate real-time traffic management. This integrative architecture includes traffic management strategy, physical framework and main functions of ATMS. The ATMS for YanAn elevated expressway is regard as the first set of modern urban expressway traffic management system in China.

2 Traffic management strategies

The elevated expressway network of Shanghai is a complicated transportation network. For a great deal of long distance travelers in this network, there are many feasible paths to be chosen. The traffic management measures for this network should be regional-oriented and dynamic. In fact, when the scale of network reaches considerable degree, no matter for convenience of management or requirement of quality control, the network should be divided into several layers which carry out different functions. Considering the complexity of urban expressway network, a new transportation management strategy was proposed for planning and designing of Shanghai Elevated expressway ATMS. It can be described as "Controlling in small region and Guiding in large region".

According to the transportation management strategy, the ATMS is design to two levels in logic. The superstratum is Overall Coordination Level (OCL) and the substratum is Cell Control Level (CCL). For the local traffic problem in a control cell, CCL can provide solutions directly. For the serious accident which should impact the whole network, OCL will consider the situations of all control cells to provide management plan. Based on analysis of serious congestion areas and relationship of sections, the whole network of Shanghai elevated expressway is divided into 4 control cells. They are North Area, Central Area, South Area and West Area. Owing to the logic division, the control cell can be easily adjusted when characters of local traffic flow change.

Considering it's the first time that the urban expressway ATMS is applied in China, the stability and accuracy of ATMS equipments should be verified. In the other side, the operators need time to accommodates to a new working circumstance. The traffic management process of Shanghai elevated expressway ATMS is described as "Four Steps Principle", including "Finding Problem", "Affirming Problem", "Providing Solution" and "Evaluating Effect". The "Four Steps Principle" traffic management process is shown in Fig. 2.

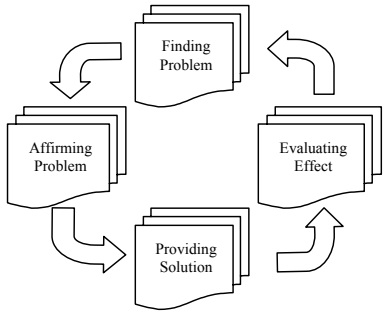


Fig. 2 Four steps principle

3 Physical framework and main system functions

Based on network traffic management strategy and the "Four Steps Principle" working process for Shanghai elevated expressway netwrok, the physical framework of ATMS was proposed and is presented in Fig. 3. The framework includes three parts: control center, field-equipment system and communication system which connect the control center and field-equipment system. The whole field-equipment system can be separated into three subsystems according to the functions: real-time traffic data collection subsystem, traffic guidance and control subsystem and real-time surveillance subsystem.

The control center is the hub of Shanghai elevated expressway ATMS, which is responsible for real data collection and transaction, such as traffic and meteorological environment, and provide traffic guidance information trough comprehensive real-time traffic analysis. Thus, the control center plays a critical role in the successful operations of this ATMS. The central traffic control software is the key part of the control center. The main functions of traffic control software are shown below.

- Data collection: Gathering and fusing various kinds of real-time traffic data that different field equipments measured;
- Real-time video surveillance: Receiving video images that the field cameras send and offering the controlling methods of the field cameras;
- Data pre-treatment: Forming the traffic flow parameter from the original information;
- Data management: Managing traffic data information that has been gathered or/and produced, including implement managing and historical data management;
- Video management: Offering video images and cameras management methods;
- Traffic state identification: Utilizing pre-treatment traffic information to identify the traffic state of expressway sections;
- Travel time prediction: Utilizing traffic flow model to predict travel time according to real-time traffic data;
- Traffic information dissemination management: disseminating traffic information to various kinds of Variable Message Sign (VMS) according to different requirements.

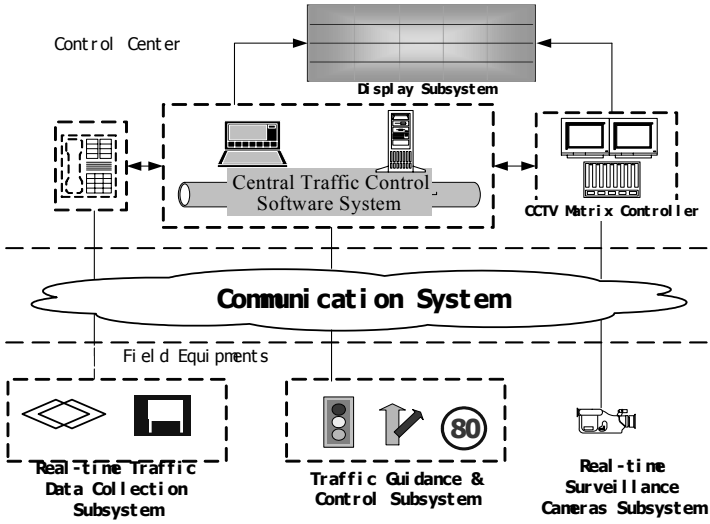


Fig. 3 Physical framework of Shanghai elevated expressway ATMS

4 Features of field-equipment subsystems

As described above, the part of field-equipment system of Shanghai elevated expressway ATMS consists of real-time traffic data collection subsystem, traffic guidance and control subsystem and real-time surveillance subsystem. Their features are described as follows.

4.1 Real-time traffic data collection subsystem

The inductive loop detector is used in real-time traffic data collection subsystem. Loop detectors can be used to get volume, mean vehicle speed and occupation proportion. The interval of loop detectors in main stem is between 400 m to 500 m. In the sections close to ramp or interchange, the interval is shorted to 250 m. The dense arrangement of loop detectors ensures that control center can find traffic congestion in time.

4.2 Real-time surveillance subsystem

Surveillance video cameras are used to monitor traffic and give operators the better capability to immediately confirm traffic problems. There are two kinds of camera was used in real-time surveillance subsystem: tower camera and pole camera. Tower camera can cover 2 km to 5 km, whereas pole camera just can cover 500 m to 1000m. With the combination of different cameras, the surveillance area covers the whole elevated expressway network. The operators in control center can get the real-time video image of anyplace in the elevated expressway network.

4.3 Traffic guidance and control subsystem

The infrastructure of traffic guidance and control subsystem is an integrated collection of different kinds of VMS under the control of traffic control center. The VMS can be categorized into three groups depending on the different objects of traffic management. These groups are (1) Type-L VMS, (2) Type-M VMS, and (3) Surface-Street VMS. Specific features are discussed as follows:

(1)Type-L VMS

Type-L VMS is a kind of graphical VMS, which is illustrated in Fig. 4. They are installed on the upstream of the key decision points (such as interchanges) elevated expressway network and provide the real-time color-coded Level of Service (LOS) information of sections in a large region of the expressway network. Type-L VMS is used to assist the driver to make on-road route selection to avoid congestion expressway sections.

Fig. 4 Type-L VMS

(2)Type-M VMS

Type-M VMS is a kind of hybrid VMS, which is illustrated in Fig. 5. The upside of this VMS is graphic VMS and the downside of this VMS is textual VMS. They are installed on the upstream of off-ramps and provide the real-time color-coded Level of Service (LOS) information and accident textual information of downstream sections in a moderate region of the elevated expressway network (usually covering 3 off-ramps). Type-M VMS is used to assist the driver to go down from off-ramps to avoid serious congestion of elevated expressway downstream sections.

Fig. 5 Type-M VMS

(3)Surface-street VMS

Surface-Street VMS is a kind of textual VMS, which is illustrated in Fig. 6. They are installed on the surface streets near the on-ramps of elevated expressway and provide the information about the traffic condition of ramps, including three statuses: “Open”, “Congestion”, “Closed”. Surface-street VMS is used to assist the driver to decide whether use elevated expressway or not in advance.

Fig. 6 Surface-street VMS

5 Effect of ATMS in YanAn elevated expressway

After the YanAn Elevated Expressway ATMS was put into operation, a traffic impact analysis has been conducted to evaluate the effect of ATMS. Three typical elevated expressway sections: NX10 (Loushanguan Road – Yanxi interchange), NX24 (Jiangsu Road – Maomin Road), and NX31 (Yandong interchange – Xizang Road)

were selected and six days (before ATMS operation: 16-Oct, 17-Oct, After ATMS operation: 20-Oct, 21-Oct, 10-Nov, 11-Nov) traffic data were collected to analyse traffic impact of ATMS

The field measurement data show that the time proportion of free flow and the mean travel speed of elevated expressway represent a significant improvement after the operation of ATMS. The time proportion of free flow has a great improvement of 22% after the operation of ATMS. Meanwhile, the time proportion of jam has an average decrease of 15% in for days after the operation of ATMS. In another side, the mean speeds of different sections in YanAn elevated expressway increase from 6.6% to 21% after the operation of ATMS. For another important traffic parameter, the traffic volumes in different sections of YanAn elevated expressway have no obviously change.

The comparison data indicate that the traffic situation of YanAn elevated expressway is improved obviously after the operation of ATMS.

6 Conclusions

Shanghai is one of the 10 ITS demonstration cities in China and keeps ahead in urban traffic management. This paper introduces a practical application of ITS in elevated expressway management, focusing on traffic management strategy, system framework, features of field subsystems. The result of the before-and-after study on the Shanghai Elevated Expressway ATMS shows that the traffic situation of elevated expressway is improved obviously with the total traffic volume being almost unchanged. The quantified benefits to be obtained from the before-and-after study will not only provide the base to justify the Shanghai Elevated Expressway ATMS, but also can be used to estimate the benefits of future ITS projects.

In spite of the positive effects of the Shanghai Elevated Expressway ATMS, there remain some problems which need to further explored. Some of them are discussed below.

- How to coordinate information dissemination with different kinds of VMS in elevated expressway to achieve the control objectives, such as system optimization, user equilibrium, and so forth.
- How to extend the dissemination field of traffic information colleted by ATMS, such as in-vehicle, internet, and so forth.

Acknowledgement

This paper was based on the result of the research project "Study on Shanghai Expressway Traffic Management System and the Piloting Project" which was jointly supported by a research grant (2002BA404A08) form "Ministry of Science and Technology of the People's Republic of China (MOST)" and a research grant (012112039) from "Shanghai Science and Technology Commission (SSTC)". The authors take sole responsibility for all views and opinions expressed in the paper.

References

Cisco Systems Inc. (2004) UTAH DEPARTMENT OF TRANSPORTATION CREATES COMMUTERLINK TO HELP UTAH RESIDENTS "KNOW BEFORE YOU GO", San Jose, CA, USA.

Du, Yuchuan and Sun, Lijun (2005) The Research on Advanced Traffic Management System of Shanghai Urban Expressway. China Municipal Engineering, 6, pp. 4-8.

Feng, Qi and Zhou, Xiaopeng and Du, Yuchuan, etc. (2004) The Research on System Framework of ATMS for Urban Expressway, Computer and Communications, 23, pp 67-70.

Gan, Hongcheng and Sun, Lijun (2004) Design and Application of a new kind of Variable Message Sign, Computer and Communications, 22, pp 93-96.

ITS America (1997) National ITS Architecture-Physical Architecture. Federal Highway Administration, Washington D. C.

IVHS America (1992) Strategic Plan for Intelligent Vehicle-Highway Systems in the United States, Federal Highway Administration Washington, D.C.

Kazuyoshi Sakamoto and Tsuyoshi Sakaino (2001) Provision of Metropolitan Expressways Information by Graphic before Entering Expressways. Proceedings of the Eighth World Congress on ITS, Sydney, Australian, pp15-27.

Sun, Lijun and Du, Yuchuan etc. (2003) Study on Shanghai urban expressway transportation management system and the piloting project, Tongji University, China Shanghai.

CAPACITY AT A SIGNALIZED INTERSECTION APPROACH WITH SHORT LANES RELATED TO THE CYCLE LENGTH

Jinyong JIANG[1] and Peikun YANG[2]

[1]School of Transportation Engineering, Tongji University, Shanghai, 200092, P.R. China; PH (86)13622345076; email: jy_jiang@21cn.com
[2]School of Transportation Engineering, Tongji University, Shanghai, 200092, P.R. China; PH (86)21-65982878; email: pk_yang@vnet.citiz.net

Abstract: In order to increase the capacity at a signalized intersection, cycle lengths in many large cities in Mainland China are extended continually to "match" the traffic demand. However, longer cycle time does not always bring up increasing capacity, especially at those signal intersections with short approach lanes. This research attempts to derive a crucial factor to help deciding whether the extension of the cycle length for a certain signalized intersection would be effective or not. It was found that for an approach with short lanes, the relationship between the capacity and the cycle length depends on the saturation flow rate of the normal lanes versus that of the short lanes, the effective length of the short lanes, the lose time and the cycle length itself. The criteria to estimate the above relationship were presented and validated through numeral computing.
Key words: *short lanes, capacity, cycle length, signalized intersections*

1 Introduction

It is normally thought that the capacity increases with the increasing of the cycle length at a signalized intersection. By this reason, signal cycle lengths in many large cities in Mainland China are extended continually to "match" the increasing traffic demand, which led to a ridiculous phenomenon that lots of intersections are operating with cycle lengths longer than 300 seconds. However, longer waiting time does not always bring up increasing capacity especially at those signal intersections with short approach lanes. While a longer cycle length can reduce the proportion of the lose time, the saturation flow rate of the short-lane approach would decrease with the increasing of the cycle length after it is long enough.

Although a significant amount of research has been devoted to studying signalized intersections, a limited number of studies addresses the effect of the short lanes on signalized intersection operations, such as the impact on intersection capacity and delay. Akcelik (1978) analyzed the operation of a signal intersection approach with short lanes and provided a simple equation to calculate the saturation flow rate. In the Highway Capacity Manual (HCM) (TRB 2001), lane utilization factors are provided to deal with short through-lane additions, and exclusive turn lanes are assumed to have enough bay length. Tian and Wu (2006) presented a probabilistic

model to calculate a signalized approach capacity with a short right-turn lane. Few studies emphasized the effect of the cycle length on the capacity (not just the saturation flow rate) of an approach with short lanes.

The purpose of this paper is to explore the relationship between the capacity and cycle length for a signalized approach with short lanes. The authors attempted to find out some crucial factors to help deciding whether the extension of the cycle length for a certain signalized intersection would be effective or not.

2 Approach Capacity Model

The HCM provided a comprehensive concept and methodology for analyzing and estimating capacity, signal time and LOS for lane groups and intersection approaches or the intersection as a whole. This paper follows those concepts and methods as the basis for the study. The variables and parameters used in this paper are given in the Notation.

The capacity of a signalized approach with short lanes is defined as the maximum flow rate (measured in vehicles per hour) that the approach can service under prevailing geometry, traffic flow, and signal timing conditions (Tian and Wu 2006).

To simplify the model, the signalized approach is assumed to be a through lane group with the same signal phase including *M* normal lanes and *N* short lanes (see Fig. 1). Under the case where the normal lanes and the short lanes are not in the same direction, some parameters would be applied to reflect the adjustments on the lane-usage balances. That would not affect the conclusions of the study.

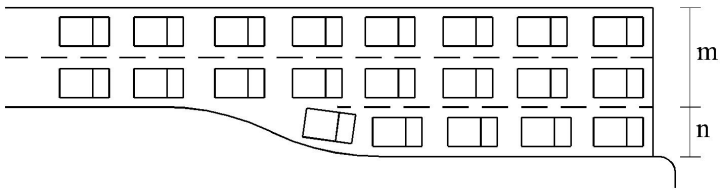


Figure 1. Layout of a signalized approach with short lanes

The effect green interval for the approach is divided into two portions. In the first portion, both of the normal lanes and the short lanes are used to discharge vehicles. The saturation flow rate of the first portion can be stated as:

$$S_1 = S_m + S_n \tag{1}$$

The length of the first portion of green, which is the time required to discharge the queues in the short lanes when they are crammed, can be calculated from the following equation:

$$g_n = \frac{D}{j}\left(\frac{3600}{S_n}\right) \tag{2}$$

In the remaining portion of the green interval, only the normal lanes would be used to discharge vehicles. The saturation flow rate of the portion is equal to S_m.

Then, the capacity of the saturation flow rate of the approach can be calculated from Eq. 3 and Eq. 4 respectively:

$$c_a = \frac{(S_m + S_n)g_n + S_m(g_m - g_n)}{C} = \frac{(S_m g_m + S_n g_n)}{C} \tag{3}$$

$$s_a = \frac{(S_m g_m + S_n g_n)}{g_m} \tag{4}$$

3 Capacity Related to Cycle Lengths

The capacity model of Eq. 3 implies the relationship between the capacity and the cycle length. The effective length of the green interval varies with the cycle length. The green interval can be calculated using the following equation:

$$g_m = (C - L)k \tag{5}$$

The variable k is the portion that the effective green time allocated to the approach with short lanes. It depends on the policy employed in allocating the available green time of the cycle. A common policy is to allocate the green such that the v/c ratios for critical movements in each phase are equal. Thus,

$$k = \frac{y_a}{y_a + y_b} = \frac{\left(\frac{v_a}{s_a}\right)}{\sum_i \left(\frac{v}{s}\right)_i} \tag{6}$$

Merging Eq. 4 through Eq. 6, the following relationship emerges:

$$y_a = \frac{v_a}{s_a} = \frac{v_a g_m}{S_m g_m + S_n g_n} \tag{7}$$

Merging Eq. 5 through Eq. 3, the following relationship emerges:

$$c_a = S_m k - \frac{S_m L k}{C} + \frac{S_n g_n}{C} \tag{8}$$

$$\frac{\partial c_a}{\partial C} = S_m \frac{\partial k}{\partial C} - \frac{S_m L C \frac{\partial k}{\partial C} - S_m L k}{C^2} - \frac{S_n g_n}{C^2} \tag{9}$$

Merging Eq. 5 and Eq. 7 through Eq. 6, the following relationship emerges:

$$k = \frac{\frac{v_a (C-L) k}{S_m (C-L) k + S_n g_n}}{\frac{v_a (C-L) k}{S_m (C-L) k + S_n g_n} + y_b} = \frac{1}{v_a + y_b S_m} \left(v_a - \frac{y_b S_n g_n}{C-L} \right) \tag{10}$$

$$\frac{\partial k}{\partial C} = \frac{y_b S_n g_n}{(v_a + y_b S_m)(C-L)^2} \tag{11}$$

Merging Eq. 10 and Eq. 11 through Eq. 9, the following relationship emerges:

$$\frac{\partial c_a}{\partial C} = \frac{y_b S_n g_n S_m}{(v_a + y_b S_m)(C-L)C} + \frac{S_m L \left[v_a (C-L) - y_b S_n g_n \right]}{(v_a + y_b S_m)(C-L)C^2} - \frac{S_n g_n}{C^2}$$

$$= \frac{(S_m L - S_n g_n) v_a}{(v_a + y_b S_m) C^2} \tag{12}$$

If $S_m L < S_n g_n$, then $\partial c_a / \partial C < 0$, the capacity of the approach decreases with the increasing of the cycle length.

If $S_m L > S_n g_n$, then $\partial c_a / \partial C > 0$, the capacity of the approach increases with the increasing of the cycle length.

It should be noted that Eq. 12 exits only when $g_m > g_n$, otherwise the short lanes are not "short" and $s_a = S_m + S_n$.

Obviously, when the "short" lanes are not short, the capacity of the approach increases with the increasing of the cycle length. In this case, Eq. 13 should be satisfied:

$$C \le \frac{y_2(S_m + S_n)g_n}{v_a} + g_n + L = C_0 \tag{13}$$

The relationship between the capacity and the cycle length for an approach with short lanes is summarized as Table 1.

Table 1. Approach capacity and cycle length relationship

Criterion 1	Criterion 2	Capacity related to cycle length
$C > \frac{y_2(S_m + S_n)g_n}{v_a} + g_n + L$	$S_m L < S_n g_n$	$\frac{\partial c_a}{\partial C} < 0$, the capacity decreases with the increasing of the cycle length.
	$S_m L > S_n g_n$	$\frac{\partial c_a}{\partial C} > 0$, the capacity increases with the increasing of the cycle length.
$C \le \frac{y_2(S_m + S_n)g_n}{v_a} + g_n + L$		the capacity increases with the increasing of the cycle length.

4 Numerical Validation

In order to validate the criteria in Table 1, numerical computing is conducted, considering a signalized approach consisting of two normal lanes and one short lane, and the related parameters are given as follows:

$s_m = 3600veh/h$; $s_n = 1800veh/h$; $v_a = 1600veh/h$;

$y_b = 0.5$; $L = 16s$; $j = 6m$.

Assuming the length of the short lane is 150m, 120m, 90m, 60m, and 30m respectively, the capacities of the approach according to different cycle lengths from 80s to 210s are then computed. The results are listed in Table 2 and illustrated in Figure 2.

In the results, for *D=150m*, the capacity increases from 1607veh/h to 1795veh/h while the cycle length increases from 80s to 150s; then it decreases to 1767veh/h

while the cycle length increases to 210s. The tendency with the condition of *D=120m* is similar to that of *D=150m*, and its largest capacity is 1746veh/h at the cycle length of 130s. For *D=90m*, *D=60m*, or *D=30m*, the capacity increases with the increasing of the cycle length all through.

These results are consistent with the relationship illustrated in Tab.1. For instance, while *D=150m*, g_n*=50s*, the short lane is not really short until the cycle length get to $C_0 = 150s$ (see Eq. 13), and after that, the capacity decreases with the increasing of the cycle length, for $S_m L = 3600 \times 16 < S_n g_n = 1800 \times 50$; while D=60m, g_n=20s, $S_m L = 3600 \times 16 > S_n g_n = 1800 \times 20$, so the capacity decreases with the increase of the cycle length all over.

From Fig. 2, we can also see that, although the capacities of the approaches with "longer" short lanes are more likely to decrease with the increasing of the cycle length than those with "shorter" short lanes, the former approaches clearly have higher capacities than the latter.

Table 2. Result of the Capacity under different conditions

Cycle length (s)	Capacity of the approach (veh/h)				
	D=150m or W*=0.64	D=120m or W=0.80	D=90m or W=1.07	D=60m or W=1.60	D=30m or W=3.20
80	1607	1607	1607	1567	1461
90	1652	1652	1652	1581	1487
100	1688	1688	1677	1592	1508
110	1717	1717	1679	1602	1525
120	1741	1741	1680	1609	1539
130	1762	1746	1681	1616	1551
140	1780	1743	1682	1622	1561
150	1795	1739	1683	1626	1570
160	1789	1736	1684	1631	1578
170	1784	1734	1684	1634	1584
180	1779	1732	1685	1638	1591
190	1774	1730	1685	1641	1596
200	1770	1728	1686	1643	1601
210	1767	1726	1686	1646	1605

Note: $W = S_m L / (S_n g_n)$

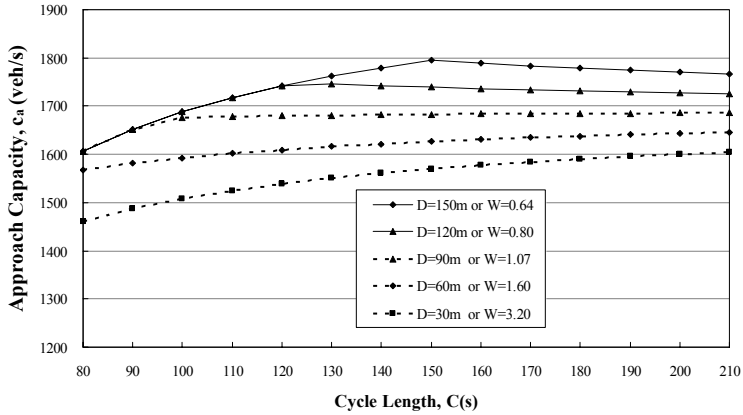


Figure 2. Approach capacity versus cycle length with various short lane lengths

5 Conclusions

This research found that for a signalized approach with short lanes, the relationship between the capacity and the cycle length depends on the saturation flow rate of the normal lanes versus that of the short lanes, the effective length of the short lanes, the lose time and the cycle length itself.

The capacity increases with the increasing of the cycle length until the cycle length reaches a critical value. After that, there are two different development routes: increase or decrease with the increasing of the cycle length. The criteria presented in this paper could be used to estimate the critical cycle length and the development route.

It should be noted that capacity is just one of the important features for the operation of an approach or an intersection, and the choice of an optimal cycle length would take into account more comprehensive considerations.

Notation

The following symbols are used in this paper:

D = effective length of the short lanes that vehicles use to queue (m);

C = cycle length (s);

C_0 = the cycle that the green time allocated to the studied approach equal to g_n;

c_a = capacity of the approach (veh/h);

g_n = effective green length required to discharge the vehicles queuing in the short lanes when they are fully utilized (s);

g_m = total effective length of the green interval (s);
j = average queue space of a vehicle (m/veh);
k = the portion that the effective green time allocated to the approach;
L = lost time per cycle (s);
S_1 = saturation flow rate of the approach in the first portion of green interval;
S_m = saturation flow rate of the normal lanes in the first portion of green interval;
S_n = saturation flow rate of the short lanes in the first portion of green interval;
S_a = saturation flow rate of the approach as a whole (veh/h);
v_a = traffic volume of the approach (veh/h);
$(v/s)_i$ = flow ratio for lane group i;
W = a dimensionless factor, $W = S_m L/(S_n g_n)$
y_a = flow ratio of the studied approach with short lanes;
y_b = sum of the flow ratio of the critical groups except the studied approach;

References

Akcelik, R. (1981). "Traffic signals: capacity and timing analysis." *Research Report ARR No. 123*, Australian Road Research Board, 1981.

Transportation Research Board (TRB). (2001). *Highway capacity manual 2000*, Transportation Research Board, Washington, D.C.

Tian, Z., and Wu, N. (2006). "Probabilistic Model for signalized intersection capacity with a short right-turn lane." *Journal of Transportation Engineering*, 132(3), 205-212.

OPTIMIZATION OF SIGNAL TIMING BY EQUILIBRIUM TECHNIQUE

Ning WU

Institute for Traffic Engineering, Ruhr University, D-44780 Bochum, Germany; PH: +49 (234) 3226557; FAX: +49 (234) 3214151; e-mail: ning.wu@rub.de; http://homepage.rub.de/ning.wu

Abstract: Optimization of signal timing plans is normally based on operation research methods. In this paper, a new optimization technique is introduced. As a basic idea, the signal timing plan is considered as an analogy to a construction structure. Therefore, all of the basic concepts and methods of the structural analysis can be used for solving the optimal timing of the signal timing plans. The only difference compared to the conventional structural mechanics is the non-linear "elasticity" of the signal timing plans. As an objective function of the optimization, either the sum of delays or the fuel consumption or other variables like air pollutants can be used. For the solution of this equilibrium problem, algorithms from structural engineering can be applied. These algorithms are very efficient in term of computation converge very quickly. Therefore, rather complex intersection structures can be optimised with quite short computer times. Also an on-line rolling optimization of signal timing plans with measured traffic flows can easily be realised. This technique is proven to be very effective for practical applications. The new method is capable of extending the optimization from single intersections (including optimization of the cycle time) to network of traffic signals and to traffic-actuated controllers.
Keywords: *Signalized intersection, Optimization of signal timing plans, delays at traffic signals*

1 Introduction

Optimization of signal timing plans is a frequently discussed topic in traffic engineering. Optimization means, above all, minimization of the sum of delays. Webster was the first one to formulate an analytical solution for minimization of the sum of delays for a two-stage traffic signal in 1958 (Webster, 1958). Using some simplifications, he was able to give the optimum cycle time and the optimum stage green times. Since then, numerous papers have been published on theoretical analyses and practical applications of signal timing optimizations. In the field of theoretical researches, publications from Allsop (1971, 1992) and Tully (1976) should be mentioned at the first place. Concerning the practical applications of signal timing optimization, different models and computer programs were developed. The most among them are e.g. the AMPEL program package in German-speaking countries and the SIGSIGN package (Silcock and Sang, 1990) from Great Britain.

The existing procedures for signal timing optimization can be distinguished into two groups: stage-oriented optimization and signal-group-oriented optimization. Most of the previous authors first treated the stage-oriented optimization due to its simple applicability. This simple applicability was necessary and made sense, since traffic signals were stage-oriented themselves until the seventies. Due to development of

modern signal controllers, signal groups can nowadays be controlled individually. Thus, an optimization which takes into account the green time and traffic characteristics of each individual signalized movement at the intersection, becomes necessary.

Most of the existing models for signal-group-oriented optimization were based on the theory of "Operations Research" (OR method) or on "Trial and Error" methods. Since the maximization of capacity is a linear optimization problem, it can easily be handled with OR methods. The minimization of the sum of delays, however, is a non-linear problem. Here, most of the current methods are based on numeric solutions. These numeric solutions are normally quite time-consuming and can, therefore, only be used for off-line calculations of signal timing plans.

This paper presents a new optimization model in which the signal timing plan is treated as a mechanical system and its optimization is achieved by determining the state of equilibrium of the mechanical system. The optimization, i.e. the state of equilibrium, is determined using the so-called moment distribution method from Cross (cf. Beaufait, 1972) in classical structure mechanics. In this mechanism the connections (connection between two or more signal groups) of signal groups (=constraints of the structure) are used as the variables for the optimization in place of the green times. Thus, the free degrees of the system can be drastically reduced. Since this procedure is very efficient (because of less free degrees) and since all characteristic information of the signal timing plan is passed on with every further step of optimization, this optimization procedure is particularly suitable for a dynamic optimization, i.e. for an on-line optimization (adaptive control) of signal timing plans. The principle of the new model can also be transferred to optimizing large scaled, coordinated networks.

This paper is the partial result of a project sponsored by the German Research Community for a two year period.

2 The Signal Timing Plan as a Mechanical System

In the theory of elasto-mechanics (cf. Lehmann 1979), Dirichlet formulated the theorem on the minimum of total potential in stable equilibrium:

> *If a system is in a state of stable equilibrium, the total potential (or total potential energy) of this system has its minimum value*

Normally, one of several possible local minima of the total potential in the mechanical system can be determined if the corresponding state of equilibrium is ascertained. If one compares the objective function (e.g. sum of delays) for the optimization of a signal timing plan with the total potential of a mechanical system, the minimization of the objective function can be achieved by transferring the signal timing plan into the "state of equilibrium".

Figure 1 shows a basic element of a signal timing plan and a basic element of a mechanical spring system. It demonstrates that, in analogy to mechanics, the green time can be described as an elastic spring.

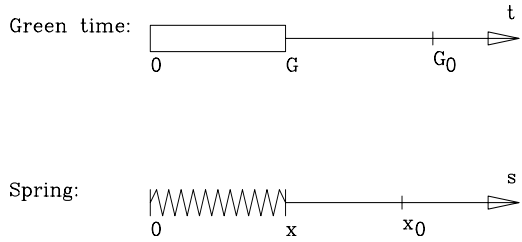


Figure 1. Comparison between a green time and a spring

Denote B the demand for more green time in a signal group. With increasing green time, the capacity is increased and the delay W is reduced. Therefore, B can be defined as the conservative force of the potential function W. Thus, B is the derivative of W towards G, i.e. $B=W'(G)$.

Therefore, the following analogies between the two elements can be established:

Green time		**Spring system**	
Duration of the green time	G	Length of the spring	x
Cycle time	C	Length in original state	l
Traffic volume	q	Specific elasticity	k
Delay	W	Potential	U
Green time demand	B	Spring force	F
Slope of B	M	Stiffness coefficient	K
Point of reference	G_0	Point of reference	x_0

G_0 or x_0 are the points of reference, at which the potential (W or U) is defined as zero. They are chosen so that the force F and the "green time demand" B is zero. That is,

$$G_0 = C \quad (\text{or } G_0 = \infty) \qquad\qquad x_0 = l$$

It is true

$$B = f(q,G,C) \qquad\qquad F = f(k,x,l)$$
$$W = f(q,G,C) \qquad\qquad U = f(k,x,l)$$
$$M = f(q,G,C) \qquad\qquad K = f(k,x,l) \quad ,$$

where *f* is the abbreviation for "function". It can be seen that

$$B = \frac{dW}{dG} \qquad F = \frac{dU}{dx} \tag{1a}$$

$$M = \frac{d^2W}{dG^2} \qquad K = \frac{d^2U}{dx^2} \tag{1b}$$

or

$$W = \int_{G_0}^{G} B \cdot dt = -\int_{G}^{G_0} B \cdot dt \qquad U = \int_{x_0}^{x} F \cdot ds = -\int_{x}^{x_0} F \cdot ds \tag{2}$$

$$B = \int_{G_0}^{G} M \cdot dt = -\int_{G}^{G_0} M \cdot dt \qquad F = \int_{x_0}^{x} K \cdot ds = -\int_{x}^{x_0} K \cdot ds \tag{2b}$$

With the Eq. (1) and (2) we are able to transform a problem of investigating the maximum (or minimum) of the potential function (objective function) into a problem of investigating the equilibrium of the corresponding virtual forces and vice versa. Under some conditions we want to investigate the maximum (or minimum) of the potential function for achieving the equilibrium of the virtual forces and under other conditions we want to investigate the equilibrium of the virtual forces for achieving the maximum (or minimum) of the potential function.

Figure 2 schematically shows the typical curves of the delay *W* and the green time demand *B* of a signal group in analogy to the potential *U* and the spring force *F* of a spring system. Both have the same characteristics as following:

(1) *B* and *F* are monotonically falling, i.e.: $f'(G)<0$

(2) *W* and *U* are strictly convex, i.e.: $f''(x)>0$

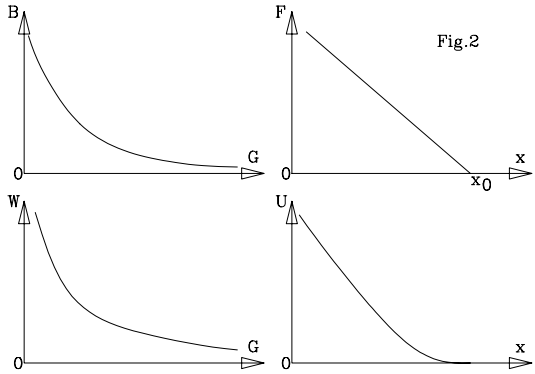


Fig.2

Figure 2. Characteristics of a green time and a spring

The second characteristic is sufficient but not necessary for the existence and uniqueness of an optimum solution and for the convergence of the optimization procedure.

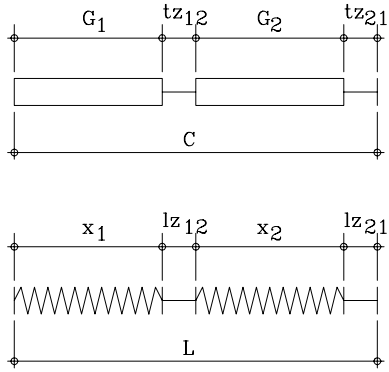


Figure 3. Signal timing plan with two incompatible and competing signal groups

According to the analogy defined above, every signal timing plan can considered as a mechanical system of springs. The springs can be linked in a serial way, one after another, or in a parallel way, side by side. The arrangement of such spring systems (analogy to signal groups in a signal timing plan) can be defined by the sequence of stages and by the restrictions (intergreens etc.)

The simplest example in this analogy is a signal timing plan with two incompatible and competing signal groups. Figure 3 shows this example. Here G_1 and G_2 are the green times of the two signal groups, tz_{12} and tz_{21} are the intergreens between the signal groups, and C is the cycle time. Respectively, the spring system can be described by x_1, x_2 (length of the springs), l_{12}, lz_{21} (inflexible connections between the springs) and L (total length of the spring system).

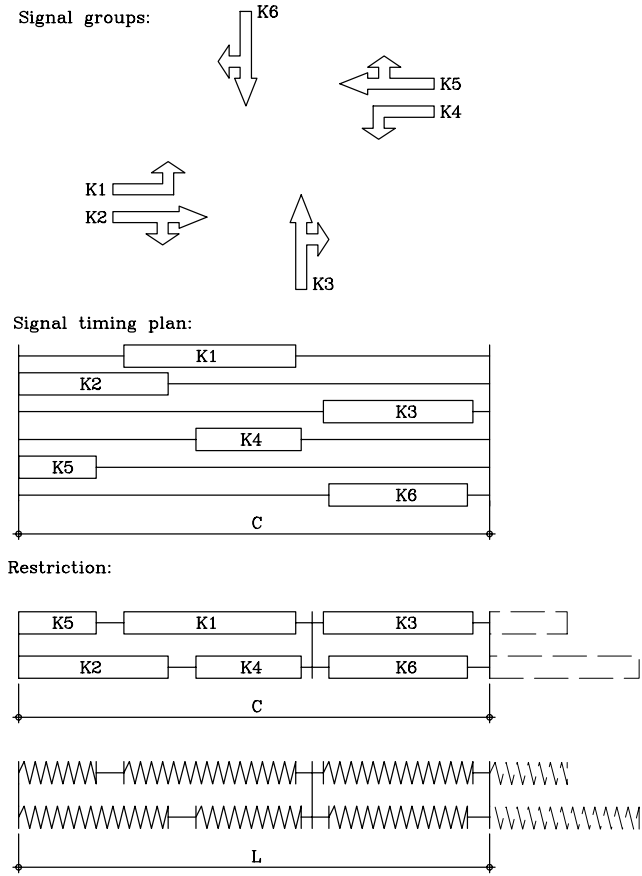


Figure 4. Simplified intersection with 6 signal groups

Figure 4 depicts a simplified intersection with 6 signal groups as a realistic example. The signal timing plan shows that the signal groups in a stage have not always the same length. For example, the signal group K1 has extended beyond the stage area and has displaced the green time of signal group K5. However, the signal timing plan is defined by the restriction of intergreens. Due to those restrictions, this signal plan can be described as a mechanical system with 6 springs.

This spring system contains serial as well as parallel connections between the springs (signal groups). The force of the springs makes the mechanical system strive for the equilibrium by itself, and thus achieves the minimum total potential (sum of

delays). The task of optimizing the signal timing plan has therefore become a task of determining the state of equilibrium of the corresponding mechanical system.

The determination of the state of equilibrium requires the knowledge of the given structure of the signal plan. The optimum signal timing plan can be found by checking all possible signal plan structures and finding out the optimum structures (stages and stage orders) with respect to the objective function. The possible signal timing plan structures can be listed e.g. according to the procedure by Tully (1976). In this paper, this procedure is not discussed in further detail. The structure of the signal timing plan is considered as predefined.

It can be proven mathematically that the objective function (the potential function U or the delay function W) has exact one minimum, if it is strictly convex in the considered interval (Allsop, 1992). Since the functions which are defined as the potential (delay etc.) during the optimization of the signal timing plan fulfil this condition, it can be assumed that a unique solution for the optimization of the signal timing plan can be found. Also by some non-convex functions the unique optimization can be found (Wu, 1999).

3 Objective Functions and Equilibrium Conditions for Minimization of the Sum of delays

With given traffic volumes and boundary conditions, the objective of the optimization is to obtain the cycle time and the corresponding green times by minimizing the sum of delays over all signal groups. The function for the delay calculation is used as the potential function. The objective function for the minimization of the sum of delays is

$$U_g = \sum_{i=1}^{n} W(q_i.G_i.C) \tag{3}$$

The force function of the signal group F_i (=green time demand B_i), of which the equilibrium is to be found, is

$$F_i = \frac{\partial W(q_i, G_i, C)_i}{\partial G_i} \tag{3b}$$

The stiffness coefficient of the signal group K_i (=M_i) is respectively

$$K_i = \frac{\partial B_i}{\partial G_i} = \frac{\partial^2 W(q_i, G_i, C)_i}{\partial G_i^2} \tag{3c}$$

The parameters in Eq. (3) are

n	= number of signal groups
G_i	= green time of the signal group i
C	= cycle time

$W(q_i, G_i, C)$ $= w(q_i, G_i, C) \cdot q_i$
$= $ delay of signal group i
$w(q_i, G_i, C)$ = average delay per vehicle of signal group i
q_i = traffic volume of signal group i

The boundary conditions for the optimization are

a) $G_i > G_{i,min}$

b) $C_{min} < C < C_{max}$

with

$G_{i,min}$ = predefined minimum green time for signal group i
C_{min} = predefined minimum cycle time
C_{max} = predefined maximum cycle time

and in case of stationary traffic

c) $s_i \cdot G_i > q_i \cdot C$

with s_i = saturation flow of signal group i

If a delay formula defined for temporary over-saturations (e.g. Akcelik, 1980 or Wu, 1990) is used, the boundary condition c) can be omitted.

Furthermore,

d) all boundary conditions which are necessary to ensure safe traffic operations, e.g. intergreens, restricted overlap of green time for permitted left turns, parallel pedestrian crosswalks with permitted right or left turners etc. (for German conditions cf. FGSV, 1992, 2003)

must be hold.

All common delay formulae (Webster, 1958; Miller, 1968; Akcelik, 1980; Kimber and Hollis, 1979; Wu, 1990 etc.) for $W(q_i, G_i, C)$ fulfil the convex condition over the interval $(0,C)$ because

(1) $\frac{\partial W_i(q_i, G_i, C)}{\partial G_i} < 0$

(2) $\frac{\partial^2 W_i(q_i, G_i, C)}{\partial G_i^2} > 0$

(3) $\frac{\partial^2 W_i(q_i, G_i, C)}{\partial C^2} > 0$.

The optimization of the objective function (Eq. (3)) can be carried out for the optimum distribution of green times $G_{i,opt}$ with a fixed cycle time C. In addition, also the cycle time C can be optimized.

4 Determination of the Equilibrium for the Optimization Procedure

According to the principle of virtual work from Dirichlet (cf. Lehmann, 1979)

A mechanical system is in its equilibrium only if the virtual change of the total potential for any virtual displacements is equal to zero, that is

$$\delta U_g = \Sigma \delta U_i = 0$$

For a mechanical system with n degrees of freedom, Dirichlet's theorem corresponds to

$$\frac{\partial U_g(x_i)}{\partial x_i} = 0 \quad i = 1 \text{ to } n \tag{7}$$

with x_i = coordinates of the i-th degree of freedom

Eq. (7) is nothing else but the condition of equilibrium of the conservative forces. If all forces F_i of this mechanical system are proportional to the coordinate x_i - as in a real spring system - a solution for the equation system Eq. (7) can be found for every force F_i and all coordinates. Unfortunately, the virtual forces (green time demand B_i) of a signal timing plan with Eq. (3), (4) and (5) as the potential function (objective) do not fulfil this condition of linearity. Most of the delay formulae are functions of higher orders. Some even contain transcendental functions. An analytical solution of Eq. (7) is therefore very difficult or even impossible to find. Up to now, optimization procedures with the sum of delays as the objective function have been carried out only numerically with an enormous calculation effort.

The state of equilibrium of a mechanical system can also be ascertained iteratively. One of the well-known procedures is the moment distribution method from Cross (cf. Beaufait, 1972) which is based on the stiffness method. This procedure can be explained by Figure 5a.

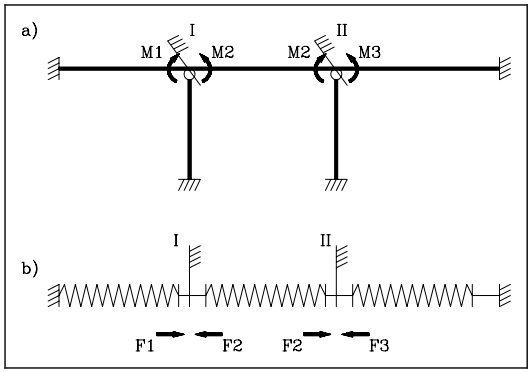


Figure 5. Principle of the moment and force distribution method

Figure 5a shows the structure of a bridge. The bending moments of the beam near the columns are to be determined. The degrees of freedom of this mechanical system are the slope deflections of the beam at both columns. First, these two slope

deflections are fastened in their original state by fictitious constraints. Putting weight on the beam causes a surplus of bending moments (*M2-M1, M3-M2*), which is fictitious as well. The bending moments are in a state of imbalance. To determine the state of equilibrium, the constraints are released alternately, i.e.:

(1) Constraint *I* is released first while constraint *II* stays fastened. Constraint *I* turns as an effect of the surplus of the bending moments *M2-M1*. When this surplus is distributed, the bending moments at constraint *I* have achieved the state of equilibrium again. The constraint *I* is fastened again. As a consequence of the turning movement at constraint *I*, constraint *II* gets a new surplus of bending moments.

(2) Then constraint *II* is released while constraint *I* stays fastened. Constraint *II* turns as an effect of the surplus of the bending moments *M3-M2*. When this surplus is distributed, the bending moments at constraint *II* achieve their state of equilibrium. Constraint *II* is fastened again. As a consequence of the turning movement of constraint *II*, constraint *I* gets a new surplus of bending moments.

(3) Steps 1 and 2 are repeated.

The bending moments at the constraints are alternately set into the state of equilibrium. The surplus of the bending moments decreases as the number of iterations increases. The iteration is stopped when the surplus of the bending moments has become small enough for practical application at all constraints. This method is also called the moment-distribution procedure. It can be used for a system with any number of constraints. All constraints are repeatedly released and fastened until all of them have achieved the state of equilibrium.

The state of equilibrium of a mechanical spring system can be determined analogously (cf. Figure 5b). Instead of the surplus of the bending moments M_i, the surplus of the spring forces at the connections F_i is distributed. The connections are first fastened and then released alternately (which causes horizontal displacements). Then they are fastened until the surplus of the spring forces (*F2-F1, F3-F2*) falls below a certain minimum at the connections. This procedure can be called the force-distribution procedure.

The force-distribution procedure does not start with the determination of the state of equilibrium for the total system, but with that of the individual constraints. In a signal timing plan as an analogy to a system of springs, the constraints are clearly defined by the restrictions of the signal timing plan (intergreens, permitted left turners or minimum green times etc.). Thus, a system with n springs (analogue to n signal groups) is simplified as a system with m constraints. In most of cases, m is always smaller than n. The example in Figure 4 has $n = 6$ signal groups, but only $m = 3$ constraints if the cycle time is fixed. If the cycle time C is regarded as a variable as well, this example has $m = 4$ constraints.

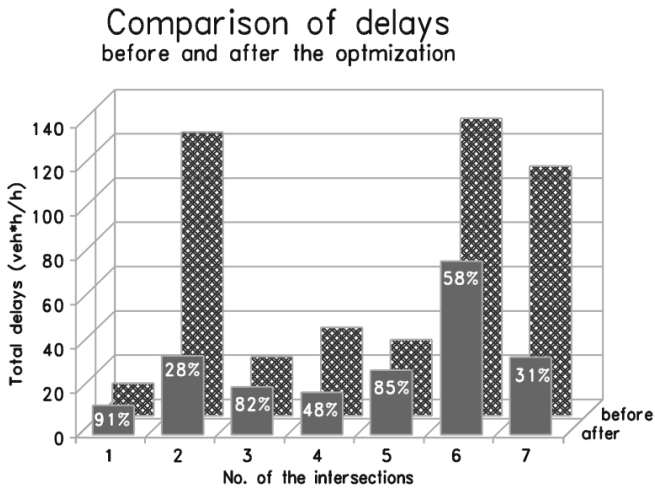


Figure 6. Optimization results for 7 intersections

The new procedure was tested at 7 intersections in Düsseldorf, Germany. The local authorities provided the existing signal timing plans. The existing signal times were used as a basis for the optimization. The result is presented in Figure 6. It shows that the signal timing plans, some of which were several years old and for which the traffic volumes have undoubtedly changed in the meantime, need to be improved. The intersections 2 and 7 were controlled by so-called advanced signals. Here, the green times of some movements were intentionally restricted to limited capacity (capacity < demand) with the purpose of keeping congestion out of the city area. Therefore, the before/after comparison is irrelevant for these two intersections. For the other 5 intersections, improvements could be achieved. The significant improvements at intersections 4 and 6 are results of overloaded left turning traffic streams.

5 Conclusion and outlook

The optimization procedure for a signal timing plan according to the principle of equilibrium has the following advantages compared to the existing procedures:

- Fast computation due to reduction of variables (in place of number of signal groups n the number of constraints m is used). An optimization procedure for a fixed-time controlled standard intersection normally takes less than 1s.
- Minimum work before the optimization (only the signal timing plan and the corresponding intergreens and the traffic volumes are needed).
- Applicable for any convex objective functions.

- Continuity of the signal timing plan during the optimization (the signal timing plan is not produced totally anew but is only developed in an innovative way).
- Dynamic character (adaptation to changed traffic volumes, minimum green times etc. is possible during optimization).
- The procedure can be influenced manually (new definition of green times, reduction of cycle time etc.).

Regarding the general theoretical background, this procedure can also be easily transferred to the following traffic control strategies:

- Signal timing control with consideration of pedestrian delays
- Signal timing control with special stages for public transport
- Adaptive signal timing control
- Coordinated signal timing control in networks

In the future, the new model will be formulated in details.

References

Akcelik, R. (1980). Time-dependent expressions for delay, stop rate and queue length at traffic signals. Australian Road Research Board, No. 361-1, 10/1980.

Allsop, R.E. (1971). Delay-minimising settings for fixed-time traffic signals at a single road junction. Journal of the Institute for Mathematics and his Applications, 8 (2), 164-185.

Allsop, R.E. (1992). Evolving Application of Mathematical Optimisation in Design and Operation of Individual Signal-Controlled Road Junctions. In: Mathematics in Transport Planning and Control, edited by J.D. Griffith, Clarendon Press, Oxford.

Beaufait, F.W. (1972). Basic Concepts of Structural Analysis. Prentice-Hall, Inc. Englewood Cliffs, New Jersey 07632.

Brilon, W. and Wu, N. (1990). Delays at fixed-time traffic signals under time-dependent traffic conditions. Traffic Engineering and Control. 12/1990.

FGSV (1992, 2003). RiLSA - Guidelines for Traffic Signals - English Version of with minor Modifications. FGSV-Verlag, Köln.

Kimber, R.M. and Hollis, E.M. (1979). Traffic queue and delays at road junctions. TRRL Laboratory Report, LR 909.

Lehmann, Th. (1979). Elemente der Mechanik IV: Schwingungen, Variationsprinzip. Friedr. Vieweg & Sohn, Braunschweig/Wiesbaden 1979.

Miller, A.J. (1968). Signalised intersections - capacity quite. Australian Research Board Bull, No. 4/1968. Reprinted as ARRB Research Report ARR No. 79/1978.

Möller, K. (1987). Signalgruppenorientiertes Modell zur Optimierung von Festzeitprogrammen an Einzelknotenpunkten. Schriftenreihe des Instituts für Verkehrswesen der Universität Karlsruhe, Heft 37.

Rose, R. (1972). Foundations of Mathematical Biology. ACADEMIC PRESS, New York and London.

Silcock, J.P.; Sang, A. (1990). SIGSIGN: a phase-based optimisation program for individual signal-controlled junctions. Traffic Engineering and Control, 5/1990.

Tully, I.M.S.N.Z. (1976). Synthesis of sequences for traffic signal controllers using techniques of the theory of graphs. OUEL Report 1189/77, University of Oxford.

Webster, F.V. (1958), Traffic signal settings. RRL Technical Paper No. 39, HMSO, London.

Wu, N. (1990). Wartezeit und Leistungsfähigkeit von Lichtsignalanlagen unter Berücksichtigung von Instationarität und Teilgebundenheit des Verkehrs (Delay and Capacity at Signalised Intersections under non-stationary and partly-bunched traffic conditions). Schriftenreihe des Lehrstuhls für Verkehrswesen der Ruhr-Universität, Heft 8 (A short English version is given in Brilon, Wu, 1990).

Wu, N. (1999). Optimierung von Signalzeitenplänen nach dem Gleichgewichtsprinzip (Optimization of Signal Timing by Equilibrium Technique). Schriftenreihe des Lehrstuhls für Verkehrswesen, Heft 20, Ruhr-Universität Bochum.

Simulation Research about the Number of Lines in a Passenger Rolling Stock Servicing Yard on ARENA

ZHOU, Lixin[1] and YANG, Xiongbin[2]

[1]Transportation Engineering School of Tongji University, Shanghai 201804, China; PH (021) 68589372; FAX (021) 69599372; email: lxzhou@mail.tongji.edu.cn
[2]Transportation Engineering School of Tongji University, Shanghai 201804, China; PH (021) 69589171; FAX (021) 69589372; email: yxb1983@gmail.com

Abstract: At the design phrase,it is difficult to determine a reasonable number of tracks of passenger rolling stock in a servicing yard. This article focuses on a new idea of simulation technology to find a solution. Using ARENA software, a simulation model for handling passenger train maintenance and overall preparation services is established on the basis of a fixed position working procedure. Several modules are developed and statistical indexes are designed. Finally, an operational evaluation of the design of a 14 track yard is made with a practical case of the passenger rolling stock servicing yard of Shanghai-South Railway Station. Fifteen or more tracks should be equipped with suitable service groups to synchronize all work.

Key words: *passenger rolling stock servicing yard, track, simulation, Arena*

1 Introduction

Passenger rolling stock servicing yards (briefly named "PSY") are important parts of the infrastructure for guaranteeing the safe operation of passenger trains. The main functions of a PSY are as follows:

(1) Technical checking, maintenance, cleaning, etc. for serviceable carriages in advance;

(2) Service preparation for departing passenger trains;

(3) Provision of a temporary stop yard (tracks) for carriages belong to the passenger rolling depot.

The number of tracks designed for the above three tasks in a PSY is the main factor to decide the scale of a PSY. This problem is related to the number of trains in the rolling stock depot, the procedures in the PSY (moving or fixed position working procedure), the location of the PSY relevant to a passenger station and a passenger rolling stock depot, the timetable of passenger trains, etc. The latest “Railway Station and Terminal Design Criterion” (GB-50091-2006) does not give a clear number of tracks in a PSY. Though the “Rail Engineering & Techniques Manual” (No 4.Railway Design and Research Institution, 2004) has listed a table based on train pairs in/out of a PSY (given in Table 1), the reference values do not agree with the actual circumstance. For

instance, in the PSY equipped for the Shanghai railway station, the daily in/out train pairs are up to 100, and the existing number of tracks in the PSY is only 16, which is far lower than the recommended value in Table 1.

Table 1. Reference number of tracks in passenger rolling stock yard

Trains in/out(pairs per day)	6～7	8	9～10	11～13	14～15	16～18	19～22	23～26	27～29	30～31	32～34	35～37	38～40	41～43	44～46	47～49	50～52
Number of tracks	5	6	7	8	9	10	11	12	13	14	15	17	17	18	19	20	21

A reasonable design scheme should provide the minimum number of tracks for the technical checks and service tasks of passenger trains without delay. On the other hand, any ineffective investment derived from an excess number of tracks should be avoided. In addition to the technical checking of a train, various operation schedules of passenger trains also have an important impact on the take-up time of the tracks. It is very difficult to decide the scale of a passenger rolling stock depot with a simple calculation formula.

Because PSYs differ in the number of tasks as well as in the content and demand of tasks, it is fruitless to determine the number of tracks in a PSY only according to Table 1. Therefore a new idea with a simulation technique is discussed in the paper to seek for a valid solution to this kind of problem. A case-study of the Shanghai-South railway station is studied in detail in this paper.

2 Model Conceiving

2.1The technical task in PSY

There are various types of rolling stock belonging to a depot and the different carriages may have different task procedures. Generally, they could be divided into three types based on technical procedures: [1]

(1) Type A train

A train does not require a check of running parts of carriages with a ditch and maintenance for air-conditioning package, e.g. type-22 carriages.

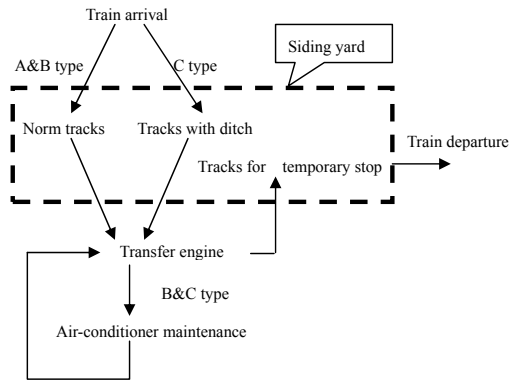


Figure 1. An operations chart of a PSY

[1] Here we simply define any carriage groups in PSY as trains.

(2) Type B train

A train requires maintenance for air-conditioning except checking running parts of carriages with a ditch, e.g. type 25K and 25G carriages.

(3) Type C train

Both tasks above are necessary for a train, e.g. type BSP carriages.

For fixed position working procedure, entering PSY trains are hauled into different siding tracks according to their types for checking running parts and equipment, and a check of the appearance of carriages or maintenance of other devices. For example (shown in Figure1), type A and B trains are first hauled into norm tracks; type C trains are hauled into tracks with ditches which are supplied with the equipment for checking the running parts of carriages. For air-conditioning package maintenance, type B and C trains are then transferred onto an air-conditioner maintenance shed. Finally, all of them are transferred onto a temporary stop yard waiting for departure.

2.2 Modeling Principle

The tasks of a train in/out of a PSY are special technical services. If we refer to the trains as "customers", and the checking groups, tracks and maintenance groups as a "service station", then the service procedure of the train can be abstracted as a typical queuing system. The characteristics of this system are as follows:

(1) The headways of arrival time of trains are asymmetric.

(2) The time used for check and maintenance of each train is not the same.

(3) The time-consuming length to use a track is uncertain.

Under the comprehensive impact of the above three random factors, there likely is a train queuing phenomenon in PSY which could seriously influence the entrance of the arrival trains from a passenger station. Considering the construction cost, it is the core goal of PSY design to decrease or ultimately eliminate train queuing time as far as possible within a limited investment scale.

2.3 Model Design

2.3.1 The Steps of Modeling Based on Arena

Arena is object-orientation simulation software. Aimed at the requirement of process simulation of PSY work, the main steps of modeling are as follows:

(1) Affirm the makeup and number of trains arriving at a PSY per day.

(2) Investigate the service station amount and service procedures according to Figure1.

(3) According to actual investigation and statistics analysis method, define the arrival interval of train flow, the probability distributions of each service station's work time and corresponding confidence interval.

(4) Establish Arena model and set up the service parameters of corresponding service stations. For different procedures (such as type A, B and C trains), the

relationship between service stations is regarded as parallel and their tasks can be done simultaneously; For the same procedures (such as the same type of train),the relationship between service stations is regarded as series-wound and their tasks are done in succession.

(5) Set up control condition of simulation, such as the times of the model running, the period of each simulation, the amount of train checking groups and tracks, etc.

(6) Design input and output of simulation. A simple and transparent interface for input/output data is designed in the simulation program. The contents of input and output include the utilization of each service station (track), the load of each checking group, the amount of delayed trains in the system, etc.

(7) Analysis of simulation result. If the distinction between the current goal and simulation result is in the allowable range, the scheme is feasible, and the optimal scheme can be picked out among several feasible schemes.

2.3.2 The features of Modules

(1) Train producing module

Regarding train arrival as a stochastic process, a stochastic vector X is defined to represent it. （x_1， x_2， …， x_n）is a sample from collectivity X, X～E(λ)，λ＞0. The scale of a sample is decided by the period length of simulation. A new train is created based on exponential stochastic distribution.

(2) Assigning module

Each arriving train is evaluated with a special arriving probability value and three attributes must be decided as follows:

- Assign a type ID to each arriving train according to probability. The probability of type A, B or C train respectively is P_1, P_2, P_3, and $\sum_{i=1}^{3} P_i = 1$, the distribution function is F(X):

$$F(X)=P(x\le X)=\begin{cases} 0 & X\in(-\infty, 0] \\ P_1 & X\in(0, 0.1] \\ P_1+P_2 & X\in(0.1, 0.5] \\ 1 & X\in(0.5, 1.0] \end{cases} \quad (1)$$

- Assign an operational flow ID to each arriving train before the simulation starts, to control which services and what sequence of the train should be done. A system procedure control variable Sequence [N] (N□ (A, B, C)) is designed with a value of a rank statistic. Assumption: rank statistic 1 or 2 or 3 or 4 respectively represents a train required to be checked /overall prepared on a norm track or on a track with ditch or on a track for temporary stop or at an air-conditioning package maintenance shed; rank statistic 5

represents a transfer by a shunting locomotive. That is: Sequence[A]＝(1,5,3), Sequence[B]=(1,5,4,5,3)；Sequence[C]=(2,5,4,5,3).

- Assign an initial time to every arrival train, then calculate the service time or delay time of each working procedure according to the train arrival interval. Considering every task performed on tracks, a "series & shunt-wound" sub-system is designed. When a train service group is free, an arriving train is directly hauled into "a service station" (a track) and the simulation system automatically records the take-up times of both the train service group and the tracks. When all the train service groups are busy, an arriving train is first hauled into a temporary stop track waiting for service. Both waiting time and train service time all are recorded automatically by the simulation system.

(3) Service station module

There are six sub-modules. According to the content of tasks in the PSY, there are five types of service stations: ditch track service, norm track service, air-conditioning package maintenance service, transfer shunting service, and temporary stop service. All services except for transfer shunting represent corresponding tracks. The time of a train service obeys certain probability distributions. Assume service time as a stochastic variable Y, through fitting history data，norm track service, ditch track service, air-conditioning package maintenance service and transfer shunting service; all can be represented by triangle distributions, that is Y～Triangle (Min, Mid, Max) (min)(E. de Jong, 1997). Other temporary stop service times obey normal distribution, that is Y～Norm(μ,σ^2) (min), in light of the impact of a passenger train timetable.

3 ARENA Simulation Model of a PSY

3.1 Main Functions

Combined with the procedure of a PSY, a simulation program is developed by the ARENA tool. The main functions are listed in Table 2.

3.2 Indexes of Simulation

There are three main indexes defined in the simulation model:

- The utilization ratio of tracks. With the objectives of keeping the rhythm of tasks in the PSY, the flexibility to handle sudden incidents and the availability of equipment, the reasonable utilization ratio of tracks should be in the range of [50%, 70%].
- The load ratio of service groups. In order to avoid having service personnel too tired or confused to deal with multiple train arrivals, the load of service groups should not be too heavy. Based on investigation, the ratio should be in the range of [50%, 60%].
- The amount of trains delayed in a system. When trains arrive at the PSY, in addition to valid maintenance (or overall preparation) tasks, the lack of equipment and service personnel would influence the delay time of trains. In

principle, the maximum number of trains delayed should not exceed the number of tracks in the PSY; otherwise it is unreasonable to increase waiting time of a train or use other tracks as temporary stop tracks.

Table 2. Function of each sub-model

Sub-model	Function
Train arrival	Create trains according to a certain time interval, assign type ID (needle) to the trains, and send them into different service schedules.
Acceptance handling	Make a statistical table of take-up time of each track (or service station), according to the type of train.
Check or air-conditioning package maintenance	Make overall preparation or air-conditioning package maintenance for type B or C trains, and record the working time of each service group
Train shunting	There are two circumstances: (1) A type B or C train is transferred to an air-conditioning maintenance shed; (2)After overall preparation work, all types of trains are transferred to temporary stop tracks. Record the working time of the shunting locomotives.
Train waiting for departure	Record temporary stop time of a train waiting for departure instruction.
Simulation data analysis	Calculate all delay time of trains and all working time of each service group or station (or tracks) in the PSY. Cease the current simulation program to stop condition and go on next running cycle.

4 Case Analysis – Shanghai-South Railway Station

4.1 Background

According to the blueprint data (No 4. Railway Design and Research Institution, 2002), Shanghai-South railway station would accept 63 pairs of passenger trains per dayin which there are 25~33 pairs of passenger trains to be overall prepared in a PSY. Both passenger rolling stock depots and locomotive depots are located in the east throat of the passenger station. There are 14 tracks planned in the PSY.

4.2 Initial Parameter of Simulation

(1) Trains and their tasks

Types A, B and C respectively have a proportion of 10%, 40%, 50% in total (shown as Table 3). The maximum number of trains arriving at the PSY per day is 33. The mean arrival interval (λ) of trains is 43 min. Corresponding probability distributions of service time are shown in Table 4.

Table 3. Types of trains and arrival probability

Train type	Ordinary	Rolling parts	Air-conditioning package	Probability
A	√			10%
B	√		√	40%
C		√	√	50%

Table 4. Probability distributions of each task Unit：min

Types of service	Probability distribution
Ordinary check	Tria(80,100,120)
Wheelbase check	Tria(80,100,120)
Air-conditioning package maintenance	Tria(60,90,110)
Shunting transfer	Tria(15,20,25)
Temporary stop	Norm(200,1)

(2) Schemes comparison

Trains are towed from Shanghai-South Railway Station, the appearance cleaned on cleaning tracks, and sent onto corresponding tracks for technical tasks and overall preparations. According to the demand of design, the simulation schemes of tracks and service groups are shown in Table 5.

(3) Period of simulation

The period is defined as 24 hours and the program is stopped after 10 repeats.

Table 5. Schemes of tracks in a PSY

Schemes	12 tracks	13 tracks	14tracks	15 tracks	16 tracks
Norm tracks	3	3	4	4	4
Tracks with ditch	3	4	4	4	4
Air-conditioning package maintenance tracks	2	2	2	2	3
Temporary storage tracks	4	4	4	4	5
Norm tracks checking groups	2	2	2	2	2
Tracks with ditch checking groups	3	3	3	3	3
Air-conditioning	2	2	2	3	3
package maintenance groups					

Note: one shunting locomotive equipped in the PSY

4.3 Result Analysis

The statistical indexes by the simulation are shown in Table 6.

Table 6. Utilization ratio of equipment and load ratio of service personnel

Options	12 tracks	13 tracks	14 tracks	15 tracks	16 tracks
Trains delayed: number	16.9	16.1	15.6	13.9	13.6
Norm tracks	0.565	0.595	0.428	0.407	0.311
Tracks with ditch	0.486	0.519	0.542	0.448	0.597
Air-conditioning package maintenance tracks	0.689	0.620	0.622	0.427	0.395
Temporary stop tracks	0.521	0.666	0.677	0.689	0.512
Norm tracks service personnel groups	0.654	0.669	0.629	0.621	0.481
Tracks with ditch service personnel groups	0.461	0.559	0.550	0.497	0.586
Air-conditioning package maintenance service groups	0.689	0.620	0.622	0.427	0.395
Shunting locomotive	0.669	0.612	0.606	0.624	0.550

The simulation results are analyzed as follows:

- The increase of norm tracks and tracks with ditch (e.g.12~14 tracks) doesn't influence the amount of delayed trains obviously; the main cause of the delays is that the limited train service groups would produce an additional waiting time in a peak period. The amount of train service groups should be matched with train arrival density. The number of delayed trains decreases with the addition of an air-conditioning package maintenance service group.
- Considering schedules of ordinary train checks, air-conditioning package maintenance and the dwell time of departure trains, the number of tracks of the three sections should match each other so as to make an equilibrium among schedules and decrease the probable waiting time. As shown in Table 6, 13~16 track schemes work very well.
- The load ratios of schemes with one shunting locomotive are all below 0.7 which is a reasonable range.
- In Table 6, the number of tracks in the PSY of Shanghai-South Railway Station should be up to 15 tracks. Comparing schemes of 15 tracks to 16 tracks, the latter has a superfluous capacity, such as the load ratio of norm

tracks and air-conditioning package maintenance tracks , both of which are below 0.4.

In summary, the PSY of the Shanghai-South Railway Station should add one more track in addition to the existing design scheme.

References

E. de Jong(1997). "Software architecture for large control systems: a case study ." Proc. of COORDINATION '97, LNCS, volume 1. Springer Verlag.

No 4. Railway Design and Research Institution(2002). *The preliminary design of Shanghai-South Railway Station.* Wuhan, China.

No4. Railway Design and Research Institution(2004). *Rail Engineering & Techniques Manual.* Beijing: China Railway Press.

PRC industry standards (GB-50091-2006). *Railway Station and Terminal Design Criterion.* Beijing: China Plan Press,2006.

Seydel, R(1994). "Practical Bifurcation and Stability Analysis: From Equilibrium to Chaos". Springer–Verlag, New York.

Serve Sequence Optimization Approach for Multiple Bus Priority Requests Based on Decision tree

Wanjing MA[1] and Yu BAI[2]

[1]ITS Research Center of Tongji University, 1239# Siping Road, Shanghai, P.R.China
PH :86-21-65537571, FAX : 86-21-65537571, email :mawanjing@gmail.com
[2]ITS Research Center of Tongji University, 1239# Siping Road, Shanghai, P.R.China
PH :86-21-65537571, FAX : 86-21-65537571, email : rabbityu@263.com

Abstract: Transit signal priority is a key method to improve the reliability and level of service of transit systems. Under the condition of multiple priority requests, the serve sequence dramatically affects the efficiency of transit signal priority strategies. The multiple priority requests are classified into two kinds: multiple requests for single phase and multiple requests for multi phases. The decision tree is used to optimize the service sequence of these two kinds of multi requests. Minimizing the average person delay of all priority requests is adopted as decision objective. The results of a case study show that the conventional response sequence, First-Come, First-Served, is not guaranteed to be the best solution. Based on the decision objective of minimizing the average person delay of multi requests, the decision tree method can generate optimal serve sequence.

Key words: *Bus signal priority; Multiple priority requests, Decision tree, Average person delay*

1 Introduction

Many traffic signal control systems have bus priority logic already embedded in their software. Such systems provide signal priority at the local or the systems level. The SCATS system (Lowrie, 1982) transit priority logic includes green extension, special phase sequences and compensation to the non-transit phases. The SCOOT system (Hunt P.B. et al, 1981) grants priority to buses (phase extension and recall) based on user specified intersection degree of saturation to avoid excessive delays for the rest of the traffic. Field evaluations in London showed bus delay savings ranging from 5 to 10 seconds/signal with no disruption to the rest of the traffic. Turin's UTOPIA system (Nelson, J.T, 1993) is an adaptive control system that provides

absolute priority to transit by continually optimizing the signal settings over a short time interval (rolling horizon). Reported benefits include a 20 percent increase in the average bus speeds without disruption to the rest of the traffic. Beside these, there is much literature (Alexander, 1998) on transit signal priority theories and strategies.

One problem with existing bus priority algorithms is few of all the bus priority strategies discussed how to handle situations when buses arrived simultaneously on conflicting approaches. It was assumed in most priority techniques that there were buses only on major roads and in the peak direction (Meenakshy, 1999). It is assumed that there can be one or more requests active at any time and that priority requests are not, by policy, handled on a first-come, first-served basis (Larry, 2006). Under the condition of high bus volume and complex bus network which is very common in Chinese cities, the situation of multiple priority requests at the same time is ubiquitous.

A Coordinated and Conditional bus Priority (CCBP) approach framework has been proposed to solve the priority control problem for adjacent intersection groups with the condition of heavy bus volume by the authors (Wanjing MA, 2007). This paper is an extension of that research. A decision tree method for solving multiple priority requests (MPR) problem is proposed. The MPR problem is divided into two kinds: multiple requests for single phase and multiple requests for multiple phases. Minimizing average person delay of all priority requests is adopted as decision objective. A numerical example validates that the proposed decision tree method can decrease average person delay of total requests dramatically.

2 Basic Concept

In order to find the best serve sequence for multiple priority requests, a decision tree method is proposed. Four key problems are discussed:

(1) Alternative priority control strategies and constrains. In order to avoid interrupting the signal coordination between adjacent intersections and congestion no-priority phases, usually there are only few strategies that could be used and some constraints should be followed. These affect the optimal serve sequence dramatically. It is also very important to select a best priority strategy with these constraints.

(2) The characteristic of multiple priority requests. The relationship between the priority requests should be analyzed. Some priority requests may be served to certain degree or denied while a serve sequence is selected. The effect of each

serve sequence to each priority request should be considered in the decision making process.

(3) Design of decision tree of each kind of multiple priority requests. This is very important to implement the method.

(4) Selection of decision objective. The criteria of optimal serve sequence should be studied. Bus delay, bus schedule delay, number of passengers and so on should be taken into considered.

Figure 1 shows the total flowchart of the proposed method. Only a "late" bus can generate a priority request. It is assumed that the arrival time of all priority requests can be predicted and then the decision should be made before the start of next signal cycle. The optimization of the serve sequence is also a process of signal priority strategies optimization.

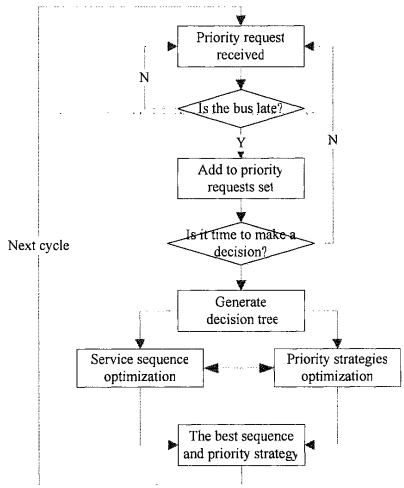


Figure 1. Optimization logic

The following three parts of this paper were developed based on the functional requirements to solve the four key problems listed above. Firstly, alternative bus signal priority strategies are introduced, and the delay calculation models are also given. Secondly, decision tree methods for multiple priority requests are discussed. Finally, a numerical example is given and the proposed decision method is tested and validated.

3 Alternative Bus Signal Priority Strategies and Bus Delay Calculation Model

Three methods, Green extension, phase insertion and red truncation, are proposed as shown in Figure 2. Different priority method educes different bus delay. Let t_{ba} represent bus arrival time and for a given intersection, g_{ge} represent maximum green extension time, g_{rt} represent maximum red truncation time, g_{in} represent the length of insertion phase, g_i represent the green length of ith phase.

Delay of Green Extension d_{ge} For a given intersection, only if $g_i < t_{ba} \le g_{ge} + g_i$, green extension can be selected. After a successful green extension, the bus can pass the stopline at green signal, that means d_{ge} =0.

Delay of Phase Insertion d_{in} For a given intersection, if $g_i < t_{ba} \le g_{ip} + \sum_{j=1}^{p-1} g_{ij} + I_{ij}$ green extension can be selected. A special bus phase can be inserted after phase p or between the rest phases as shown in Figure2. If the bus arrives at stopline after the insertion phase starting, the bus delay is 0 second, if the bus arrives at stopline before the insertion phase starting, the bus delay can be calculated as:

$$d_{in} = t_{in} - t_{ba}$$

Where，t_{in}: Starting time of the insertion phase, (s),

There are several time windows between different phases when a bus phase can be inserted as shown in Figure 2. The most appropriate insert time should be decided by the model built in the next part.

Delay of Red Truncation d_{rt} For a given intersection, if $g_i < t_{ba} < C$, red truncation can be selected. It is same as the phase insertion method: if the bus arrives after the red truncation time, the bus delay is 0 second, if the bus arrives at stopline before the red truncation time, the bus delay can be calculated as:

$$d_{re} = t_{rt} - t_{ba}$$

Where, t_{rt}: Red truncation starting time, (s).

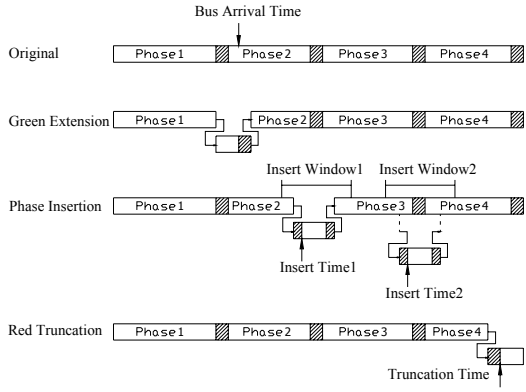


Figure 2. Alternative bus signal priority strategy

4 Design of Decision Tree For Multiple Priority Requests

4.1 Two Kinds of Multiple Priority Requests

According to the attributes of the applied phase, the multiple priority requests can be allocated into two kinds, multiple priority requests for single phase (MPRS) and multiple priority requests for multiple phases (MPRM) as shown in Figure 3.

Let B_j^i represent the i^{th} priority request for j^{th} phase. For multiple priority requests for single phase, as shown in Figure 2 (up), two priority requests, B_1^1 and B_1^2, both of them request signal phase one. It is obviously that B_1^1 would be served at the same time if B_1^2 is served. However, if the B_1^1 is served, B_1^2 would be denied in one signal cycle.

For multiple priority requests for multiple phases, as shown in Figure 2(down), three priority requests, B_1^1, B_3^2, B_4^3, respectively apply for phase 1, phase 3 and phase 4. All requests are independently done, one request could not benefit from the service of a request.

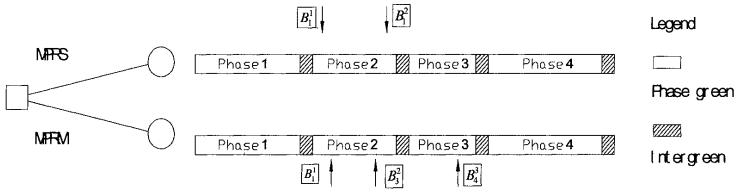


Figure 3. Two kinds of multiple priority requests

4.2 Design of Decision Tree

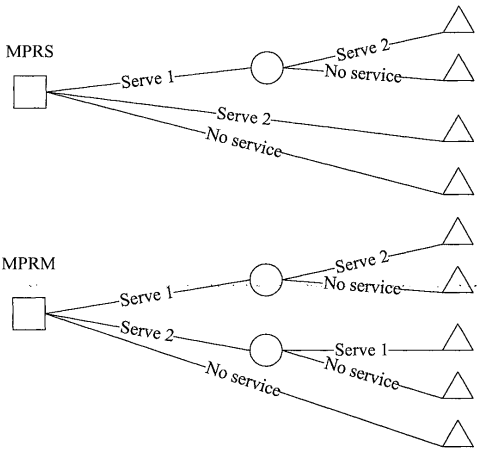


Figure 4. Decision trees of two kinds of multiple priority requests

According to the characteristics of each kind of multiple priority requests, the decision tree can be designed as follows:

(1) The original signal plan is taken as root node of the decision tree,

(2) Every whole branch represents a serve sequence of multiple priority request,

(3) The weight of every whole branch is the average weighted bus schedule delay.

(4) The decision objective is average weighted bus schedule delay as explained above.

Take two signal phases control signal plan as an example, the decision of MPRS and MPRM is shown in Figure 4.

4.3 Decision Objective

The three main factors which affect the urgency of a priority request are bus delay, schedule delay and number of passengers. The adverse effects on no-priority phases or coordinated signal plan of adjacent intersection groups also should be taken into considered.

Bus schedule delay: transit priority should not be provided if buses may result being ahead of schedule. So the strategies proposed in this paper provide priority to only those buses that are behind schedule. That means the all the priority requests are "late" buses.

Bus delay: the potential delay of a bus at intersection with different signal plan including priority signal plans and pre-priority signal plans should be estimated. It is the base of making a decision to give priority to a bus.

The number of passengers: favoring a "late-runner" bus may not be beneficial if it is empty. So the number of passenger should be taken into consideration..

Based on the analysis above, the decision objective for multiple priority requests can be presented as a mathematical programming formulation as:

$$\text{Minimize } D = \sum_{i=1}^{n} Q_i d_{si} \Big/ \sum_{i=1}^{n} Q_i$$

Where, D: Average weighted bus delay,

Q_i: Number of passengers of priority request i,

d_{si}: Bus schedule delay of request i.

This decision objective is equal to minimize average person delay of all "late" bus. Two factors affect the delay of buses: serve sequence and priority strategies. The optimization of two factors is integrated into a same process.

4.4 Decision Constrains

Priority frequency constraints: There is a constraint to limit the priority strategies selection: Only one phase insertion strategy can be adopted in one cycle. Two of all priority methods can be adopted simultaneously in one cycle at most. This constraint is used to avoid the undesirable adverse effect of priority strategies on no-priority phases or coordinated signal plan.

Cycle length constraint: The cycle length constraint is used to keep signal progression of the intersection group. It is a simple equation constraint, and for a given intersection i, it can be shown as follow:

$$\sum_{j=1}^{M} g_j + I_j = C$$

Where, M : phase number after bus priority of the intersection,

I_j : green interval between jth phase and the next phase.

The phase duration constraint: It is used to ensure pedestrian safety and to maintain the level of service of other motor vehicle. It would be reasonable to assume that the traffic engineer can set the initial minimum and maximum green time. For a given intersection i, the following non-equation constraint should be satisfied:

$$g_{j\min} \le g_j \le g_{j\max}$$

Where, $g_{j\min}$: Minimum green of j^{th} phase,

$g_{j\max}$: Maximum green of j^{th} phase.

5 Evaluation

A four phase signal controlled intersection is adopted as an example to test the proposed decision tree approach. In the original signal plan, cycle length is 100 seconds, intergreen is 3 seconds, minimum and maximum green time of each phases is 15 seconds and 35 seconds. Maximum green extension time (red truncation time) is 10 seconds. Two situations are tested:

Situation A: one bus (with 30 passengers) arrives at the intersection of 30th and Second, another bus (with 50 passengers) arrives at the intersection of 38th and Second. Both of them apply for signal phase 1.

Situation B: one bus (with 30 passengers) arrives at the intersection of 24th and Second and applies for signal phase 2, another bus (with 50 passengers) arrives at the intersection of 29th and Second and applies for signal phase 1. The original signal plan and priority requests status are shown in Figure 5.

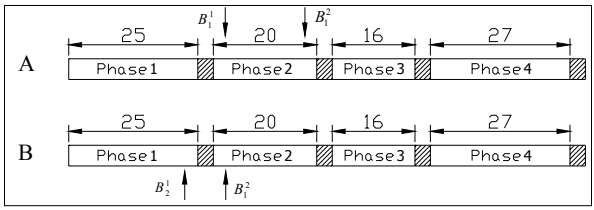


Figure 5. Original signal plan and priority requests status

Based on the method proposed above, decision trees are generated and relative priority strategies are selected as shown in Figure 6.

Figure 6 shows the effects of different serve sequence on every request and average weighted bus schedule delay. Buses can benefit dramatically from signal priority in contrast to no priority situation. The benefit of giving priority to buses is changing with the change of serve sequence. Under the optimal serve sequence and priority strategy, the average delay is 2.25 (situation A) and 2.625 (situation B). The first-come, first served sequence does not guarantee best performance (in this example, it is not the best sequence obviously. the person delay is 8.125 for situation A and 30.625 for situation B). The average person delay of the decision tree method is lower 72.3 percent at situation A and 91.4 percent at situation B than that of first-come, first served rules. The best priority strategies is to insert phase 1 at 36 seconds for situation A and extension phase 1 to 29 seconds for situation B.

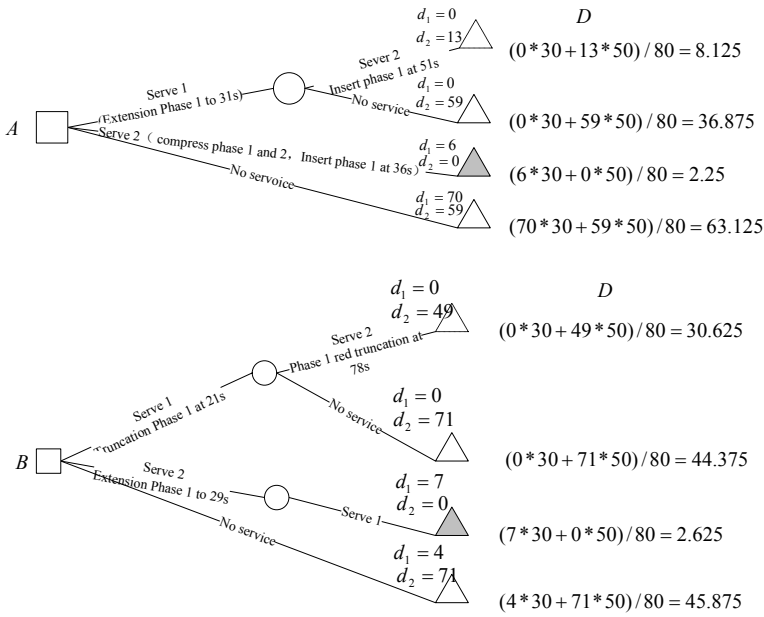


Figure 6. Decision trees and selected priority strategies

6 Conclusion

For the multiple priority requests situation, the serve sequence is one of several key factors that affect the efficiency of bus signal priority. The results of this study have shown that the approach of decision tree could be developed. The proposed decision tree method generates better sequences as well as better signal priority strategies. And the first-come, first served method hsd proved that it does not guarantee the best performance.

Although this study is limited to the efficiency of bus system, it nevertheless demonstrates the improvements that can be gained by decision method. In addition, while the study provides an applicable solution to overcome the problems posed by multiple priority requests, further studies are required to address the usefulness of the proposed method after motor vehicle movements are taken into considered.

Acknowledgments

This study is supported by The Hi-Tech Research and Development Program of China (863 Program, Grant No.2006AA11Z205)at the Institute of Transportation Engineering of Tongji University.

References

Lowrie, P.R., The Sydney Corrdinated Adaptive Traffic System SCATS)-principles, methodology, algorithms," International Conference on Road on Road Traffic Signalling, London, 1982.

Hunt P.B. et al., "SCOOT, a traffic responsive method of coordinating signals," Transport and Road Research Laboratory, TRRL laboratory report 1014, 1981.

Nelson, J.T., et al, "Provision of Priority for Public Transport at Traffic Signals: A European Perspective," Traffic Engineering & Control, Vol 34(9), 1993.

Vincent, R. A., B. R. Cooper, and K. Wood. Bus-Actuated Signal Control at Isolated Intersections-Simulation Studies of Bus Priority. Transport and Road Research Laboratory Report 814, Crowthorne, U.K., 1978.

Alexander Skabardonis. Control Strategies for Transit Priority(UCB-ITS-PRR-98-2), 1998

Hongchao Liu, Alexander Skabardonis, and Wei-bin Zhang, A Dynamic Model for Adaptive Bus Signal Priority, TRB, CD-ROM, National Research Council, Washington, D.C. January 2003

François Dion, H. Rakha. Integration of Transit Signal Priority With in Adaptive Traffic Signal Control Systems. 84th Annual Meeting of the Transportation Research Board, Washington, D.C., January, 2005

Meenakshy Vasudevan, Gang-Len Chang. Design Framework for Integrating Real-Time Bus Priority Control with Robust Arterial Signal Progression. Transportation Research Board, January 2000, Washington D.C. November 17, 1999. pp: 2.

Larry Head, Douglas Gettman and Zhiping Wei. A Decision Model for Priority Control of Traffic Signals. Transportation Research Board. Washington D.C. 2006.

Wanjing MA, Xiaoguang YANG and Meiping YUN. Development and Evaluation of a Coordinated and Conditional Bus Priority Approach. TRB, National Research Council, Washington, DC, 2007

IMPROVING CHINA TRAFFIC SIGNAL OPERATIONS THROUGH SIGNAL SYSTEM RETIMING

Edmond Chin-Ping Chang, Ph.D., P.E.[1]

[1]H President & CEO, EDCPC, Inc. 8720, Snowhill Ct., Potomac, MD 20854 USA
Tel: +1-301-983-4008 Fax: +1-301-983-4007 Cell: +1-240-643-8988 (USA)
e-Mail: ecc2005@yahoo.com or ecc2005@EDCPC.com

Abstract: Significant development has been made by many operating agencies throughout the China in the last decades, in deploying the urban Intelligent Transportation Systems **(ITS).** This paper describes efforts needed to (1) develop the practical implementation framework, (2) develop education program and training packages, (3) devise practical traffic signal operation evaluation mechanisms, (4) deploy the education and training workshop(s), and (5) develop web-based Community of Practice (COP) and peer-to-peer support network, in order to improve the skills for attacking the traffic signal timing, timing synchronization, and traffic congestion management problems currently exist throughout cities/counties in China.
Key words: *ITS, Traffic Management, Signal Timing, Traffic Models*

1 Introduction

Significant development has been made by many operating agencies throughout the China in the last decades, in deploying the urban Intelligent Transportation Systems **(ITS).** These include the development and deployment of various vehicular/pedestrian detector systems, standardized controllers, communication systems, information exchange networks, and field implementation of ITS platforms. There is an immediate need to elevate the traffic signal system control and operational practice into the next level by maximizing system/equipment/personnel already available. Since most cities/counties are now well equipped with new generation of control centers, traffic controllers, operational staffs, and domestic capabilities, the results can be very significant. As demonstrated successfully in USA, such as in the state of California during 1970s, in the state of Texas during 1980s, in the state of Florida during 1990s, significantly benefits can yield the benefits of 20% travel time saving, 15% stops saving, and 12-18% fuel saving. These tangible operational benefits can be readily obtained simply by retiming the existing traffic signal timing plans at the various local intersections, arterial streets, and downtown networks with updated traffic demands, traffic signal control strategies, and traffic signal timing parameters and updated plans to provide updated and smoother operations.

Various implementation guidelines and education workshops are available in the United States and Asia that can be used as the starting point for developing practical education/training packages, and implementation/evaluation programs in China.

2.ADVANCED SIGNAL SYSTEMS

Traffic signal system operations are typically modified through cycle lengths, signal phases, phase sequences, and offsets. Cycle length is total time required for a complete signal phase sequences and is typically between 60 to 120 seconds for a four-legged intersection. Signal phase defines how the signal operates. Phasing plans can be simple, two-phase plans (one phase per approach) or be tailored to allow protected/permitted movements and lead/lag phases. An intersection with heavy left turning traffic and heavy opposing through movements would probably include a protected left turn phase either before the opposing traffic is released (lead phase) or after the opposing traffic is stopped (lag phase). Offset between successive traffic signals is the time difference between the start of the green phase at an upstream intersection as related to the start of the green phase at an adjacent downstream intersection under isolated, arterial, and network control modes.

Signal Operation Modes

There are five (5) different signal controls, from individual intersection with pre-timed control to the network with adaptive control with various complexity.

Pre-timed Operations - Under pre-timed operation, the master controller or master intersection sets signal phases and the cycle length based on predetermined rates. These rates are determined from historical data. Pre-timed signal control is appropriate for areas where traffic demand is very predictable.

Progression Operations - A progression operation is a simple way of coordinating signals along an arterial, which is common in many urban areas. The signals can be set manually to run in a constant, synchronous manner. These include simultaneous progression, alternative progression, and flexible progression.

Actuated Operations - An actuated controller operates based on traffic demands as registered by the actuation of vehicle and/or pedestrian detectors. There are several types of actuated controllers, but their main feature is the ability to adjust the signal's pre-timed phase lengths in response to traffic flow. If there are no vehicles detected on an approach, the controller can skip that phase. The green time for each approach is a function of the traffic flow, and can be varied between minimum and maximum lengths depending on flows. Cycle lengths and phases are adjusted at intervals set by vehicle actuation of pavement loops. **Semi-Actuated Control -** A semi-actuated controller provides for traffic actuation of all phases except the main phase. A continuous green is maintained on the major street except when a demand is registered by the minor street detector. The right of way always returns to the major street when no vehicles are present on the minor street or a timing limit has been reached. Semi-actuated operation is best suited for locations with low volume minor street traffic. It may also be used to permit pedestrian crossings at mid street.

Full Actuated Control- Under full actuated control, the function of the controller is to measure traffic flow on all approaches to an intersection and make assignments of the

right of way in accordance with traffic demand. Full actuated control requires placement of detectors on all approaches to the intersection. The controller's ability to respond to traffic flow provides for maximum efficiency at individual locations. This type of control is appropriate for intersections where the demand proportions from each leg of the intersection are less predictable.

Traffic Responsive - In traffic responsive mode, signals receive inputs that reflect current traffic conditions, and use this data to choose an appropriate timing plan from a library of different plans. An individual signal or a network of several signals may be traffic responsive. Capabilities include: Vehicle Actuated - uses data from presence detectors and modifies the phase splits based on vehicle actuation and gaps. This procedure addresses current traffic and does not require traffic projections. Future traffic prediction - control system uses the volume data from system detectors and projects future conditions. Pattern Matching - the volume and occupancy data from system detectors are smoothed and weighted and compared with profiles in memory. This enables identification of stored profile most closely matching existing traffic. When a pattern is identified, appropriate parameters are needed.

Adaptive Control Strategies (ACS) – the ACS systems are currently the most advanced and complex control systems available. They are similar to traffic responsive signals in that they receive real-time data through detectors, but instead of matching current conditions to an existing timing plan, the system uses an online computer to create an optimal timing plan. No library of timing plans is needed, which works well for areas with high rates of growth, where libraries of timing plans would need to be updated frequently.

Offline Signal Timing Optimization Models

As computer technology progressed, computer models have replaced manual setting that optimize signal timing settings. These powerful off-line and on-line computer models use the historical data and simulation models to create optimal signal timings that either maximizes bandwidth or minimizes total delay. These models typically include (a) traffic flow model(s), and (b) algorithm(s) for optimizing specified operational performance, such as delays, stops, and/or queues. Several signal timing optimization programs are available through public domain, or from private companies (3, 4). These include UTCS, SOAP, TRANSYT, MAXBAND, PASSER II-80, PASSER III, SIGOP, MOTION, and various adaptive control strategies.

Adaptive Control Strategies

As opposed to the models outlined above, which use historical data to create one or more optimized timing plans, adaptive control strategies use real time data from detectors to perform constant optimizations on the signal timing plan for an arterial or a network with signals adapted non-recurring congestion, incidents, events.

UTCS Control Strategies – As mentioned above, the second and third generation control strategies developed by the FHWA are adaptive control strategies:

Table 1. Comparison of 1-GC, 2-GC, 3-GC UTCS Control Strategies (1).

表一 UTCS控制策略比較

特色	第一代控制(1-GC)	第二代控制(2-GC)	第三代控制(3-GC)
更新週期	15分鐘	5-10分鐘	3-5分鐘
控制計畫產生	離線最佳化 依每日時間、交通感應或手動模式自時制計畫資料庫選擇	線上最佳化	線上最佳化
交通預測	無	基於歷史資料	平滑目前資料
週期長度決定	區段內固定	變動路口群組內固定	於時間空間皆變動。於控制期間前決定

Second Generation Control (2-GC) - 2-GC control uses an online strategy that implements signal timing plans based on real time surveillance data and predicted values. The optimization process can be repeated every five minutes. However, to avoid transition disturbances, new timing plans cannot be implemented more than once every 10 minutes. The software also contains a traffic prediction model, CIC, and a transition model to minimize transition time between two plans.

Third Generation Control (3-GC) – Similar to 2-GC, 3-GC is a fully responsive, online traffic control system. Similar to 2-GC, the control plans to minimize a network wide objective using predicted traffic conditions. It differs from the 2-GC model in that the period after which timing plans are revised is shortened to 3 to 5 minutes, and the cycle lengths are allowed to vary during the control period.

Distributed Intelligence Traffic Control System (DITCS) - DITCS is a control system in which intersection controllers use timing plans but can dynamically adjust the splits to suit traffic conditions at the controller level. DITCS are closed loop systems providing real-time traffic adaptive control. The central system sends synchronization pulses, but most functions are performed at the intersection level maximizing the use of computing power. Some well known DITCS are Sydney Coordinated Traffic Adaptive System (SCATS) and TracoNet, described below:

SCATS - Developed by the New South Wales Department of Main Roads, SCATS is a dynamic control system with a decentralized architecture. SCATS updates intersection cycle length using the detectors at the stop line. SCATS allows for phase skipping. Offsets between adjacent intersections are predetermined and adjusted with the cycle time and progression speed factors.

TracoNet - TracoNet is a distributed intelligence closed loop network control system used for coordinating, controlling and facilitating the flow of vehicular traffic. It can operate in all control modes, including fully actuated. Traffic responsive algorithms based on pattern matching are also available.

SCOOT - Split Cycle and Offset Optimization Technique (SCOOT) - SCOOT is an off-the-shelf centralized computerized traffic control model developed at the Transportation Road Research Laboratory in the U.K. It is an enhancement over first generation UTCS systems and provides real-time adaptive control. SCOOT uses system detectors to measure traffic flow profiles in real time, and along with predetermined travel times and the degree of saturation (the ratio of flow-to-capacity), predicts queues at intersections. Cycle length, phase splits, and offset adjustments are made to operate at a preset degree of saturation (usually 90%). SCOOT is most effective when demand approaches, but is less than, capacity, where demand is unpredictable, and when distances between intersections are short.

RT-TRACS - Real-time Traffic Adaptive Signal Control System (RT-TRACS) - In 1991 the FHWA solicited proposals for the development of a real-time, traffic adaptive signal control system called RT-TRACS. Shortly thereafter, the FHWA contracted with PB Farradyne to develop and implement RT-TRACS. The RT-TRACS control logic assesses the current status of the network with forecasting capabilities, allowing proactive, not reactive, response. The most fundamental requirement of this system is to effectively manage and respond to rapid variations in traffic conditions. RT-TRACS consists of a number of real-time control prototypes that each function optimally under different traffic and geometric conditions. When conditions dictate, RT-TRACS can automatically switch to another strategy. The FHWA realizes that this control logic must be integrated with freeway performance data and provide network wide control. A thorough understanding of past experience with advanced traffic signal control strategies is critical to the development of effective RT-TRACS strategies for ITS. Features of the RT-TRACS design include: distributed & centralized traffic control; dynamic priority control on selected routes; interact with dynamic traffic assignment to implement proactive control; improved fallback capabilities in case of surveillance system failure; effective use of the accumulated experience with real-time control. Five (5) different prototypes real-time, traffic adaptive signal control strategies have been developed and evaluated for use in the RT-TRACS program. The RHODES evaluation indicated a reduction in delay, stops, and fuel consumption of 24 percent, 9 percent, and 6 percent, respectively, while maintaining the same throughput as baseline (vehicle actuated control) in Seattle, Washington, and Chicago, Illinois.

ATSAC (Automated Traffic Surveillance and Control) - The city of Los Angeles created ATSAC-based originally on UTCS is one of the earliest and most extensive advanced traffic management systems, including centralized, adaptive traffic signal

control. The system includes surveillance via loop detectors and closed circuit television, signal optimization software, and real-time remote control of signals.(14)

3.OPERATIONAL EVALUATION AND PERFORMANCE ASSESSMENT

Traffic signal control can provide significant benefits for traffic flow on a surface street network. However, it seems that the most advanced systems are not always the most effective. Careful attention must be paid to implement a traffic signal control system that is appropriate and cost-effective for the area. In addition, it is important to assess the current state of the existing traffic signal control system when projecting results of an improvement to the system. If a system is currently pretimed, and ACS is installed, there will probably be a significant improvement. However, if the current system is already fairly updated, the improvements will generally not be as great. After extensive field tests in the 1980s, which compared each generation of UTCS on an arterial and a grid network to a standard, pre-timed system, showed that the simpler methods performed better on average (See Table 3). 1-GC, in its various modes of operation, performed best overall, and demonstrated that it can provide measurable reductions in total travel time over that which could be attained with a well maintained fixed time system. In 2-GC and 3-GC, the effectiveness of the control system response depends entirely on the quality of the prediction model. The traffic responsive plan selection method was generally better than the time of day method. The results of the 2-GC method were mixed, but overall inferior to the 1-GC. The 3-GC strategy was unsuccessful in responding to traffic flows and degraded performance under almost all traffic conditions. Counter-intuitively, the more responsive strategies resulted in poorer performance than fixed cycle, non-responsive strategies. A close examination of the experiments reveals that expectations were not fulfilled because the models and procedures used in the UTCS study failed. **Table 3 "Comparison of Results of UTCS Strategies"** presents a summary of some extensive field tests conducted on UTCS control systems in the United States in the early 1980's, comparing each generation on an arterial and grid network. The UTCS strategies are compared to standard pre-timed traffic control.

Table 2: Comparison of Traffic Control Systems (2}.

表二交通控制系統比較

運作	UTCS	SCOOT	SCATS	TracoNet
控制	集中	集中	分散	分散
時差	預定	動態	預定	預定
交通感應	配對計畫 使用系統偵測器	交通預測演算法 使用系統偵測器	既有交通之時比調整 使用佔有率偵測器	型態配對 使用系統偵測器
軟體成本($1994)	$100,000	$350,000	$130,000	$8,000
聯絡	Honeywell, Sperry, 等.	TRRL, GE, Fortran系統	新南威爾斯道路部, Phillips, AWA	Traconex

Table 3. Comparison of Results of UTCS Strategies (1).

表五ACS組件估計成本

系統	中央硬體($)	中央軟體($)	現場控制器($)*	偵測器($)*
SCATS**	30,000	40,000-70,000	4,000-6,000	5,000-7,000
SCOOT	30,000	NA	NA	5,000-7,000
OPAC	20,000-50,000	100,000-200,000	4,000-6,000	NA
RHODES	50,000	500	NA	NA
ATSAC	40,000-50,000	1,000+版權	8,000-10,000	5,000-10,000

Operational Benefits

Traffic signal control improvements are very effective at reducing congestion. In fact, they generally provide the greatest payoff compared with any other method for reducing congestion on surface streets. signals do not need to become “state-of-the-art” to realize great improvements. one simple improvement, such as interconnecting signals that were previously operating independently, can produce significant results. Projects in the United States have found that (12):

Interconnecting previously uncoordinated signals or pretimed signals, and providing newly optimized timing plans and a central master control system can result in a travel time reduction of 10-20 percent. Installing advanced computer control has resulted in about a 20 percent travel time reduction when compared to interconnected pretimed signals using old timing plans. Installing advanced computer control has resulted in a 10-16 percent travel time reduction when compared to non-interconnected, traffic actuated controls. Installing advanced computer control, when compared to

interconnected pre-timed control with relatively active signal timing management, has resulted in an 8-10 percent travel time reduction.

Optimizing traffic signal timing plans, when compared to previously interconnected signals with various master control forms and varying previous signal timing qualities, has resulted in a 10-15 percent reduction in travel time. In addition to significantly reducing travel time, traffic signal control improvements also reduce stops, fuel consumption, and emissions. For example, the Texas Traffic Light Synchronization Grant Program II (TLS II) achieved reduced fuel consumption, delay and stops by 13.5 % (20.8 million gallons/year), 29.6% (22 million hours/year), and 11.5% (729 million stops/year), respectively. The total savings to the public in the form of reduced fuel, delay, and stops was approximately $252 million in the following year alone. More significantly, however, the study indicated that an average of 10 gallons of fuel was saved for every dollar that was spent on the retiming project (8). An aggressive signal retiming effort in California resulted in a benefit-cost ratio of 58 to 1. The program improved 3,172 signals across the state, resulting in a 15% reduction in delay, 16% reduction in stops, and a 7.2% reduction in travel time throughout the system. The reduced fuel consumption (8.6%) alone returned the total cost of the program 18 times over. (12)

Adaptive Control Strategies (ACS) have added benefits, such as increased safety. ACS can reduce stops through improved signal coordination, which in turn reduces the chance of rear-end collisions. Comparing to fully optimized fixed-time systems, SCATS has been shown to reduce stops by up to 40 percent (11). Since implementing SCATS, Broward County, Florida has seen a 28 percent decrease in stops, and Oakland County, Michigan showed a 33 percent reduction in stops. ATSAC in Los Angeles has reduced stops by 41 percent (11). ACS have the added advantage of being able to grow with a community. The ITS deployment tracking database shows that few areas re-time their signals each year. This is especially beneficial to high growth areas, where the best timing plans quickly become out-of-date.

4.BENEFITS & COST CONSIERATIONS

Traffic signal control improvements vary widely in cost, and can be quite expensive. Besides initial cost, signal operations & maintenance costs can be significant, and their implementation must be carefully considered. It is important to consider Preventative Maintenance and Design Maintenance (12). Preventative Maintenance, to be performed at regular intervals to avoid unnecessary problems. Response Maintenance, which includes quick response to emergency situations as well as trouble-shooting. Design Modification monitor new equipment and new signal locations in order to ensure safe and effective operation (11).

Los Angeles, CA Automated Traffic Surveillance and Control (ATSAC) -
With its large population and its auto-dependent urban form, Los Angeles experiences extremely heavy congestion on its arterials. The situation is exacerbated by activity centers, such as the coliseum and the airport, which create large and less predictable surges of traffic. In response to this problem, the city of Los Angeles created ATSAC, one of the earliest and most extensive advanced traffic management systems, including centralized, adaptive traffic signal control. The system, first utilized around the Coliseum for the 1984 Olympic Games, was initially based on the FHWA's UTCS signal control software. The system includes surveillance via loop detectors and closed circuit television, signal optimization software, and real-time remote control of signals.(14) ATSAC has had tremendous success in reducing system-wide congestion, as well as in clearing event traffic. Since the system was implemented, coliseum traffic clears within an hour after a big concert, compared with over two hours previously. (14). In addition, the system has been found to reduce stops by 35%, intersection delay by 20%, travel time by 13%, fuel consumption by 12.5%, and air emissions by 10%. The benefit/cost ratio was found to be 9.8:1, and the system paid for itself in less than one year. (15)

Oakland County, Michigan - Faster and Safer Travel through Traffic Routinand Advanced Controls (Fast-Trac) -
Oakland County looked for a more innovative approach. The result was the Faster and Safer Travel through Traffic Routing and Advanced Controls (Fast-Trac) system. Fast-Trac integrates advanced traffic management with advanced traveler information systems, with the SCATS adaptive control strategy at the core. With their SCATS system, Oakland County can claim many "firsts": the first adaptive traffic control system in the U.S., the first SCATS application in the western hemisphere, and the first to use video image processing with SCATS. Oakland County chose to use video surveillance instead of loop detectors with SCATS for several reasons. Video cameras are installed on any surface and in any weather conditions-- a very important advantage in the Michigan climate. Fast-Trac has been very successful on several fronts. There has been an 89-percent drop in the number of accidents at the most dangerous intersections, a 100 percent decrease in the number of serious injuries at those same intersections, and 40-plus hours a year trimmed from the average commute time (16). For example, at the World Cup soccer matches held in Detroit's Silverdome--and since then, at other major concerts and special events--tests showed that the traffic management system eased traffic flow and reduced the need for police to manually direct traffic. Overall, the program is responsible for a 19 percent increase in rush-hour travel speed and a significant decrease in accidents. Studies suggest that Fast-Trac could potentially reduce the average number of vehicle stops by one-third, decreasing the incidence of rear-end collisions and reducing carbon monoxide emissions by 12 percent (17),

Anaheim, California Field Operational Test with SCOOT -

Anaheim, best known for Disneyland, also houses many other large event centers, including a convention center; a professional baseball stadium, Anaheim Stadium; and a professional hockey ice rink, Arrowhead Pond. These event centers have a collective maximum attendance of 200,000 people; when combined with Anaheim's 300,000 residents, major traffic congestion results. With these unpredictable surges, it seemed that Anaheim was a perfect candidate for an ACS implementation. As part of the federally funded Anaheim Advanced Traffic Control System Field Operations Test (FOT), a version 3.1 SCOOT system was installed by Siemens for a portion of the City of Anaheim network near Arrowhead Pond and Anaheim Stadium. From fall 1994 to spring 1998, PATH researchers conducted a study comparing SCOOT to the previous UTCS system, which was already considered state-of-the-art. Contrary to expectations, SCOOT was not found to be an improvement over the UTCS system. The SCOOT system produced lower intersection delays in some cases, but more often it produced higher delays. In cases where there was improvement, the improvement was less than 5%, and in cases where conditions worsened, the increase in delay was less than 10%. There were several problems that led to SCOOT's less-than-ideal performance. Among others, SCOOT predicts traffic conditions using input from loop detectors located upstream of the intersection. The loop detectors in Anaheim were located closer than usual to the intersection, and therefore did not give SCOOT completely accurate traffic conditions.

5.CONCLUSIONS AND RECOMMENDATIONS

The most common challenge to implement traffic signal control improvements is the initial financial cost. As has been seen in California, Texas, Florida in USA, the operational benefits from a well-designed improvement program far outweigh the initial cost. It is crucial to use pilot studies and other evaluation techniques in selecting a system that will work well for a particular area. A limited SCOOT installation in Anaheim, California, produced little improvement, and even increased delay in some cases. According to a recently completed USDOT signal evaluation study, the detector placement may have been the cause of the sub-optimal performance of many signal system operations (11). Also, areas with fairly predictable traffic demand and low growth, a well-maintained fixed-time/time-of-day signal may perform just as well as the ACS system as being increasingly used in the United States. Staff training is normally required for implementing adaptive signal systems due to the increasing complexity to set up and fine-tune. Adaptive signal control operations is highly dependent on the communications and detectors used.

6.REFERENCES

(1982) "Computer Controlled Traffic Signal System." USDOT, Federal Highway Administration,1982.

(1994) “Transportation Infrastructure - Benefits of Traffic Control Signal Systems Are Not Being Fully Realized.” US General Accounting Office, 1994.

Busch, Fritz, “MOTION – a new approach to urban network control.” Traffic Technology International.

Dahlgren, Joy, et al., (1996) “Lessons from Case Studies of Advanced Transportation and Information Systems.” California PATH.

Fambro, Daniel, et. al., (1995) “Benefits of the Texas Traffic Light Synchronization (TLS) Grant Program II.” Texas Transportation Institute, 1995.

Gartner, Nathan H., Stamatindius, Chronis, and Tarnoff, Philip, J., (1995) “Development of Advanced Traffic Signal Control Strategies for Intelligent Transportation Systems: Multilevel Design.” Transportation Research Record 1494.

Genovese, Joseph, A., (1994) “SCOOT in the USA.” Institute of Transportation Engineers 1994 Compendium of Technical Papers.

Gordon, Robert, et al., (1996) “Traffic Control Systems Handbook.” US Department of Transportation, Federal Highway Administration.

Institution of Civil Engineers, (1988) “Electronic Traffic Control, How Does UK Compare?.”

Meyer, Michael, (1997) “A Toolbox for Alleviating Traffic Congestion and Enhancing Mobility.” Institute of Transportation Engineers.

Moore, Jayakrishnan, McNally, MacCarley (1999) "SCOOT Performance in Anaheim Advanced Traffic Control System", Intellimotion - Research Updates in Intelligent Transportation Systems, Vol. 8, No. 3.

Rowe, Edwin, (1998) “The Los Angeles Automated Traffic Surveillance and Control (ATSAC) System.” Los Angeles Department of Transportation.

ROUTE DIVERSION PROBABILITY MODEL BASED ON GUIDANCE UTILITY AND ITS APPLICATION

Meiping YUN and Shoupeng TANG

School of Traffic and Transportation Engineering, Department of Traffic Engineering, Tongji University, 200092, Shanghai, Siping Road No.1239, PH (8621) 6598- 5368; FAX (8621) 6598-5368; email: yunmp@mail.tongji.edu.cn

Abstract: Variable Message Signs (VMS), a collective route guidance information provision, are widely used in urban traffic in most metropolitan areas in China, which actually affect drivers' route choice behavior. In this paper, a disaggregate choice analysis method is used to describe changes in drivers' route choice behavior. Logit model, based on guidance utility, is applied to determine probability of route diversion, in which, not only traffic condition information characteristics are included, such as survey time, congestion level, but also drivers' characteristics are taken into account, such as gender, age, driving experience. A case study is presented by taking drivers' responsesto VMS on Shanghai out-ring road, where a lot of stated preference data and revealed preference data is surveyed. Parameters of the above logit model are determined using those SP and RP data. The result shows congestion level and age are important factors that affect route diversion probability.

Key words: *Guidance utility, route diversion probability, logit model, congestion level*

1 Introduction

Nowadays travel-related messages are widely used on urban roadways. These messages may be derived from static or dynamic information about the network. Static messages provide fixed information about the network and the destinations that it serves, and may be of use in tasks such as way-finding or preliminary trip planning; however, they do not recognize actual traffic conditions and so cannot respond to them. Dynamic messages reflect either prevailing or predicted conditions on the network, and require capabilities for collecting and possibly processing network data in real time. Such dynamic messages may describe the network conditions, or make recommendations based on the conditions, or both. A variety of presentation media (graphical, spoken or text) and levels of quantitative or qualitative detail in the messages are possible. Variable message signs (VMS) is one of the main information presentation displays of traffic condition information, which are used in urban expressways in Shanghai.

With these recent enhancements and application of traveler information systems how to understand travelers react to information provided in this way is really important for either evaluation or optimization of the system. For example, if VMS

display traffic congestion or traffic jam on a decision point in road networks , how drivers react to this information, either choosing a different route or keeping the original route. This paper focuses on this question and studies drivers' responses to guidance information considering both information content and different driver characteristics.

2 Literature Review

Many surveys and travel choice simulator studies have demonstrated the ability of ATIS to influence route choice. For example, Khattak, Yim (et al. 1999) presented survey results in which over 50% of respondents reported that they had made travel route changes in response to pre-trip information received by radio, television or telephone. Owens (1980) describes an early travel choice simulator study that demonstrated drivers' willingness to divert in response to highway advisory radio (HAR) messages about incidents. Some researchers have estimated sophisticated econometric models of route choice or route switching probabilities in response to ATIS, for example Uchida, Iida (et al. 1994) and Polydoropoulou, Ben-Akiva (1999).

There are many studies on the relationship between VMS messages and driver response. Peeta (et al. 2000) examined the effect of different message content on driver response under VMS. The issue was addressed through an on-site stated preference (SP) user survey. Logit models were developed for drivers' diversion decisions. The analysis suggested that content and level of detail of relevant information are factors that significantly affect drivers' willingness to divert. Other significant factors included socioeconomic characteristics, network spatial knowledge, and confidence in the displayed information. Results also indicated differences in the response attitudes of semi-trailer truck drivers compared to other travelers. These results provide substantive insights for the design and operation of VMS-based information systems.

A somewhat similar study was performed by Wardman(et al. 1997). Although drivers' response to VMS information will vary according to the availability of alternative routes and the extent to which the routes are close substitutes, the research findings showed that route choice can be strongly influenced by the provision of information about traffic conditions ahead. This has important implications for the use of VMS systems as part of comprehensive traffic management and control systems.

Most disaggregate-level studies of drivers' response to VMS use stated preference (SP) survey data rather than actual traffic data. It is generally not possible to infer from traffic measurements the effects of a VMS on individual driver behavior, since the drivers' intentions prior to receiving the messages are not usually known. One study encountered used aggregate traffic data, that is Yim and Ygnace (1996) who used loop detector data in Paris to investigate the effects of VMS on link flows. Time-series traffic data were analyzed to measure changes in mean flow rates at a selected link. It was found that VMS influences drivers to choose less congested

routes when the drivers are provided with real-time traffic information, and that a driver's decision to divert is closely associated with the information pertaining to the level of congestion. In the Paris region, drivers received prevailing queue length information from the VMS. According to the data analysis, a reported queue length of 3 km seems to be a threshold at which a significant number of drivers choose to divert to an alternative route.

Based on analysis of driver route choice responses to both VMS and radio information, Emmerink, Nijkamp et al. (1996) have suggested that some people have a natural propensity to use traffic information of any kind and from any source. Nonetheless, there is considerable evidence that the nature of the guidance information, and the conditions experienced prior to its dissemination, can strongly affect driver route choice response to it.

A number of generally idiosyncratic factors also condition a driver's route choice response to ATIS messages. A freeway bias has been observed in several studies (Hato, Taniguchi et al. 1995; Kitamura, Jovanis et al. 1999). Because of this bias, drivers receiving messages that suggest diverting from a non-freeway to a freeway facility are considerably more likely to switch than drivers who receive the opposite message, other things being equal. The influence of habit or inertia on route choice response has also been noted in a number of studies (Uchida, Iida et al. 1994; Srinivasan and Mahmassani 2000). Drivers tend over time to establish a preferred route for particular trips. Guidance messages that suggest switching from the preferred route to another are less likely to be accepted than messages that suggest the opposite.

However, as stated above, from these various surveys and modeling efforts, it is difficult to extract generally applicable quantitative conclusions regarding traveler responses to information. The state of knowledge does not yet allow the development of a general model capable of predicting that, on a given network, X% of drivers will divert to route Y if they receive message Z while driving. Unfortunately, sufficient experience with and data about these systems is still lacking. And the result of this research and survey can not directly guide the better way VMS works in Shanghai. This paper focuses on the relationship between route diversion behavior and VMS application in Shanghai out-ring urban expressway.

3 Model Formulation

3.1 Random utility theory

Disaggregate (i.e., individual-level) choice analysis methods based on random utility models have been widely applied to model drivers' decision making processes, which has a basic hypothesis that users make their decisions according to maximum utility. Here we propose guidance utility as travelers' decision rule. Suppose traveler n's choice set is A_n and the utility of choosing scheme j is U_{jn}. Only if the following inequation exists, traveler would choose scheme i:

$U_{in} > U_{jn}, i \neq j, i, j \in A_n$

Utility U is composed by deterministic part and random item. Then, $U_{in} = V_{in} + \varepsilon_{in}$.

The probability of traveler n choosing scheme i is P_{in},

$$p_{in} = \Pr ob\left(U_{in} > U_{jn}, i \neq j, i, j \in A_n\right) = \Pr ob\left(V_{in} + \varepsilon_{in} > V_{jn} + \varepsilon_{jn}\right)$$

$$0 \le P_{in} \le 1, \sum_{i \in A_{in}} P_{in} = 1$$

3.2 Logit models formulation

Suppose random item $\varepsilon_{jn}(j = 1,2,\mathrm{K}\ ,J_n)$ is an independent double exponential distribution. Taking the distribution parameter (η, ω) as (0, 1), then $U_{in} = V_{in} + \varepsilon_{in}$ is double exponential distribution with parameter is $(V_{in},1)$. Then the probability of choosing scheme i is

$$p_{in} = \Pr ob\left(U_{in} \ge U_{jn}, i \neq j, j \in J_n\right) = \Pr ob\left(V_{in} + \varepsilon_{in} \ge \max\left(V_{jn} + \varepsilon_{jn}, (j \in J_n)\right)\right)$$

Note $U_n^* = \max_{j \in J_n}(V_{jn} + \varepsilon_{jn})$. U_n^* is double exponential distribution with parameter $\left(\ln \sum_{j=2}^{J_n} e^{V_{jn}}, 1\right)$. If U_n^* is described as $U_n^* = V_n^* + \varepsilon_n^* = \ln \sum_{j \neq i}^{J_j} e^{V_{jn}}$, then ε_n^* is double exponential distribution with parameter (0, 1). So,

$$P_{in} = \Pr ob\left(V_{in} + \varepsilon_{in} \ge V_n^* + \varepsilon_n^*\right) = \Pr ob\left((V_n^* + \varepsilon_n^*) - \left(V_{in} + \varepsilon_{in}\right) \le 0\right)$$

So, the general multinomial logit model formulation is,

$$P_{in} = \frac{e^{V_{in}}}{\sum_{j=1}^{J_j} e^{V_{jn}}} = \frac{e^{V_{in}}}{\sum_{j \in A_n} e^{V_{jn}}}$$

3.3 Probability of route diversion

Drivers' route choice behavior under traffic condition information can be described as two kinds of situations, which are diverting to an alternative route or keeping the original one. Binary logit model is adopted here to predict probability of route diversion as follows.

$$P_{1n} = \frac{e^{V_{1n}}}{e^{V_{in}} + e^{V_{2n}}}$$
$$P_{2n} = 1 - P_{1n}$$

.

4 Case Study

Based on the above model formulation, we take VMS on Shanghai out-ring urban expressway as a survey object, where SP data and RP data are collected considering driver characteristics and information content. Driver characteristics include age, gender, and driving experience (how many years the driver has), and information content includes survey time, congestion level (here there are three levels, free, crowd, and jam).

4.1 Probability of route diversion formulation

Taking above general BL model formulation,

$$P_{1n} = \frac{e^{V_{1n}}}{e^{V_{in}} + e^{V_{2n}}}$$
$$P_{2n} = 1 - P_{1n} = \frac{e^{V_{2n}}}{e^{V_{1n}} + e^{V_{2n}}}$$

P_{1n}, probability of route diversion; P_{2n}, probability of keeping original route; V_{1n}, deterministic utility function of driver n from changing a alternative route; V_{2n}, deterministic utility function of driver n keeping the original route.

The utility function can be determined by many variables, in which characteristics of different choice and that of different driver are the most important and we call them characteristic variables.

Note that $X_{in} = [X_{in1}, \Lambda\ , X_{ink}]'$ is characteristic vector, K is the number of the variables. Let's set $\theta = [\theta_1, \Lambda\ , \theta_k]'$ is unknown parameter vector. The key problem is to estimate unknown vector θ.

V_{in} is function of θ and V_{in}, that is $V_{in} = f(\theta, X_{in})$. Suppose that V_{in} and V_{in} are liner function, $V_{in} = \theta' X_{in} = \sum_{k=1}^{K} \theta_k X_{ink}$

The formulation of probability of route diversion is as follow.

$$P_{1n} = \frac{1}{1+e^{-\theta'(X_{1n}-X_{2n})}}$$

$$P_{2n} = 1 - P_{1n} = \frac{e^{-\theta'(X_{1n}-X_{2n})}}{1+e^{-\theta'(X_{1n}-X_{2n})}}$$

4.2 Selection of characteristic variables

Based on description data analysis of the surveyed SP data and RP data on Shanghai out-ring urban expressway, we select gender, age, drive years and congestion level as characteristic variables as table 1 shows.

Table 1. Characteristic variables of BL model

options	Information characteristic			Driver characteristic		
	Dummy variable	Survey time	Information on congestion level	gender	age	Drive years
Variables	X_{in1}	X_{in2}	X_{in3}	X_{in4}	X_{in5}	X_{in6}
Route diversion V_{1n}	1	Rush hour：1 other：0	Jam, 2 Congestion, 1 Free, 0	Male,1 Female, 0	<30,1 31-50,2 >51,3	1year,1 2years,2 3-5years,3 6-10years,4 >10years,5
Keep original route V_{2n}	0	0	0	0	0	0
Unknown parameters	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6

So the utility difference is,

$$V_{1n} - V_{2n} = \theta_{1(1-0)} + \theta_{2(1/0-1)} + \theta_{3(2/1/0-0)} + \theta_{4(1/0-1)} + \theta_{5(age-0)} + \theta_{6(drivers-0)}$$

4.3 Data process

In order to estimate unknown parameter vector θ, descriptive data analysis is performed here for the collected data on Shanghai Out-ring urban expressway as following table.

Table 2. Data analysis for the unknown parameter estimation

Driver No.	Route diversion or not	Driver decision	Characteristics of VMS information			Characteristics of drivers		
			Dummy variable	Survey time	Congestion level	gender	age	Drive years
			X_{in1}	X_{in2}	X_{in3}	X_{in4}	X_{in5}	X_{in6}
1	I=1 I=2	0 1	1 0	1 0	1 0	0 1	3 0	5 0
2	i=2 i=2	1 0	1 0	1 0	1 0	1 0	2 0	4 0
3	I=1 I=2	0 1	1 0	1 0	1 0	0 0	1 0	1 0
…	…	…	…	…	…	…	…	…
n	I=1 I=2	δ_{1n} δ_{2n}						
…	…	…	…	…	…	…	…	…
N	I=1 I=2	δ_{1N} δ_{2N}						
Unknown parameter vector			θ_1	θ_2	θ_3	θ_4	θ_5	θ_6

In this paper we adopt the maximum likelihood function to estimate θ, where Newton-Raphson algorithm is applied. We select the sample size as 200 and use Matlab to solve algorithm programming. The estimated value of $\hat{\theta}$ is as follow.

$$\hat{\theta} = \begin{pmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \\ \theta_5 \\ \theta_6 \end{pmatrix} = \begin{pmatrix} -10.967 \\ -0.77653 \\ 1.3308 \\ 2.6175 \\ 3.8319 \\ 0.56279 \end{pmatrix}$$

The result of T test, standard deviation, Mcfadden coefficient is listed in the following table.

Table 3. Estimation result

Unknown θ vector	θ_1,……,θ_k,……,θ_K	
Estimated value	0.74362,-0.98162,2.4632,0.95339,-2.0116,0.20662	
t value	-1.8882,0.55521,2.1176,1.2887,1.8334,0.84231	
Standard deviation	5.80784,1.398606,0.628458,2.031157,2.090024,0.668147	
N=50 (sample size)		ρ^2 =0.5556 (Mcfadden coefficient)

5 Conclusions

The result in above table shows precision of the model is certain to be adopted here, as ρ^2 is 0.5556 more than the threshold value 0.2. Then we study the effect of information characteristics and driver characteristics respectively.

(1) effect of information characteristics

Dummy variable

Dummy variable is a virtual variable which may influence route choice probability but not considered respectively as characteristic variable in logit model. θ_1 meets the t test demand under significance level of 90%. That means some other factors do influence driver route choice probability, such as fuel consumption, drivers' knowledge of road network.

Survey time

Parameter θ_2 is a coefficient of survey time to route choice probability. T test result of θ_2 does not meet the demand under significance level of 90%, which shows that the survey time is rush hour or not does not influence route choice probability so much. From the estimated value of θ_2, probability of route diversion in rush hour is smaller than at other times. The possible reason is in rush hour, most drivers are on commuting trips and have relatively more strict time restriction compared to non-commuting trips who won't take a risk changing to a new route.

Congestion level

Parameter θ_3 is a coefficient of congestion level to route choice probability and meets the t test requirement under significance level of 90%. The result shows that congestion level remarkably influences drivers' route choice. When the information displayed on VMS is a traffic jam, route diversion probability is highest, and when VMS display traffic congestion, diversion probability is higher than that VMS display free.

(2) effect of information characteristics

Driver gender

Parameter θ_4 is a coefficient of driver gender to route choice probability. Based on the collected data, t test result does not meet the demand of significance level 90%. That revealed influence of driver gender to route choice probability is not so notable. From the actual value of θ_4 we can conclude that when VMS display traffic jam or congestion ahead, men are more likely to change to a new route than women.

Driver's age

We set four groups of surveyed drivers according to their ages. T test results shows that age does distinctly influence route choice probability. The estimated value of θ_5 is -2.0116 that reveals that the older the drive is, the less likelihood to change route.

Driver experience

According to the collected data of drive years, we named five groups on the sequence of length of drive years. The result of t test is less than 1.64 which does not meet the demand of significance level 90%. We can conclude from the estimated value of θ_6 that drivers who have driven for many years are more likely to change to a new route compared to those who have driven relatively fewer years. The possible reason is that drivers who have driven many years have more knowledge of road network, so there is little risk when they choose to divert to another route.

Reference

Emmerink, R., P. Nijkamp, P. Rietveld and J. N. Van Ommeren (1996). "Variable Message Signs and Radio Traffic Information: An Integrated Empirical Analysis of Drivers' Route Choice Behavior." Transportation Research A 30(2): 135-153.

Hato, E., M. Taniguchi and Y. Sugie (1995). Influence of Traffic Information on Drivers' Route Choice. Proceedings of the 7th World Conference on Transportation Research. Sidney, Australia. 1: 27-40.

Kitamura, R., P. P. Jovanis, M. Abdel-Aty, K. M. Vaughn and P. Reddy (1999). Impacts of Pretrip and En-route Information on Commuters' Travel Decisions: Summary of Laboratory and Survey-based Experiments from California. Behavioural and Network Impacts of Driver Information Systems. R. Emmerink and P. Nijkamp, Ashgate. Chapter 11: 241-267.Khattak, A. J., Y. Yim and L. Stalker (1999). "Does Travel Information Influence Commuter and Noncommuter Behavior? Results from the San Francisco Bay Area TravInfo Project." Transportation Research Record 1694: 48-58.

Owens, D. (1980). Traffic Information Broadcasting: Driver Reaction to Two Kinds of Traffic Message -- A Pilot Study, Transport and Road Research Laboratory.

Peeta, S., J. L. Ramos and R. Pasupathy (2000). Content of Variable Message Signs and On-line Driver Behavior. Transportation Research Board 79th Annual Meeting.

Polydoropoulou, A. and M. Ben-Akiva (1999). The Effect of Advanced Traveller Information Systems (ATIS) on Travellers' Behaviour. Behavioural and Network Impacts of Driver Information Systems. R. Emmerink and P. Nijkamp, Ashgate. Chapter 14: 317-352.

Srinivasan, K. and H. S. Mahmassani (2000). Modeling Inertia and Compliance Mechanisms in Route Choice Behavior under Real-Time Information. Transportation Research Board 79th Annual Meeting.

Uchida, T., Y. Iida and M. Nakahara (1994). Panel Survey on Drivers' Route Choice Behavior under Travel Time Information. Proceedings of the Conference on Vehicle Navigation and Information Systems (VNIS): 383-388.

Wardman, M., P. Bonsall and J. Shires (1997). "Driver Response to Variable Message Signs: A Stated Preference Investigation." Transportation Research C 5(6): 389-405.

Yim, Y. and J.-L. Ygnace (1996). "Link Flow Evaluation Using Loop Detector Data: Traveler Response to Variable Messages Signs." Transportation Research Record 1550: 58-64.

ITS Architecture for EXPO 2010

BAI, Yu[1] , XUE, Kun[2]

[1 2] School of Transportation Engineering, ITS Research Center, Tongji University, Shanghai. PH (021) 65985368; FAX (021)65985368; email: rabbitby@263.net

ABSTRACT: ITS technology applied in Expo is not only in favor of meeting the traffic demands by Expo participants and solving the traffic problems, but also an important approach in demonstrating the development and promoting advancement of ITS. Based on the ITS technology development and traffic characteristics analysis of Expo, this paper analyzed the different user groups' demands for ITS after defining the Expo users' requirements. Meanwhile, starting from the existing and potential possibilities of adopting info exchange technology, we are building the architecture model and information flow of ITS for World Expo 2010 Shanghai.
Key words: *ITS Architecture, EXPO, User demands, Information flow*

1 EXPO traffic review

The development of Expo traffic could be divided into following phases according to the history of Expo (Wu, 2006):

- **Formation period**（1851～beginning of 20th century）, the relatively small area featured in subject halls by single traffic means of on-foot.
- **Developing period**（beginning of 20th century ～before World War □）, exhibition areas expanded gradually and functional areas appeared, traffic facilities of mass transit were used inside the exhibition halls.
- **Perfecting period** (after World War II), exhibition function were clearly classified, areas enlarged and internal traffics diversified and networked. Special designs and plans were made in terms of transportation for Expo areas, with the precondition of taking consideration of the existing most advanced traffic modes and exhibiting delicately the future trip modes.

By the review on the development history of Expo, it is very obvious that Expo by itself is a product of history, which could be better and completely understood under certain social background and historical environment, while the transportation model of expo is also influenced by development environment of society. 2010 World Expo is merely an important stage to demonstrate the development of ITS technology.

Over the recent 10 years, ITS has been adopted in the special events of Expo, Olympic Games and World Cup, etc. which forms key means and safeguard measures in transportation organization and management. The Aichi Expo 2005 in Japan provided real-time traffic information service. (Hirata, 2004)

The typical function of ITS for Aichi Expo in Japan was to collect and manage traffic information and then use them in places which were thought most needed for

the moment, helping participants to reach their destinations safely and comfortably. The system structure was mainly composed of traffic information management center, navigation supporting system, parking guidance system, video monitoring system, bus location system, communication system and so on. Description of each system is as following:

- **Traffic information management center:** to collect and manage all kinds of traffic information and to supply them to the equipments of Internet and Plasma Display Panel (PDP), etc., by which the participants will choose their non-backup path. The system played a core part in ITS center of Aichi Expo, it gathered information from subsystems and supply them to the relevant transportation performers
- **Navigation supporting system:** to collect travel information and to supply participants the path to the exhibition site. The system offered service in the following two aspects: (a) trip arrangement: to reserve accommodations and to make trip arrangement (choice among different traffic tools) at the related service terminals. (b) mobile navigation: to supply guiding convenience to the in-coming and at out-going participants.
- **Parking guidance system:** to supply parking guidance by the guiding signs arranged around the outside spaces of the Expo areas. Meanwhile, the information could be transmitted to VICS and drivers would get the information as soon as possible via automotive navigation facilities.
- **Video monitoring system:** to collect and to distribute HDTV video image information via web camera head.
- **Bus location system:** to help operational management of bus lines by the intelligent transportation technology of public traffic priority, vehicle operation control and so on.
- **Communication system:** to supervise and to monitor network equipments as well as to supply image information to all the PDP inside the event place.

By the ITS appliance analyzing in Hanover and Aichi Expo, we can conclude that being an advancing front applied science which fuses all kinds of technical means of high-new science, ITS also plays role in the important event of Expo in the way of rationally organizing the traffic flow and balancing traffic demand as well. The scale of Shanghai Expo is far larger than that of the former two; the event location is on the central city edge of megalopolis crossing the two sides of the Huangpu River. It is rather unpredictable for the potential traffic flow, which will definitely bring about severe traffic problems. Further more, in respect of transportation information service system for this Expo, there are no mature experiences to be borrowed. There exists great necessity for Shanghai Expo to do an overall research on ITS from entire point of view.

2 Introductions to EXPO 2010

The World Expo 2010 lies in the central part of Shanghai, extending to the both sides of Huangpu River. The area is about 5.28 km2. During the Expo, the total participant volume will be 70 to 80 million person-times(pts), average daily volume will be 400,000 pts, peak day volume will be 600,000 pts and extreme peak day will reach 800,000 pts. According to the quantitative analysis for Expo transportation demand, the capacity target of Expo will meet the daily volume of 600,000 pts and will have a potential of ensuring 800,000 pts per day under strict traffic restrictions.

According to the demand, corresponding countermeasures will be taken separatedly in three different ways of arrival, inter and outer traffic of the Expo area. From the point of view of both travelers and administrators, the Traffic Plan for World Expo2010 Shanghai (The Bureau of Shanghai World Expo Coordination, 2006) mixes the various characteristics of time and space with the entire journey and it tends to build up corresponding countermeasure systems suitable for different transportation situations, realizing the intellectualization of transportation system and the systematization of transportation information service.

Table 1 show that Expo Traffic Plan defines the space relationship between urban districts and Expo area into four zones, i.e. green zone, buffer zone, comfort zone and home zone. Based on the different traffic areas, the traffic operation management makes analysis about the functional demand and determines the theme content.

Table 1. Content of traffic operation management for 2010 Expo

Area	Area coverage	Objective of operation management
Green zone	Main body of Expo site	Traffic operation management facing the Expo area
Buffer zone	Protective shell of Expo traffic	implementing in-out traffic control and linking management for purpose of reaching optimal state of Expo transportation
Comfort zone	Zone inside the outer ring roads	For operation management objective of best harmonizing transportation between Expo and urban areas
Home zone	Yangtze Delta with Shanghai as the head	Transportation experiencing zone serving Yangtze Delta area, China as well as the whole world for cooperative objective of multi-model transportation system

Base on Expo Traffic Plan, this paper studied on the ITS Architecture of World Expo 2010. The objective of this study is to establish a comprehensive Expo area related ITS system, making an adequate introduction to ITS structuring method, service target, system composition, logical framework as well as the information flow and so on.

3 ITS architecture for EXPO 2010

3.1 STRUCTURING METHOD OF ITS ARCHITECTUREFOR EXPO 2010

Similar to the structuring method of national ITS system architecture (Yang, 2001), we adopt the same Hatley-Pirbhai method (H-P method for short) for 2010 Expo. By defining the Expo users and demands analysis at first hand, we summarize demands for ITS from the different user groups, and then establish user requirement model. At the same time, based on the information exchange technology that exists for the time being or might arise in the future, we establish architecture model of Expo ITS system (master plan of architecture and information flow chart). Refer to Fig 1 for details.

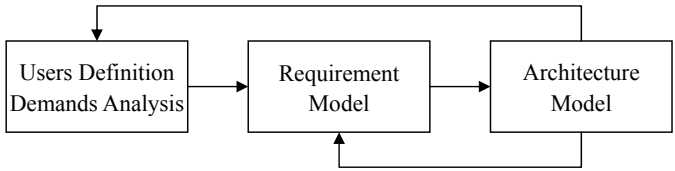


Figure 1. Structuring Method of ITS Architecture for Expo

3.2 USERS DEFINITION

The ultimate purpose of structuring ITS system is to serve all the personnel and institutions relevant to Expo. From this point, our initial task is to clearly define all the users to be served by ITS, and the following three aspects would be taken into consideration:

- **Utilizers:** users of ITS service (beneficiaries from the system output information)
- **Service provider:** provider of ITS service
- **Administrator:** to utilize ITS supplied by service provider for a better service to users

We conclude the following after analyzing users in the three respects. See Table 2 for details:

Table 2. ITS Users Group of Expo 2010

User Group	User
Utilizers	Bus drivers, passengers, private cars (native, non-native and foreign), bicycles, pedestrians (home and abroad), specially served people (troops make up of the old and weak), Expo area service staff and logistic vehicles
Service provider	Transport companies, bus companies, parking lot operators, Metro operation companies and water-way companies
Administrator	Expo manage department, commonality safety department, environmental protection institution, research institution, other government department

3.3 USER DEMANDS

User demands for ITS during Expo will vary respectively. The investigations we have done describe the different demands of various users and the result could be seen from Table 3.

Table 3. ITS Demands of Expo 2010 Users

Demand Type	Demand Content
Traffic safety info	Safety driving system of automotive vehicles, assistant signs of parking , safety information of pedestrians, etc
Exchange of static info	Bus route planning, info of stops and nearby location, fare collecting info of parking lot, gas station info, Expo related info, bicycle parking lot info, etc
Exchange of dynamic info	Real-time parking lots info, parking guidance info, real-time info of bicycle parking lots, roadway guidance info of Expo area, incident arrangement info, demand forecast of pedestrian flow, Electronic location, information releasing, etc
Public transport info	Position info for bus after N shift, arrival time prediction for bus after N shift, saturation info for bus after N shift, automatic bus-stop broadcasting, etc
Electronic toll collection (ETC)	ETC from bus passengers, ETC from parking, etc
Parking Management info	Parking guidance, parking lot reservation, parking lot info releasing, parking lot location, convenience of parking area identification and driving-out, real-time monitoring of parking lot, etc
Public safety management	Real-time monitoring of Expo area, early warning and handling of disasters, gathering and handling of emergencies, aid to unexpected events, etc
Intelligent scheduling	Intelligent scheduling for bus, intelligent scheduling for rail traffic, scheduling for logistics distribution, etc
Employee management	Management on vehicle drivers, management on bus company employees, management on parking lot employees, management on Expo area employees, etc
Database building and management	Dynamic information gathering, collection and analysis of Expo area operation data, database building, incident management data info, etc

3.4 REQUIREMENT MODEL

According to the analysis and summary to users' demand, we can draft out a general plan of demand. Entirely speaking, the general plan demonstrates all demands from users and defines the boundary range of the system, See table 4 for details

Table 4. Demand Analysis of ITS User Groups

Demand type / User group	Utilizer	Service provider	Administrator
Traffic safety info	●		●
Exchange of static info	●	●	●
Exchange of dynamic info	●	●	●
Public transport info	●	●	●
ETC	●	●	●
Parking Management info	●	●	●
Public safety management			●
Intelligent scheduling		●	●
Employee management		●	●
Database building and management			●

Note: ● *The user group have a requirement for this kind of service*

3.5 ARCHITECTURE MODEL AND INFORMATION FLOW

3.5.1 Architecture Model

This architecture model mainly describes the key sub-systems included in ITS, information handling process, info gathering, all kinds of info service systems as well as the related releasing methods. See Fig 2 for details. Descriptions to each model are as following:

Expo Information Exchange Center: Public data center of each business system in Expo area mainly supplies data service to all kinds of systems in the area. Meanwhile, being one of the nodes over Shanghai information network, the centers serve the information center of city level under the requirement of Expo information, providing effective data information to the area information center on actual time basis. All of these will supply possible reference in analyzing and policy making of the area transportation. Information Exchange Center is responsible for the collection, processing, storage, mining and maintenance, etc. of fundamental geography-spatial data, fundamental transportation data, urban and Expo area data as well as the dynamic transportation information data. The center also helps to offer fundamental and real-time data to the Expo Traffic Control Center as well as to offer unified data resources for the issuing of all kinds of information for ITS sub-systems, Shanghai Emergency Response Center, Shanghai Traffic Control Center and for the related government departments for result of actual and unified data management in keeping

the concordance and integrality of data as well as avoiding overlapping investment and construction. In another sense, the Expo Information Exchange Center provides a unified ITS integrated information platform to all the GIS systems inside the Expo area, which enables each business department to carry out re-development on it, and to achieve resource sharing to the largest extent.

Expo Traffic Control Center: This is a general management control center established inside the 2010 Shanghai Expo area, which carries through unified management in the respect of transportation, environment, safety, emergency state, facilities and integrated information, etc. The center plays the role of monitoring all the assembly halls and important facilities, analyzing and handling route crowd, aiding pedestrians, dealing with emergency, analyzing and controlling environment as well as distributing and managing public information, etc. Furthermore, it is responsible for the coordinating, decision-making and info controlling related with ITS sub-systems inside the Expo area. Besides each of sub-systems, its service object includes government competent authorities, departments at all levels and emergency response centers, offering them the latest data and report in terms of transportation information and decision.

Traffic Management System： By applying computer and communication technology, traffic management systems and vehicles would be treated as a single system instead of two separate ones. Advanced traffic management would be implemented in Expo areas by use of real-time information such as travel time and traffic incidents.

Public Transportation Management System: It is mainly responsible for the integrated coordination and management of bus scheduling and operating in Expo areas, info and designing of public transportation, public transport demand and safe operation of public transportation, etc.

Parking Management System: It is mainly responsible for the operation and management of parking lots in Expo areas, including parking space reservation, info releasing of vacant parking space, fee collection, travel guidance and safety of parking lots, etc.

Emergency management system: This is responsible for monitoring and managing emergency incidents occur inside the Expo area, including acquirement of emergency incidents, level evaluation, making of emergency scheme as well as optimization and decision-making of measures, etc. The model could function together with Shanghai Emergency Response Center.

Logistic Management System: The duty of the system is the united arrangement and management of logistics inside the area, including facilities in each hall, layout of large-scale exhibitions, supplement of routine consumable goods, waste recycle and transport, etc.

Traffic Information ***dissemination System:*** This is responsible for info distribution and consultation related with Expo transportation, which mainly includes:

- Expo Information Service Website: Topographic map of Shanghai, parking lot info and reservation function, recommended traffic lines to the Expo area, consultation from other service places in Shanghai, recommendation of traffic lines inside the area, info consultation of event places or halls inside the area, appointment of participants, ticketing and reservation system and all kinds of historical information, etc. In addition to the above mentioned items, the website will also consider issues of service versions of multi-languages as well as relevant services to overseas participants.
- Screen-touching Information Inquiry and Service System: Functions of information inquiry related with Expo event place, ticket reservation, transfer information, recommendation of lines, weather forecasting, air quality inquiry, real-time saturation inside the event place, recharge and inquiry of all - in-one card and so on.
- PDP: Display of Expo area map, distribution of event places or halls, real-time saturation of participating groups inside the event place, emergency announcement, real-time info of road situations and so on.
- Expo FM Service: Introduction to comprehensive info of Expo, real-time saturation of participating groups inside the event place, weather forecasting, real-time info of road situations, emergency announcement and so on.
- Mobile Phone Service: Character info service, image info service of location and POI, various real-time info inquiries and function of mobile payment, etc.
- Phone-inquire Service: info service over telephone
- Vehicular Facilities: GPS monitor, Real-time route guidance and collection of real-time traffic info
- Display Wall (TOC) \ Smart Board：This is mainly used in Expo Traffic Control Center, reflecting the real-time transportation situation inside the Expo area, e.g. the area digital map, event place info, traffic flow information, traveler flow information, emergency incident info and so on. By this direct reflection, rapid and proper emergency measures will be taken respectively.

3.5.2 Top Layer Information Flow

ITS of Expo is composed of participants (traveler), vehicles, Sub- system of the Center and transportation infrastructure in Expo areas. At the top layer of the model, the chart describes the info flow relationship between these four components and physical modules as well as the info flow relationship between the components and each of the external terminals.

The information channel in the physical structure determines the way of communication and contacting in which the info flow will be carried out as possible, and by defining the information channel chart and information channel, it introduces method of communications between each of ITS part. The communication in ITS could be borrowed from the four communication media defined by UNIA (FHWA, 2003), such as Wire Communication (fixed- fixed), Wireless WAN (fixed- mobile), Dedicated Short Range Communication (fixed- mobile) and Vehicular Communication (mobile - mobile), etc. See details in Fig 3.

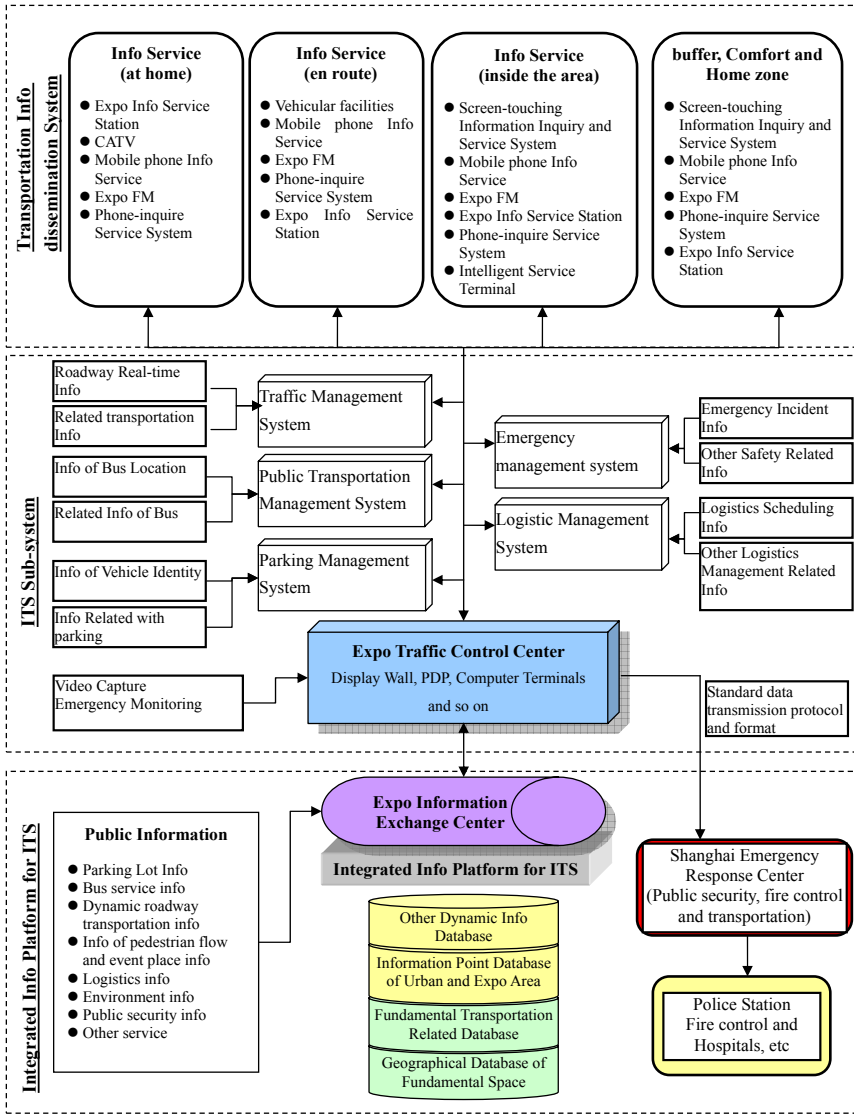


Figure 2. ITS Structure Module of World Expo 2010

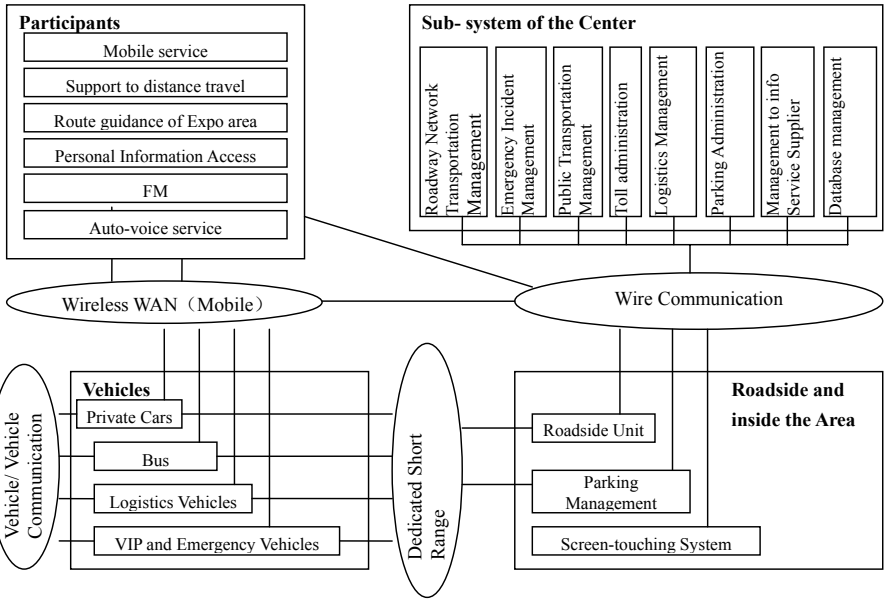


Figure 3. Top Layer Information Flow Chart

4 Applications

The further development of this paper could be done in respect of both theoretical research and practical application. In theoretical aspect, a deeper research is suggested to do in the field of sub-system, such as public transportation management system, roadway management system, parking management system, info distribution system, logistic management system, disaster prevention, terrorist prevention and emergency rescuing system, etc. as well as the data transmission standard and method. Furthermore, the issue of utilization of ITS system after the Expo 2010 should be researched. In practical application aspect, we suggest that on the basis of this research the construction of Expo Information Exchange Center and Expo Traffic Control Center be started and that the fundamental spatial geographical database, fundamental transportation database, POI of urban and Expo area as well as the dynamic transportation info database be constructed at the most priority position, reaching the further combination with Shanghai ITS.

5 Conclusions

The application of ITS technology at World Expo will not only be in favor of meeting the traffic requirement of Expo users, relieving traffic problems exist at Expo

but also be a demonstration of ITS technology development and be the important means of boosting ITS technical progress.

On the basis of analyzing ITS technology development and Expo transportation characteristics, starting from defining Expo users, this paper makes conclusions about the ITS demand of different user groups and establishes user demand module after analyzing user demand. At the same time, on the basis of the existing and future info exchange technology that might occur, this paper establishes the World Expo 2010 ITS System structure module.

We hope that on the basis of this research, further research and exploration could be done in the study and appliance of ITS, which will enable us to do adequate and perfect preparations for Expo 2010, reaching our target of hosting the most and unforgettable Expo in Shanghai.

6 Acknowledgments

This research is a component part of Research on Intelligent Transportation Management System for 2010 Expo under the project of Science and Technology Commission of Shanghai Municipality. Our utmost respect and sincere thanks go to Science and Technology Commission of Shanghai Municipality and all the project members for their support to this research. Besides, this work was supported by the Program for Young Excellent Talents in Tongji University （2006KJ027）and National Natural Science Foundation of China（50578123）.

References

FHWA, U.S. DOT, *National ITS Architecture*, Washington D.C. 2003

Hirata Tetsuya, etc., *ITS for Expo 2005 Aichi Japan*, 11th World Congress on ITS (CD-ROM), Aichi, Japan, 2004

The Bureau of Shanghai World Expo Coordination, *Traffic Plan for World Expo2010 Shanghai*, 2006

Wu Zhiqiang and Li Xin, *An Analysis about the Evolution and Features of Expos' Internal Traffic*, Urban Planning Forum, Tongji University Publishing Company, 2006(5), pp. 46-52

Yang Peikun, System *Architecture of Intelligent Transportation System*, Tongji University Publishing Company, 2001

CAN WE PREDICT ROAD SAFETY BY SIMULATION?

Henk J. van Zuylen

Delft University of Technology, P.O.Box 5048, 2600 GA Delft, The Netherlands,
TRAIL Research School, email: h.j.vanzuylen@tudelft.nl

1. Introduction

The conditions of road traffic have changed considerably in the past century. Dust roads were replaced by paved roads, pavements were improved with respect to maintainability and driver comfort, standardization of road construction has been introduced, improvements of the cars like suspension, tires, lighting, cage construction, air bags etc. has changed the characteristics of cars, regulations for vehicle maintenance, visibility of vehicles etc. and the introduction of improvements of cars like cruise control, and ABS breaking systems have an important influence on the modern traffic. Traffic safety has increased in a relative way: the number of accidents per car or per vehicle kilometer has decreased over the years, not in the last thanks to these technological improvements and to the education of drivers.
However, the safer the traffic conditions are, the more risk drivers seem to take. The introduction of ABS made driving at high speeds and slippery pavements relatively safe. The consequence has been that drivers with ABS drive faster under slippery road conditions, which partly annihilates the impact of the technological improvement (Kahahne 1994, Farmer 2001). This phenomenon of *rebound* is well known, but foreseeing this is rather difficult. The initial effect is technological, the reaction is typically behavioral. Predicting human behavior is even more difficult than inventing effective means to improve traffic safety.

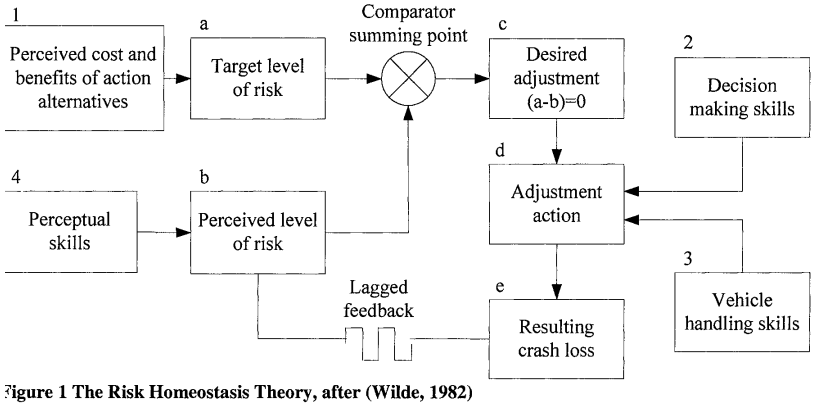


Figure 1 The Risk Homeostasis Theory, after (Wilde, 1982)

Wilde (1982) introduced the concept of *risk homeostasis* which means that if driving conditions become safer, drivers tend to take more risks, such that the risk level remains the same. There is much discussion about the validity of this proposition, but

there is no doubt that the introduction of new devices and facilities to improve traffic safety may have the consequence that drivers are intended to change their behavior in such a way that the risk level remains the same.
This conclusion makes the answer to the question of this paper rather trivial: of course it is impossible to find the impact of technological innovations or new regulations on traffic safety by using simulation alone. Still it is good to give some nuances to this statement and maybe the situation is much more positive than what we would think from a first analysis.
In this paper we will further analyze the question: "can we predict the impact of a technological innovation on traffic safety?" In chapter 3 we will analyze different kinds of simulations. Chapter 4 discusses the use of simulation programs to analyze the impact of new driver support systems on traffic safety. Chapter 5 gives some conclusions and ideas for further research.

2. Analyzing the average impact of technological innovations on traffic safety
In the 60s of the past century Smeed developed an analytical model for the number of traffic accidents in a country as a function of parameters like the total distance traveled in a country. This relationship is very abstract and is blind for effects of specific measures to improve traffic safety. Of course, the relationship contains the effect that a country takes measures to improve traffic safety when the amount of traffic is growing and apparently there are not so many differences between different countries.
In 1976 the author analyzed in an unpublished report the number of pedestrian accidents in different countries and he found that when the motorized traffic is increasing, the relative safety for pedestrians is also increasing. This is not simply due to the fact that fewer pedestrians remain when motorized traffic is growing, but by the fact that the roads become safer and the conflicts between high speeds motorized traffic and slow traffic become more seldom and better regulated and not at least that drivers as well as pedestrians get adapted to the dangerous situations. Countries with more roads per vehicle have relatively more pedestrian accidents than countries with fewer roads per vehicle. This can be explained by the fact that more developed countries have roads that are suited for more traffic and therefore also have better safety provisions.
Such macroscopic relationships give us the possibility to express the traffic safety in macroscopic quantities, which makes it possible to predict the safety improvement after certain measures, e.g. the reduction of the maximum speed.
These models are based on a statistical analysis of traffic safety under different situations. Apart from the formula of Smeed (1949), that is valid for a whole country or region, there are formulas that describe the number of accidents on a road, an intersection, e.g.

$$AA = 0.0015 \cdot ADT^{0.9711} \cdot 0.8897^{W} \cdot 0.9403^{PA} \cdot 0.9602^{UP} \cdot 1.2^{H}$$

With
AA = Total Accidents [mile/year]
ADT = average daily traffic
W = carriageway width [feet]
PA = paved shoulder width [feet]

UP = unpaved shoulder width [feet]
H = roadside hazard rating (Zeeger et al. 1987)

or

$$AA = 0.0019 \cdot ADT^{0.8824} \cdot 0.8786^{W} \cdot 0.9192^{PA} \cdot 0.9316^{UP} \cdot 1.2365^{H} \cdot 882^{T1} \cdot 1.3221^{T2}$$

AA = Total Accidents [mile/year]
ADT = average daily traffic
W = carriageway width [feet]
PA = paved shoulder width [feet]
UP = unpaved shoulder width [feet]
H = roadside hazard rating
T1 = 1 if flat terrain; 0 other
T2 = 1 if mountainous terrain; 0 other (Zeeger et al. 1987)

These formulas show the influence of shoulders on traffic risk but not the effect of other means to increase traffic safety. Similar empirical analyses have been done for other traffic conditions e.g. the curves of a road and speed differences.

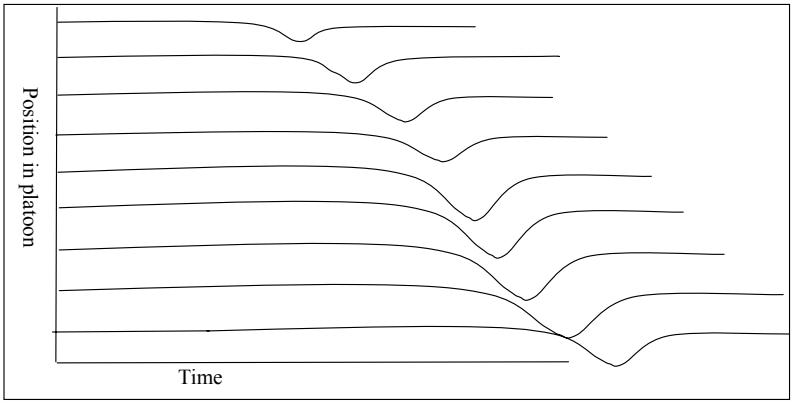


Figure 2 Instability in car following behavior. The breaking of the first car in the platoon is amplified when propagating back due to the reaction time and the over-reacting of the followers who follow on a too close distance.

The simulation of traffic states in order to assess the traffic safety has been attempted by several researchers. E.g. figure 2 shows that a simple car following model can lead to instabilities in the traffic flow by which a sudden speed reduction of a vehicle in a platoon is amplified by a delayed speed reduction of its follower, which leads to a collision of two vehicles far behind the leader. This example shows the importance of the reaction time and headway, but it does not have any predictive value, even though the situation occurs in reality that a maneuver of a vehicle leads to a collision far back in the row of following vehicles.

For intersections and roundabouts empirical data exist that shows the accident probability for different kinds of control and geometry. These figures can be used as a guideline (Polidori and Gatti 2007)

Table 1 Accident figures for different intersection types (adapted from Polidori and Gatti 2007)

	Crossroads	Crossroads (signalized, without left-turn filter and switched off at night)	Crossroads (signalized and left turn filter and 24 hours operation)	Partial grade separated	Staggered junction
Mean accident rate [acc/10^6veh]	0.93	1.31	0.86	0.94	0.84
Accident cost rate [DM/10^3veh]	80.40		36.80	38.60	37.20

Of course, these figures have to be considered with some caution, because they are the result of a limited number of intersections in one country only (Germany) and in a specific time (before 1990).
The examples show that the impact of certain (technological) changes can be estimated in advance by using certain relationships that have been derived from statistics.
The next question is whether it is feasible and reliable to use simulation for the prediction of the impact of measures on traffic safety is not easy to answer. The answer has to do with the question of validity and generality of the simulation and the safety measures that can be obtained from the simulation.

2. Traffic simulation for risk assessment

Simulations mimic reality but no simulation can deal with the whole reality. It is assumed that the relevant aspects of reality are represented in the simulation and that the properties of the simulation represent the properties of reality sufficiently faithful under different relevant conditions.
As an example we can look at a driving simulator. Of course, such a simulator should not necessarily kill the driver if he makes a fatal error and collides with high - simulated - speed with another vehicle or an object. It is even not necessary that it makes the noise of an accident in that case. Some discussion exists on the question whether the driver in a simulator should feel the acceleration and deceleration forces during his performance. But the driving simulator should give the test person a realistic cognitive task if one wants to know what the effect of distraction is on the driving performance. In such a way Abdel-Aty (2003) investigated the influence of the use of a mobile telephone on the driving task under the following assumptions:

- the driving task was well represented in the driving simulator and the task of the driver was sufficiently similar to the task during a real driving task
- the number of errors made by the driver in the simulator was a valid measure of the safety risk.

The first issue is clear, the simulator should be realistic with respect to relevant features of the system to be simulated. The second is equally relevant: the measures – the quantitative or qualitative properties of the performance of the simulated situation – should be valid predictors of the (un)safety.

The process to identify the safety impact of certain measures or technologies on traffic safety using simulation has three important steps (figure 3):

- mapping a selected part of the reality on a valid simulation model,
- simulating the effect of some measure, including the effect of human behavior,
- evaluating the effect of the measure with respect to some quality or quantity that represents traffic risk or unsafety.

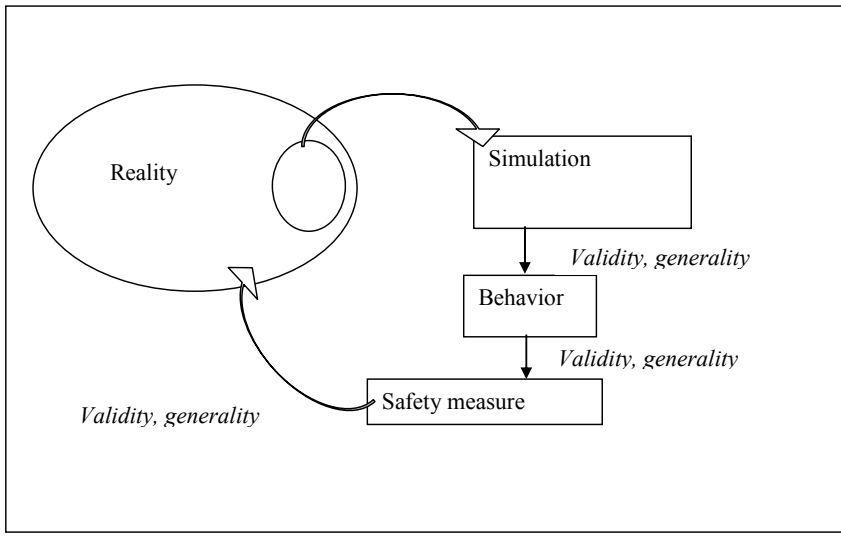


Figure 3 Mapping of the reality to a simulation

Only a part of reality is simulated and often this part is not the (un)safety itself. The outcome of the safety simulation should be a statement about the safety consequence. The naive example of figure 2 simulated directly the headway between cars in a platoon and it is evident that a distance less than zero is a collision. There is no doubt about the validity of the measure of safety, but apparently this simulation model is not valid. Of course the car following model used in figure 2 was calibrated and validated, but the generalization of that model to collision conditions was never made.

The validity of a simulation model can be seen at different levels:

- Content validity (does the model have parameters that are logically related to the characteristics of the system)
- Construct validity (are the parameters values simulated mutually consistent and consistent with observations in reality)
- Predictive validity (are the predictions made with the model consistent with the observed evolution of the system)

It is obvious that the simple car following model has content validity, but no construct validity and certainly no predictive validity.

2.1 Different kinds of simulation

The simulation of parts of the traffic systems is done in several ways. Without trying to give a complete overview of all kinds of simulations, from crash simulations to models of car dynamics, this paper will concentrate on the simulation of *traffic flows*. The question to be answered is whether it is possible to use traffic flow models for the ex ante assessment of the safety implications of traffic measures and new technologies like cruise control, narrowing of lanes or speed limitations.

2.2 Validity of the traffic simulation

The importance of a valid simulation is illustrated by the following example. In The Netherlands measures have been taken to reduce the emission on motorways close to cities. Since emissions are minimal at a speed of about 80 km/h it was decided to impose a speed limit of 80km/h on several parts of the motorways. A first trial showed that the air quality indeed improved and that the traffic performance did not deteriorate. For the new road stretches simulations were done to assess the impact of the speed reduction. The results of the assessment were positive. However, when the speed limits were indeed imposed on these road stretches, it appeared that on several places the congestion increased considerably and that the impact on the air quality was not always good (Harms 2006).

A closer analysis of the situation showed the following:

- the behavior of the drivers on the 80 km/h roads was different from what the simulation model assumed. The drivers clustered on the right lane which made merging at on-ramps difficult
- the maximum speed of 80 km/h reduced the capacity of bottlenecks. In the first trial no capacity bottleneck was present on the stretch with the maximum speed, but on the later stretches there were several bottlenecks which deteriorated the performance of the roads.

Most important is that the simulation did not show these effects, because the model did not represent the driver behavior and the capacity in the model was not affected by the maximum speed. The model that was used had been calibrated and validated, but it was not validated for the conditions for which it was used.

This is a general problem: simulation programs are validated for certain conditions and they predict certain aspects quite well, but if we want to use them for applications for which they had not been validated for, the models may not perform well. Microscopic simulation programs have in general been validated to predict travel time, capacity of road sections and intersections, and the emergence of congestion. If we want to use these models for other purposes, for instance the disappearance of

congestion, they don't automatically perform well. Also, if we want to use the vehicle trajectories to assess the emission of the vehicle during a trip, it appears that although the travel time is well represented, the trajectories are not sufficiently accurate to calculate the emissions (figure 4).

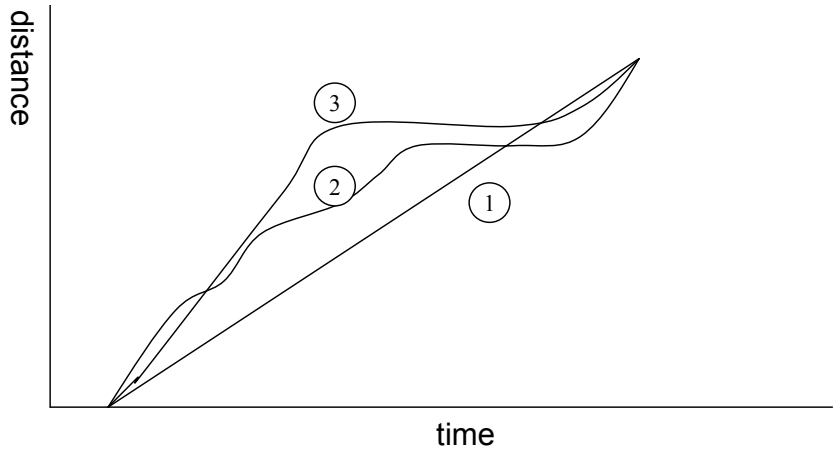


Figure 4 Three different trajectories giving the same travel time but different driving patterns: trajectory 1 with constant speed, trajectory 2 and 3 with higher maximum speed and one or more stops/decelerations.

For traffic safety assessment we have to take into account that the flow models, even though they are sufficiently valid for many evaluations of the network performance, they are not suited to assess all aspects of the traffic that determine the traffic safety. Models are always simplifications of reality. The simplification that is satisfactory for one purpose is often insufficient to make the model suited for other applications. For instance, the use of car following models in simulation programs is often based on measurements done many years ago, while in the meantime drivers, cars and roads have changed a lot. Recent research has shown that in reality different drivers follow different car following behavior with different parameters: some drivers have a longer reaction time than others and some show different following behavior (Ossen and Hoogendoorn 2007).

2.3 Measures for safety assessment

The other problem is that, even if we know the behavior of the traffic sufficiently accurately, it remains difficult to derive from that a measure for the traffic safety. Often the time-to-collision is used as a measure of traffic safety. This is the time during a maneuver that a collision would occur if the drivers would not change their speed and / or lane. There is not much evidence that this measure is a valid predictor of safety. Other measures are the macroscopic relationships between traffic parameters

and accidents like the ones presented in section 1. These measures are at least valid for situations in the past, but also here the question remains whether they are also valid predictors of unsafety. Minderhoud (1999) introduced two other measures for safety in simulation programs: the TET, the Time Exposed Time to collision, the total time that drivers are in a state where the time to collision is below a threshold, and the TIT, the Time Integrated Time-to-collision, the integration of the difference between the time-to-collision and the threshold value.
Another measure for traffic safety is the number of accelerations and deceleration per time. Saka and Glassco (2001) developed a simulation program to investigate the impact of several ITS applications on the safety of truck traffic. They showed that the number of sudden accelerations and decelerations decreased by the measures. The validity of this measure, which could be considered as the acceleration/deceleration noise has not been proved as far as the author knows, but the likeliness of the validity is quite high, since it has been shown that in queues with stop and go traffic the accident frequency as well as the acceleration/deceleration noise are quite high and correlated (van Zuylen et al. 2007).
Although there is some logic in these unsafety measures, the values that are produced by a simulation program for these measures depends strongly on the way the behavior of cars has been modeled. The simulation instrument is a main determinant for the values of the unsafety measures.
Concluding: most existing simulation models have not been intended to be used for traffic safety assessment. The validity of the parameters calculated by these models for the representation of driving behavior that affects traffic safety has not been proven. The heterogeneity of the drivers is not well represented in these models while it is likely that heterogeneity is an important factor in the emergence of conflicts in traffic. Furthermore, the measure that is believed to represent the traffic safety like time-to-collision has not been validated yet. For simulation programs that calculate traffic flows and average speeds the measures for unsafety can be derived from statistical analysis of accident data. Also here there is some doubt whether such relationships have also a predictive validity.

3. Probabilistic models
Models and reality nearly always differ. Most model developers and users name the difference between model and reality ‘noise’. It is not uncommon that this ‘noise’ is deafening. Often it is very difficult to see the supposed regular behavior of a traffic system.

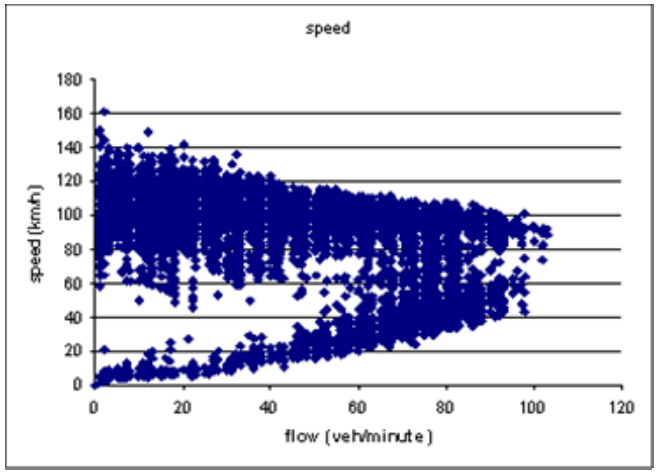


Figure 5 Observations of speeds and flows over a year. The so-called fundamental diagram is hardly distinguishable.

This 'noise' often is evidence of a flaw of the model: it may represent the average behavior of the traffic system, but the deviations from the average behavior may indicate that the model is incomplete, that some mechanisms are present in reality that are not included in the model. For instance, the model may incorrectly assume that all drivers behave the same, apart from some variations in some parameters. As Ossen and Hoogendoorn found, in reality driver behavior is not homogeneous. By the introduction of different user classes or other parameters into the model a part of the noise may disappear. However, further fitting of the model to reduce the remaining error between the model and the calibration data does not always give a better model. Over fitting may occur, where the model fits quite well the calibration data but the performance on a different dataset or the predictive value is still bad.

In traffic processes a certain level of stochasticity remains. As van Zuylen and Hoogendoorn (2007), van Zuylen and Viti (2007) and Viti and van Zuylen (2006) show, some traffic processes should be described in probabilistic terms and not as deterministic. Especially for the prediction of traffic safety aspects a good approach would be to describe the traffic process as evolving probabilities where the impact of measures can be assessed as changes in the probability distribution. So instead of assessing the time that accidents could occur, like this is done with the time-to-collision, one should assess the probability of dangerous situations.

Finally, traffic accidents are not a priori deterministic in nearly all cases. In reality accidents are due to a coincidence of different conditions. Small changes in these conditions have big consequences for traffic safety. The process before and in an accident has a chaotic character, where small different in conditions have a big impact on the outcomes. Not much work has been done until now on chaotic models in traffic.

4. Conclusions

The use of simulation programs to assess the impact of measures on traffic safety is possible, but needs a double validation: the simulation should be valid for the properties that influence traffic safety and there should be a criterion and observable measure that correlates well with traffic safety. Both requirements are difficult to satisfy for most simulation programs. Accidents happen too frequent, but still the causes of accidents are not always easy to identify and the accident process is often very sensitive to small influence factors and cooperative affects.

Therefore, the use of simulation for the determination of effects of measures and technologies on traffic safety is often based on believes and expert knowledge: if in a driving simulator the number of errors made by a driver is counted and when this is considered as a measure of the unsafety, the simulation has at least *face validity* a logical consistency with the expected relationship with traffic unsafety.

In macroscopic simulations, like e.g. traffic assignment programs and traffic flow models, the empirical relationship between macroscopic characteristics of the traffic situation and safety can be used to predict changes in accident after changes in the traffic conditions. Still this is at best an estimate of the expectation value and not accurate prediction of the change in number of accidents.

For microscopic simulations no validated measure for safety exists, as far as the author knows. The time-to-collision is often used, sometimes modified to the TIT or TET, but the validity as safety predictors has still not been proven.

A most difficult to assess is the rebound effect, the effect that drivers change their behavior after new technologies become available or certain measures have been taken. This adaptation leads to more risky behavior which may annihilate the effect of the improvement.

5. Literature

Abdel-Aty, M. (2003) Investigating the relationship between cellular phone use and traffic safety ITE Journal (Institute of Transportation Engineers), 73 (10), pp. 38-42

Chen, T., Wei, L. (2007) Driver-vehicle-road virtual simulation with interaction mode Chang'an Daxue Xuebao (Ziran Kexue Ban)/Journal of Chang'an University (Natural Science Edition), 27 (1), pp. 67-71.

Farmer C.M. (2001) New evidence concerning fatal crashes of passenger vehicles before and after antilock braking systems. Accident analysis and prevention, Vol. 33, pp. 361-369

Gettman, D. and L. Head (2003), Surrogate Safety Measures From Traffic Simulation Models, FHWA-RD-03-050. National Technical Information Service. Springfield, VA 22161

Harms, H. (2006) De oorzaken van de toename in filezwaarte na invoering van de 80 km/u maatregel (The causes of the increase of severity of congestion after the 80 km/h measure) in Dutch, Master thesis Delft University of Technology.

Hoogendoorn, S.P., Minderhoud, M.M. (2002) Motorway flow quality impacts of advanced driver assistance systems, Transportation Research Record, (1800), pp. 69-77.

Khane C.J. (1994) Preliminary evaluation of the effectiveness of antilock brake systems fro passenger cars. DOT HS 808 206.National Highway Traffic Safety Administration Washingtom DC
Minderhoud, M.M., (1999). Supported driving: Impacts on motorway traffic flow, TRAIL Thesis Series T99/4. Delft, The Netherlands
Ossen, S., S. Hoogendoorn (2007) Driver Heterogeneity in Car-Following and Its Impact on Modeling Traffic Dynamics, TRB 86th annual meeting
Polidori, C., G. Gatti (2007), Safety Handbook for Secondary Roads, Report RIPCORD – ISEREST consortium, Politecnico de Bari
Prigogine I., R. Herman (1971) Kinetic Theory of Vehicular Traffic, Elsevier New York
Saka, A.A., Glassco, R.A. (2001) Modeling traffic safety benefits of intelligent transportation system technologies at truck inspection facilities: Microscopic simulation Transportation Research Record, (1779), pp. 173-181.
Smeed, R.J. (1949) Some statistical aspects of road safety research, J. Royal Stat. Society A 112 (1) 1 - 23
Van Zuylen, H.J. (1976) A comparison of the pedestrian casualties in some European countries, unpublished report Traffic Studies Group, University College London.
Van Zuylen, H.J., V.L. Knoop, H. Tu. F. Viti. E. Jagtman and H. Smit (2007) Accessibility of the Harbor of Rotterdam, TRAIL Research school, Delft (to be published).
Van Zuylen, H.J. and S.P. Hoogendoorn (2007) A probabilistic calculation of queue lengths at signalized intersections, ITS World congress Beijing (to be published)
Viti F. and van Zuylen H. J., 2004. Modeling queues at signalized intersections. Transportation Research Record 1883, pp. 68-77. National Research Council, Washington
D.C.Wilde, G. J. S. (1982) The Theory of Risk Homeostasis: Implications for Safety and Health. Risk Analysis, 2, 209-225.
Zeeger C., Reinfurt D., Hummer J., Herf L., Hunter W. (1987) Safety effects of cross section design fort two lane roads – FHWA-RD-87/008.

Mechanism of Dynamic Multiple Ant Colony Optimization for Mass Service Routing and Assignment in Urban Rail Transit Stations

Xiao-lei ZOU[1]，Rui-hua XU[2]

(School of Transportation Engineering, Tongji Univ. Shanghai, 200092)

[1] School of Transportation Engineering, Tongji University, Shanghai, 200092 China
PH: 86-21-69589877; FAX: 86-21-69589487; e-mail: zouxiaolei@mail.tongji.edu.cn
[2] School of Transportation Engineering, Tongji University, Shanghai, 200092 China
PH: 86-21-69589487; FAX: 86-21-69589487; e-mail: rhxu@mail.tongji.edu.cn

Abstract: The study on mass performance in Urban Rail Transit (URT) stations helps to enhance efficiency and security of passenger service, and supports design and amelioration of stations. The procedure of Mass Service Routing and Assignment (MSRA) is similar to the routing of ant colonies looking for food. This paper analyzes the service network of URT stations and the hierarchical structure of the mass, and brings forward Dynamic Multiple Ant Colony Optimization (DMACO) to study MSRD. In DMACO, the mass is divided into colonies corresponding with passenger flows, and original Ant Colony Optimization is used in all the flows simultaneously. The effects, from other colonies, guidance, capacity of network, and preference of ants to pheromone, are imported in state transition rules, in order to meet MSRA in condition of multi-hierarchy, multi-process passenger service in a URT station.

Keywords: *Mass Performance, Routing, Assignment, Dynamic Multiple Ant Colonies Optimization, Mass Simulation*

1 Introduce

Stations are responsible for mass passenger service. They are the elementary unit of the operation of Urban Rail Transit (URT) as well as the point for passenger flows of arrival, getting service, and departure. Mass performance is bound up with passenger service. Therefore, studies on mass performance, passenger service and interaction between them in URT stations help to enhance the efficiency and security of passenger service, and support the design and amelioration of stations and service processes.

Passenger service processes and facility arrangement in URT stations can be drawn into a service network. Mass performance is mainly the procedure of routing and assigning in the service network, planning the path to service nodes, moving towards nodes and receiving service.

Mass service routing in the station is similar to the routing of the ant colonies looking for food. So in this paper an improved Ant Colony Optimization (ACO) is brought forward to imitate Mass Service Routing and Assignment (MSRA).

The first ACO system was introduced by Marco Dorigo in his Ph.D. thesis (1992), called Ant System (AS). Since 1995 Dorigo, Gambardella and Stützle have been working on various extended versions of the AS paradigm, such as Ant Colony System (ACS), MAX-MIN Ant System (MMAS) and new hybrid versions of ant colony optimization with a local search. In 1999, the Ant Colony Optimization metaheuristic was defined.

Since 2000, MACO is derived from ACO, namely several ant colonies making routing at the same time. Hidenori Kawamura (2000) from Hokkaido University in Japan, Kwang Mong Sim (2002) from Hong Kong University, Issmail Ellabib (2007) from Waterloo University in Canada, and others, and some research institutes and researchers have proposed MACO independently. They tried to find the optimal solutions with several colonies working in coordination.

ACO and MACO are applied to the Traveling Salesman Problem and to the quadratic assignment problem, but little to MSRA.

This paper brings forward Dynamic Multiple Ant Colony Optimization (DMACO), making an attempt to solve a MSRA problem, in which the mass is divided into flows according to the service processes and ACO is used in all the flows simultaneously. Pheromones, evaporations and state transition rules are also adopted. The factors of interaction among colonies, of the guidance and the preference of passengers to pheromones, are introduced to state transition rules, in order to meet the mass service routing in condition of multi-colony, multi-service flow and the complicated composition of the mass.

2 Passenger Service Process and Service Network in station

2.1 Passenger service processes

Passenger flows can be classified into three kinds according to the service process, which is shown in Table 1.

Table 1 Passenger Classifying and the Service Process

Passenger Flows	Service Process
Entering Flow	Entering→(Buying tickets)→Checking in→Waiting→Departure
Leaving Flow	Arrival→Checking out→Leaving
Transferring Flow	Arrival→(Checking out)→(Buying tickets)→ (Checking in)→Waiting→Departure

The stages enclosed with brackets are optional. For the entering and transferring

flows, the step of buying tickets is optional, because passengers with pre-game tickets or traffic cards do not need to buy the tickets. The other passengers could choose automatic or artificial ticket selling. And in some hub stations, passengers do not need to have tickets checked or buy new tickets when transferring.

2.2 The network of the passenger service

Areas in stations can be classified into passenger service areas (ticket areas, checking areas, waiting areas, etc.), transiting areas (entrances and exits, stairs, etc.) and passageways. The relationships among different areas are periodical and directional, with the constraint of the passenger service processes. Therefore, the service processes and areas can be described as a network structure, and passengers receiving service is a procedure of routing and moving in the network.

Take a certain station for example. Passenger service areas are the nodes of black circles, transiting areas are the nodes of white circles and passageways are used as edges. The passenger service network for transfer is shown as an example in Figure 1.

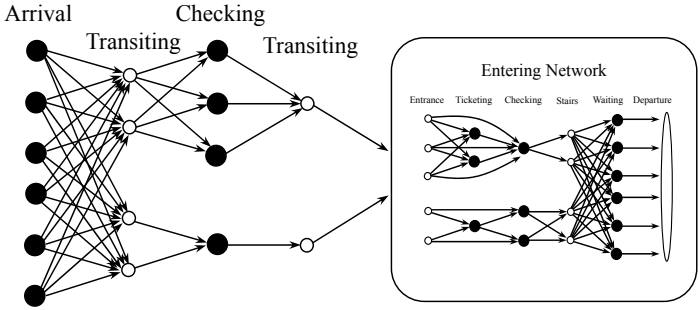


Figure 1. Demonstration of Transferring Network Structure

In the network, the service stages constitute the passenger service processes while the directional edges constitute the routes for the service. The weight of the edge is mainly decided by the variables of distance and service time. The network has its service capacity. There may be several routes in one complete service process. The best or relatively fine route can be found according to some criteria.

3 The hierarchical structure of the mass and interactions among flows

3.1 The hierarchical structure of the mass

The mass in the station can be sorted into several kinds of flows in terms of the service they receive and their direction in the network. Passengers with the same origin and destination as well as the same service process are in one flow. For example,

there are entering, transferring and leaving passenger flows. Each flow could be divided into several sections in accordance with the stages of the service process. Besides, passengers usually travel with company. Therefore, each section consists of many groups as well as single passengers. A demonstration of the masses in a station is shown in Figure 2.

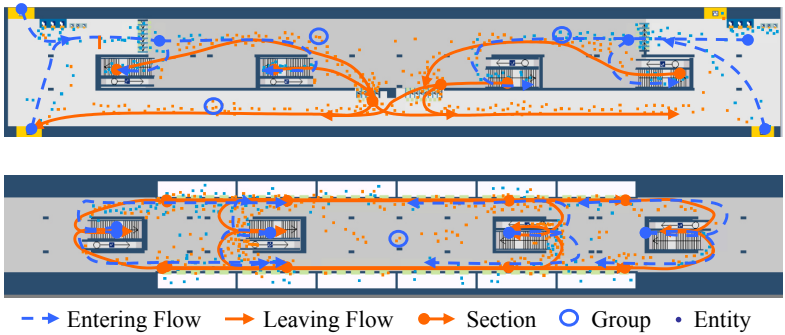


Figure 2. Demonstration of the mass in station

In Figure 2, the mass consists of two main flows, the entering (the dotted arrow) and leaving flow (the solid arrow) according with the service processes in Figure 1. Taking the entering flow for instance: it can be divided into sections such as Entering-Ticketing, Ticketing-Checking, Checking-Transiting (between stories), Transiting-Waiting, and Waiting-Departure. Each section stands for one stage in the entering flow service process. The entering flow also has several branches in some of the sections. There are many groups and entities in each section.

A hierarchical structure of the mass can be established, shown in Figure 3.

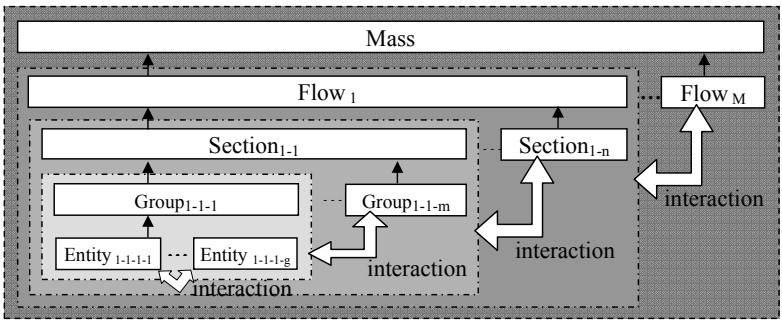


Figure 3. The hierarchical structure of the mass

The structure consists of five hierarchies: Mass, Flows, Sections (with branches),

Groups and Entities (passengers). The relationship of the hierarchies can also be described as following sets.

$$Mass = \{Flow(i) \mid i \in N\}$$

$$Flow(i) = \{Section(j) \mid j \in N\}$$

$$Section(j) = \{Group(k) \mid k \in N\}$$

$$Group(k) = \{Entity(g) \mid g \in N\}$$

3.2 Interactions among flows

The elements in the same hierarchy produce effects on each other as shown in Figure 3. Entities have strong cohesion when in a same Group, and repulsion when in different Groups. And Entities in a same Group play different roles, such as leader and followers. The same happens with Groups even much more distinct. While in a higher hierarchy, the flows interact with each other in converging, splitting, interfering and crossing.

The interactions and effects among elements are based on different performance of passengers. In hierarchies of Groups and Entities, the dynamics of pedestrians play a crucial role, while in hierarchies of Sections and Flows, the passengers' routing and assigning in service processes are the crux of the matter. On the contrary, the given and potential interactions and effects among flows are the crucial factors that influence the passengers' service routing and assigning.

The relationship between interactions among flows and routing of passengers is quite complicated. But in general the evolution of one flow is similar to that of an ant colony in food seeking. The interactions and affects among flows in stations have similarities with the interactions and effects among ant colonies when looking for different food sources.

This paper focuses on the hierarchy of flows and the interactions among flows, and tries to describe and explain the relationship between interactions among flows and MSRA by the concept of DMACO.

4 Mechanism and description of DMACO for MSRA

4.1 The mechanism of MACO and DMACO

MACO is derived from ACO, by which optimal solutions could be found with several colonies working simultaneously with the same origin and destination. The ant colonies interact with each other through colony level information exchange.

Their basic thoughts are as follows:

(1) Several ant colonies make a route choice in the same network at the same time, with the same origin and destination. In the Traveling Salesman Problem, all nodes should be visited by every ant and make a complete loop.

(2) Each ant colony has its independent routing function and works in parallel.

(3) The pheromones of all ant colonies lie and evaporate according to the same rule.

(4) The ant colonies exchange information through positive or negative pheromones. The influence is carried unidirectionally.

Since MSRA is based on the routing of each entity in the flow, which is similar with the condition in an ant colony, we propose DMACO to solve the MSRA problem in stations. There are several differences in DMACO compared with previous studies:

(1) Each ant colony has its own origin and destination, which is unidirectional and not forming a loop.

(2) The ants vary in size; therefore the volume of pheromone left corresponds with the size of the ant.

(3) The concepts of friend-colony and enemy-colony were introduced to present the different relations among different colonies.

(4) The ant has a lifecycle which can be generated and disappear. Thus the numbers of ants change with time.

(5) The weight of each edge changes dynamically.

4.2 The definition of the ant colonies

Suppose each passenger flow is an ant colony, and there are M colonies (*Colony 1*, *Colony 2*, … , *Colony M*) corresponding with the M flows. Suppose that at the point of time t, there are m groups in each passenger flow (each group contains one entity or several companions), then each ant colony has m ants. Therefore, ant k in *colony h* is described as *ant (h, k)*. At the time t, the total number of ants is $M*m$. When M=1, it is just the original Ant Optimization without colony-level interactions.

There are two types of relationships between two colonies: friend or enemy. The colonies with similar service processes are friend colonies, otherwise enemy colonies. For instance, colonies that represent entering flows are friend colonies but they are enemy colonies to the ones representing leaving flows. The friend colony has a positive attractive effect by pheromone while the enemy colony has negative exclusive effect.

The size of *ant (h, k)* is c^{hk}, represents the number of people in the group. *Ant (h, k)* is generated when *Group (h, k)* gets to its origin and disappears when *Group (h, k)* gets to its destination.

4.3 The structure of a service network

Suppose $G = (N, A)$ is the graph of a service network, where $N (|N| = n)$ is the set of nodes, A is the set of edges. The edge $A(i, j)$ is bi-directional. $\tau_{ij}^{h}(t)$ is the intensity of pheromone of *Colony h* on $A(i,j)$ at time t, through which the ants choose

the route step by step. The edge with higher intensity of pheromone will be chosen with more probability.

The weight of *A(i,j)* $L_{ij}(t)$ takes the distance l_{ij} between *i* and *j* as well as the average waiting and service time $\bar{t}_j(t)$ at time *t* into account. $\bar{t}_j(t)$ is converted to distance according to the average speed of the passenger flows: $l'_{ij}(t) = \bar{v}_{ij} \cdot t_j(t)$.

$$L_{ij}(t) = l_{ij} + l'_{ij}(t) \qquad (1)$$

There may be bidirectional passenger flows on edge *A(i,j)*. The density of passengers on edge *A(i,j)* $d_{ij}(t)$ at time *t* has influence on the routing of *ant (h, k)*. When $d_{ij}(t)$ approaches saturation, the intention of choosing *A(i,j)* by *ant (h, k)* $I_{ij}(t)$ decreases obviously.

$$I_{ij}(t) = f(d_{ij}(t))$$

The relationship can be illustrated by Figure 4. The quantizing relationship between the intention of choosing an edge and the density of passengers on the edge needs further study and experimental verification.

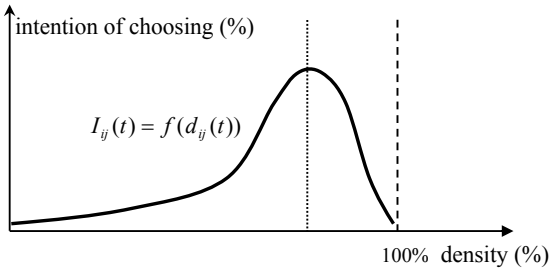


Figure 4. Demonstration of the relationship between intention and density

4.4 Pheromone and Evaporation

The intensity of pheromone of each colony is stored on the edges separately.

Let $\tau_{ij}^{h}(t)$ be the intensity of the pheromone laid by *Colony h* on *A(i,j)* at time *t*.

At the beginning, the guidance information is introduced as a constant Q_0.

$$\tau_{ij}^{h}(0) = Q_0 \qquad (2)$$

After every interval Δt, the intensity of the pheromone of *Colony h* on *A(i,j)* is updated as follows:

$$\tau_{ij}^{h}(t+\Delta t) = \tau_{ij}^{h_left}(t+\Delta t) + \Delta\tau_{ij}^{h} \qquad (3)$$

where $\tau_{ij}^{h_left}(t+\Delta t)$ is the intensity of pheromone of *Colony h* left on *A(i,j)* after evaporating in the interval Δt, and $\Delta\tau_{ij}^{h}$ is the increment of pheromone intensity laid by *Colony h* on *A(i,j)*.

$\tau_{ij}^{h_left}(t+\Delta t)$ is defined as follows:

$$\tau_{ij}^{h_left}(t+\Delta t) = (1-\rho)\times\tau_{ij}^{h}(t) \qquad 0<\rho<1 \qquad (4)$$

where ρ is a coefficient representing the evaporation rate of the pheromones between time *t* and *t*+ Δt.

$\Delta\tau_{ij}^{h}$ is defined as follows:

$$\Delta\tau_{ij}^{h} = \sum_{k=1}^{m}\Delta\tau_{ij}^{hk} \qquad (5)$$

Here, *ant(h,k)* releases the intensity of pheromone $\Delta\tau_{ij}^{hk}$ on *A(i,j)*:

$$\Delta\tau_{ij}^{hk} = \begin{cases} \dfrac{c^{hk}Q}{L_{ij}} & \text{if } ant(h,k) \text{ has passed through } A(i,j) \text{ or completed the tour} \\ 0 & \text{otherwise} \end{cases} \qquad (6)$$

where Q is a constant. The ant with a larger size c^{hk} can release more pheromone in Δt.

4.5 State Transition Rules

After getting service at a certain node, *ant(h,k)* moves to the next node according to the State Transition Rules. For *ant(h,k)* at *i*, the State Transition Rules could be defined as Hidenori Kawamura (2000) defined them, but modified as follows:

$$p_{ij}^{hk}(t)=\begin{cases}\dfrac{[\pi_{ij}^{h}(t)][\eta_{ij}(t)]^{\beta^h}}{\sum_{l\in N_i^k}[\pi_{il}^{h}(t)][\eta_{il}(t)]^{\beta^h}} & if\ j\in N_i^{hk}\\ 0\end{cases} \tag{7}$$

$$\pi_{ij}^{h}(t)=\prod_{l=1}^{M}[\tau_{ij}^{l}(t)+C(h)]^{\alpha(h,l)} \tag{8}$$

$$\eta_{ij}(t)=1/L_{ij}(t)\cdot I_{ij}(t)\cdot g_{ij} \tag{9}$$

where N_i^{hk} is the feasible neighbourhood for *ant(h,k)* from node i, which represents the set of nodes that haven't been visited before. $\pi_{ij}^{h}(t)$ is the extent that all pheromones on *A(i,j)* at *t* affecting the decision of *ant(h,k)* to choose *j* for the next node. *C(h)* represents the degree of insentistivity to the pheromones for *Colony h*, and it can prevent the ant from being overly affected when $\tau_{ij}^{l}(t)$ approaches zero. $\eta_{ij}(t)$ is the greedy heuristic term, $I_{ij}(t)$ the density of ants and g_{ij} the guidance sign on *A(i,j)*at time *t*, all of which affect the expectation of ant agents from *i* to *j* at time *t.* $\alpha(h,l)$ is the parameter indicating the influence on *Colony h* made by the pheromone of Colony *l.* It is positive when *Colony l* is friend and negative when it is enemy. β^h is the parameter indicating the degree of influence made by $\eta_{ij}(t)$ on the decision of *ant(h,k)*.

4.6 Parameter Setting

$\alpha(h,l)$ should be set reasonably to avoid passenger flows being too intensive. If $\alpha(h,l)$ is overestimated, the ant colonies will depend too much on pheromone, which may lead to the situation of colonies overly gathering in some certain routes. While underestimated, the ant colony could not take full use of the experience (pheromone) and lead to the situation of passenger flows being too scattered. $\alpha(h,l)$ among different colonies could be set at different values. According to a previous study, $\alpha(h,l)$ is usually between 1 and 5, and β^h between 1 and 10.

ρ represents the evaporating speed of the pheromone. If ρ is overestimated, useful information stored in the pheromone would be lost soon (pheromones exponentially decrease with time); if ρ is underestimated, the faulty information

would not be forgotten easily. According to a previous study, ρ is usually between 0.4~0.5.

Q represents the total amount of pheromone of each ant in a cycle and it has little influence on the best solution. In this paper, Q=100.

g_{ij} is the term of guidance information. When there is guidance on the edge, g_{ij}=0.8, otherwise g_{ij}=0.2.

5 Procedure of Simulating MSRA by DMACO

5.1 The Role of DMACO in Simulation

The routine ACO and MACO are used to find the optimal solution. But in this paper, DMACO is used in simulation to reproduce the procedure of passenger flow routing and assigning on the service network. The stage of routing and that of the passenger flow assignment on all routes are smoothly joined; thesaturation density of each edge, the guidance of heuristic information, and colony level interaction are introduced as well, in order to reflect the interactions among flows and approach a dynamic balance in the hierarchy of flows in the network.

However, there are a lot of things to do in the simulation of the whole procedure of passenger service in a station. The passengers' performance in the station contains three levels from macroscopy to microscopy: service routing, path finding in space, pedestrian dynamics, and service performance. In the simulation and through DMACO, the levels of service routing and service process control are solved by the lifecycle of ants. For an entity, after an edge has been chosen, what path to go through, how to move along the edge, and how to get service are the businesses of path finding, pedestrian dynamics and service performance.

Figure 5 shows a hierarchical structure of passenger performance and feasible solutions for imitating.

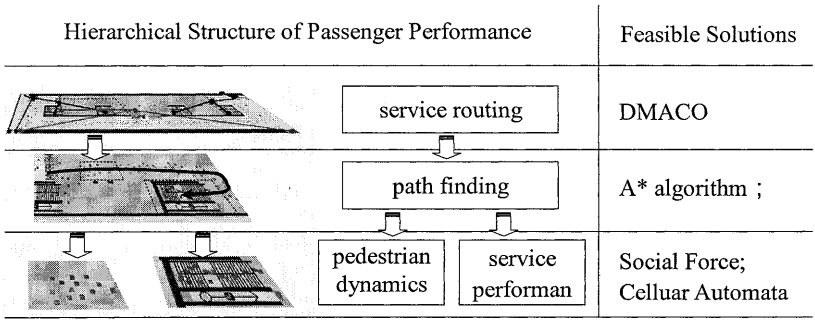


Figure 5. The hierarchical structure of passenger performance and feasible

The following stages are included in the lifecycle of an ant agent in a network: entering the network, moving along the edges, getting service at a certain node, leaving and moving to the next node, and leaving the network.

The routing of ants in a colony forms the routing of a passenger flow.

5.2 Process of DMACO for simulating passenger flow routing and assigning

The process of simulating passenger service flow routing and assigning by DMACO is as follows:

Step 1. Initialize

t=0

Generate service network $G = (N, A)$

Generate M ant colonies

For h=1 to M Do

For i=1 to n Do

For j=1 to n Do

Set the initial pheromone of *Colony h* on *A(i,j)* $\tau_{ij}^{h}(0) = Q_0$

Step 2. Dynamical control of service procedure

For t=t_0 to t_{MAX} step=Δt Do

For h=1 to M Do

For i= 1 to n Do

For j=1 to n Do

Update the pheromone of *Colony h* on *A(i,j)* is $\tau_{ij}^{h_left}(t)$.

New ants are generated in *Colony h* according to the distribution of arriving passenger flows, and the ants vary in size.

For k=1 to m Do

Case 1(*ant(h,k)* leaves i)

Choose next node j according to Stage Transition Rules

ant(h,k) moves along *A(i,j)* (turn to the path finding and pedestrian dynamics modules)

Case 2 *ant(h,k)* moves along *A(i,j)* then

Turn to the path finding and pedestrian dynamics modules

Case 3 (*ant(h,k)* arrives at j from i)

Update the pheromone $\tau_{ij}^{h}(t)$ on *A(i,j)*

Update the feasible neighborhood N_j^{hk} of *ant(h,k)*

ant(h,k) receives service at j (turn to the service performance module)

Case 4 (*ant(h,k)* arrives at the destination)

Calculate the length of the tour of *ant(h,k)* and update the shortest route of *Colony h*

Exclude *ant(h,k)* from *Colony h*

5.3 Time Complexity Analysis

Suppose M<<m, the time complexity of calculation in each step is $O(Mmn^2)$. When M=1, then it's $O(mn^2)$, the time complexity of original ACO. The number of the nodes and of the colonies in rail transit stations are not big. . Therefore the calculation is mainly decided by the number of ants in each colony. In this paper, the group of passengers traveling together is regarded as a single ant, which could reduce the number of ants and enhance the calculation efficiency.

Summary

This paper analyzed the service network of a URT station, the hierarchical structure of the mass, and the hierarchical structure of passenger performance. Based on the similarities between mass getting service in a station and ant colonies looking for food, DMACO is defined and the mechanism and process of using DMACO in simulating passenger service flow routing and assigning are introduced. If DMACO combined with path finding, pedestrian dynamics and service performance, the whole procedure of passenger service could be simulated.

Further studies on DMACO are needed as follows:

(1) the setting of all kinds of parameters

(2) the extent of interaction among the ant colonies

(3) the influence on the results by different calculation sequences of colonies and ants.

With verification and modification by programming and practical research, DMACO could be applied in a more complex network for routing and assigning of traffic flows.

References

Hidenori Kawamura, Masahito Yamamoto, Keiji Suzuki, Azuma Ohuchi.Multiple Ant Colonies Algorithm Based on Colony Level Interactions. IEICE TRANS.FUNDAMENTALS, VOL.E83-A, NO.2 FEBRUARY 2000.

Issmail Ellabib, Paul Calamai, Otman Basir. Exchange strategies for multiple Ant Colony System. Information Sciences: an International Journal, Volume 177, Issue 5 (March 2007), Pages: 1248-1264

Kwang Mong Sim, Weng Hong Sun. Multiple Ant-Q colony Optimization for Network Routing. IEEE Computer Science 2002.

A Study of Parking Behavior and Parking Information Requirements in Shanghai CBD

Yikui MO[1]，Baoyu ZHANG[2]，Kefei YAN[3]

[1]College of Civil Engineering, Shenzhen University, Nanhai Road 3688, Shenzhen,P.R.C, 518060; Phone and fax numbers: 86-21-6958-4580; email: moyikui@hotmail.com

[2]School of Civil Engineering & Architecture, Chongqing Jiaotong University, Xuefu Road 66, Chongqing, P.R.C, 400074; Phone and fax numbers: 86-21-6958-4612; email: zbyresearch@yahoo.com.cn

[3]School of Transportation Engineering, Tongji University, Siping Road 1239, Shanghai, P.R.C, 200092; Phone and fax numbers: 86-21-6598-0487; email: yankf@mail.tongji.edu.cn

Abstract: Searching for a parking space has become an unavoidable task for drivers in Shanghai CBD. Providing drivers with parking availability information will reduce their searching time and thus reduce traffic on the road during congested periods. However, there is a need for more field study/surveys in order to improve our understanding of parking behavior and drivers' reactions to the parking information provided.

The main goal of this paper is to analyze parkers' behavior in choosing a parking lot in the Shanghai CBD and to develop a better understanding of the users' need for parking information. The Xujiahui Area, one of the CBDs in Shanghai where a Parking Guidance and Information System (PGIS) was deployed in December 2003, has been chosen to conduct the case study. In April 2006, a scientifically designed parking survey was carried out to collect data about parkers' revealed-preference and stated-preference. The results of 783 drivers' responses are used in this study.

The analysis of this study shows that parking difficulty mainly resulted from the lack of parking information. The drivers' behavior in choosing a parking lot is mainly influenced by walking time, availability of parking spaces, parking fees and trip purposes. In particular, this is the first time that a study has investigated the performance of a PGIS operating in China.

Key words: *parking guidance and information system, parking behavior, survey*

1 Introduction

Shanghai is the largest city in China with a population quickly approaching 17 million. The remarkable economic developments in Shanghai have greatly increased vehicle ownership, and consequently demand for auto travel in recent years. The lack of parking spaces or capacity is becoming a huge problem in the Shanghai CBD. Providing drivers with information on parking availability will reduce their search time and thus also reduce their impact on congestion. In December 2003，a Parking

Guidance and Information System (PGIS) was deployed to alleviate parking difficulties in the Xujiahui Area of Shanghai. The system mainly relies on variable message signs to deliver information about the direction to and availability of spaces in parking lots. The signs are located some distance away from parking lots on main streets in the Xujiahui Area. The available information of the parking lots is manipulated by the operator of each parking lot. The main functions of the PGIS are reducing the travel time of looking for parking place, increasing the parking facility utilization rate, reducing the traffic caused by the drivers who are looking for a parking lot, balancing parking demand during time and parking lots. In order to achieve better targeting and operation of the PGIS, there is a need for better understanding of the parking behavior and parking information needs in the CBD.

Since understanding of parking behavior and parking information needs is fundamentally important for any parking studies, some studies have already been published on this topic. I Tsopelas, *et al.* (2001) analyzed the effectiveness of a new PGIS in Manchester City and developed a better understanding of the user needs for parking information. Hualiang Teng, *et al.* (2001) discussed the impact of levels of parking information on parking difficulty, with the development of ordered probit models. Teknomo, *et al.* (1997) analyzed parkers' behavior in choosing a parking location in the CBD of Surabaya.

Due to the differences in culture and social-economic background, parking behavior varies from country to country. Unfortunately, a very limited number of studies have been conducted on this subject in China. The paper published by Hongzhi Guan, *et al.* (2003) is the only study that deals with parking behavior. We can not find any study of how the PGIS performs in practice and how drivers react to the information provided in China. Therefore, the purpose of this study is to investigate the parking behavior and parking information needs based on the findings from a survey conducted at the Xujiahui Area, the most important CBD in Shanghai.

2 Methodology

2.1 The surveyed parking lots

Located in the southwestern part of Shanghai, the Xujiahui Area is a famous travel place with a long history and accumulation of culture. As there are many businesses, modern commercial plazas and restaurants, the Xujiahui Area is also a commercial and entertainment center. With the high volume of traffic, including many people unfamiliar with the street system and parking facilities, there is a need to efficiently guide drivers to available parking spaces. Therefore it deployed a PGIS in December 2003. Twenty-one car parks are included in the system. The survey covers six parking lots in the area, including four underground parking lots and 2 on-ground parking lots. These six parking lots were chosen because of their high capacity and important location on the main street. During the survey, drivers were interviewed in the car parks by a group of surveyors.

2.2 Data collection

In general, data collection can be divided into three sections:

(1) Drivers' general information.

(2) Drivers' preference in choosing a car park.

(3) Drivers' attitudes to parking information.

To get the data, a questionnaire interview survey was done in the six parking lots. The interview was conducted on April 7, 2006. Drivers' general information is mainly information about sex, age, occupation, monthly income, trip purpose, duration of parking and visit frequency. The drivers' preference was taken by letting them give the main reasons for the choice of a car park. Respondents were permitted to give multiple answers to this question. To analyze drivers' attitudes to the parking information, respondents were asked to provide details about their knowledge level of parking information before starting a trip, main reason for encountering parking difficulty, reactions to the PGIS, and their request for parking information. After reducing the invalid data, the behavior and preference data of 783 drivers were collected.

3 Results and Analyses

3.1 Drivers' general information

This section describes the general characteristics of the drivers that were interviewed during the survey. The sample of 783 drivers consisted of 602 males and 181 females. Age was classified into four groups: under 30, 30-45, 46-60 and over 60. The distribution of these four age groups is: 21.07%, 65.78%, 12.77% and 0.38%. It can be seen that the 65.78% of the drivers are in the age range of 30-45 years old. The drivers in the “over 60” are very few. As shown in Table 1, the majority of the female drivers (28.73%+67.96%) are younger than 45 years old.

Table 1. Percentage of females and males for each age group

Age Group	Females	Males	Total
<30	28.73%	18.77%	21.07%
30－45	67.96%	65.12%	65.78%
46－60	2.76%	15.78%	12.77%
>60	0.55%	0.33%	0.38%
Total	23.12%	76.88%	100%

Approximately 97% (100%-2.21%) of the drivers interviewed traveled to the Xujiahui Area from Shanghai City. As shown in Table 2, 41.51% of the drivers came from the Xuhui District. 56.36% of the female drivers came from this district. In particular, 30.02% drivers traveled from four districts adjacent to the Xuhui District.

Table 2. Places of origin by gender

District	Females	Males	Total
Xuhui District	56.36%	37.04%	41.51%
Changning District	4.97%	6.48%	6.13%
Minhang District	7.18%	17.11%	14.81%
Jing'An District	0.00%	2.99%	2.30%
Pudong District	9.39%	5.99%	6.78%
Other Districts	19.89%	27.07%	25.41%
Other cities	2.21%	3.32%	3.06%

Shanghai City includes eighteen districts and one county: Xuhui, Changning, Minhang, Jing'An and Pudong District. Changning, Minhang, Jing'An and Pudong District are four districts adjacent to Xuhui District. The Xujiahui Area is the center of the Xuhui District.

The survey results in Table 3 indicate that 43.55% of the drivers park for a period between 1 and up to 3 hours. This category comes as the first choice for both female and male drivers. The second choice,i "3 to 5 Hours", includes 31.16% of the responders. Adding the proportion of the drivers who park for a period less than one hour, the majority of drivers (93.49%) stay no more than 5 hours in the car park.

Table 3. Duration of stay in the car park by gender

Duration Group	Females	Males	Total
Up to 1 Hour	8.86%	21.75%	18.77%
1 to 3 Hours	49.72%	41.70%	43.55%
3 to 5 Hours	37.56%	29.24%	31.16%
5 to 6 Hours	2.21%	1.16%	1.40%
More than 6 Hours	0.00%	3.99%	3.07%
N/A	1.65%	2.16%	2.04%

Within the framework of the survey, drivers were asked the reason for visiting the Xujiahui Area the day they were interviewed. Table 4 demonstrates the results of the drivers' responses. In this table the mean of "work" trip is different from "business" trip. "Work" trip means the drivers' need to drive to there to work on weekdays. "Business" means drivers drive there only for temporary business. Overall, "Shopping" (47.64%) and "Dining or Recreation" (22.73%) were the major reasons for drivers to travel into the Xujiahui Area. The percentages of female and male drivers visiting Xujiahui Area for "Shopping" and "Dining or Recreation" are very close. In particular, only 2.94% of the drivers declared that "working" was the reason for visiting Xujiahui Area the day of the interview. A car is still a luxury and driving to work is comparatively expensive in China, so this result is reasonable,and quite different from the case in developed countries. More specifically, 12.62% of the male drivers visited the Xujiahui Area for "business", but none of female drivers. The

results of the data analysis show that all of the drivers that traveled for "work" park for a period more than 3 hours and approximately 70% of the drivers that traveled for "shopping" and "Dining or Recreation" stay no more than 3 hours in the car park.

Table 4.Trip purpose of coming to the Xujiahui area by gender

Trip purpose	Females	Males	Total
Work	1.10%	3.49%	2.94%
Business	0.00%	12.62%	9.70%
Social	2.21%	1.66%	1.79%
Dining or Recreation	28.17%	21.10%	22.73%
Shopping	52.48%	46.18%	47.64%
Traveling	7.74%	2.49%	3.70%
Others	3.32%	4.65%	4.34%
N/A	4.98%	7.81%	7.16%

As for trip frequency, six levels were surveyed: less than one time a month, one time a month, two times a month, 1-2 times a week, 3-4 times a week and 5-7 times a week. The distribution of the frequency is: 1.79%, 21.20%, 12.00%, 43.80%, 5.87% and 13.92%. It shows that more than 40% of the respondents travel into Xujiahui Area 1-2 times a week. Moreover, the survey findings suggest that drivers traveling for "work" visit the Xujiahui Area more frequently than those that do not. Drivers that stated "shopping" and "Dining or Recreation", as their usual reason for visiting the Xujiahui Area appear to have a more flexible travel pattern.

Table 5. Frequencies of visits to Xujiahui area by gender

Frequency	Females	Males	Total
5-7 times a week	8.84%	15.45%	13.92%
3-4 times a week	2.76%	6.81%	5.87%
1-2 times a week	52.49%	41.19%	43.80%
2 times a month	10.49%	12.45%	12.00%
one time a month	19.89%	21.60%	21.20%
Less than one time a month	4.43%	1.00%	1.79%
N/A	1.10%	1.50%	1.41%

In China, generally, some of the vehicles are owned by various organizations and business enterprises. The non-private car drivers use their company or organization's vehicles for which all associated costs, including parking expenses, will be reimbursed. Private owners are more sensitive to parking cost than non-private drivers are. Who pays the bill significantly affects parking behavior, which must be considered in the development of parking policy. Parking patrons can be divided into two

groups-private and non-private vehicle owners. Table 6 shows that the majority of parking fees were paid by private owners. And there is a difference in distributions between male and female drivers. The parking fees paid by private owners is 84.04%, while the most of the female drivers (95.58%) pay the parking fees themselves.

Table 6. Responses to "who is paying the parking fee?" by gender

Who is Paying the Parking Fee?	Females	Males	Total
Individual	95.58%	80.57%	84.04%
Organization	4.42%	19.43%	15.96%

3.2 Drivers' preference in choosing a car park

In order to identify and analyze the opinions and attitudes of drivers regarding the choice of car park in the Xujiahui Area, respondents were asked to provide the main reasons for their choices, and they were permitted to provide more than one answer. As trip purpose is also an important factor which influences parkers in choosing a car park, this paper analyzes the relationship between trip purpose and reasons for choosing a car park.

Table 7. Reasons for choosing a car park by gender

Reason for choosing a car park	Females	Males	Total
Location	75.69%	78.74%	78.03%
Parking fee	38.11%	18.77%	23.24%
Availability of space	46.96%	47.84%	47.64%
Ease to park	16.57%	10.80%	12.13%
Security and comfort	22.10%	16.12%	17.50%
Others	0.55%	0.50%	0.51%
N/A	2.76%	3.49%	3.32%

Table 7 reveals the preference for choosing a car park. It is clear that "Location" is the key decision factor. Overwhelmingly, about 78.03% of the surveyed people stated that the distance between a parking lot and the destination is the driving force in their selection of parking facility. The second choice, "Availability of space", includes 47.64% of the responders. The percentages for females and males are close in these two categories. More specifically, 38.11% female drivers declared that "Parking fee" was one of the main reasons for choosing a car park, while the figures show that only 18.77% of the male drivers stated that. In addition, the figures show that women consider "Security and comfort" and "Ease to park" more than men.

Table 8 shows the relationship between trip purpose and reasons for choosing a car park. "Location" is still the first choice for each trip purpose. In particular, all of the drivers that traveled for "work" declared that "Location" was the key decision

factor, while only 14.56% of them chose “Availability of space”. The main reason for this is that many drivers visiting the Xujiahui Area for “work” have reserved parking spaces which they pay for monthly.. People driving for business stated that “Location” and “Availability of space” were main reasons for their choice; the percentages were 75.35% and 21.20% respectively. The drivers that traveled for “Dining or Recreation” and “Shopping” provided more reasons for choosing a car park. “Location”, “Availability of space” and “Parking fee” were the three main reasons. More specifically, about 60% believe “Availability of space” and 30% believe “Parking fee” is one of the main reasons for their final choice. Travelers who were more unfamilar with the Xujiahui Area considered “Availability of space” more than the others. They stated that “Availability of space” was as important as “Location”.

Table 8. Reasons for choosing a car park for each trip purpose

Reason for choosing a car park	Work	Business	Social	Dinning or Recreation	Shopping	Traveling	Others	N/A
Location	100 %	75.35%	82.38%	71.27%	85.06%	36.40%	100%	76.00%
Parking fee	0.00%	0.00%	20.05%	29.38%	33.85%	12.14%	0.00%	0.00%
Availability of space	14.56%	21.20%	30.02%	61.17%	63.09%	31.80%	0.00%	0.00%
Ease to park	0.00%	0.00%	0.00%	26.31%	12.90%	0.00%	0.00%	0.00%
Security and Comfort	0.00%	0.00%	0.00%	26.18%	22.48%	22.73%	0.00%	0.00%
Others	6.26%	3.45%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
N/A	0.00%	1.79%	0.00%	2.68%	1.79%	0.00%	0.00%	24.01%

3.3 Drivers' attitudes to parking information

Within the framework of the survey, respondents were asked about their knowledge level of parking information before starting a trip. The survey results in table 9 indicate 6.13% drivers may be very new to their destination areas and have no parking information available to have a pre-choice of an area, location, or facility for searching a space. It is implied that the drivers may have to pick a direction randomly and drive to search until finding a parking space. 24.26% of the drivers may be sure that vacant parking spaces may be easier to find in one area than the others, because they may have been familiar with their destination area somehow. 69.10% of the respondents are pretty sure that an area would have spaces available when they get there. Thus, they may just go directly to their familiar area without a searching process. Unless the parking lot that they usually use is full, they avoid changes in their travel patterns and carry on using their usual car park.

Table 9. Known parking information level before starting a trip

Known Parking Information Level	Percentages
I didn't know where to park	6.13%
I partially knew where to park	24.26%
I knew exactly where to park	69.10%
N/A	0.51%

Parking Guidance and Information systems (PGIS) have been developed in China for improving parking difficulty and reducing traffic congestion. Although these systems appear to operate quite successfully, there are several areas, such as drivers' reaction to them; the effects of such systems have not been investigated. In December 2003, the Xujiahui Area deployed a PGIS. After two years of operating, its effectiveness and drivers' reactions to it were investigated.

In order to investigate the effectiveness of the PGIS, drivers were asked to identify their parking difficulty the day of the interview. In fact, 77.39% indicated that parking is not at all difficult. However, 22.61% of the drivers did indicate that parking was somewhat or very difficult. This result indicates that the PGIS is not effective enough. The drivers who stated that parking was somewhat or very difficult were asked to provided the reasons for their parking difficulty. Table 10 shows that most of the parking difficulty didn't result from the lack of parking spaces but the lack of parking information.

Table 10. Reasons for parking difficulty

Reasons for Parking Difficulty	Percentages
Didn't know where to park	37.25%
Didn't know where to get into the car park	44.07%
The most desirable park was full	8.48%
Traffic jam	9.08%
Others	1.12%
N/A	0.00%

The survey results in table 11 show that 30.71% of the drivers stated that they often used the PGIS. 52.60% of the drivers said they used the PGIS occasionally, while most of them reported that this happens only when their desirable parking lot is full. In particular, 16.69% of the drivers declared that they never used the PGIS. Drivers were also asked how satisfied they are with the performance of the PGIS. It was interesting that the majority of the responders chose "Neutral", which means they were lukewarm about the system. Only 32.27% appear to be satisfied with the PGIS and 9.43% were unsatisfied. This indicates that the performance of the PGIS is not good enough.

Table 11. Drivers' responses to the PGIS in Shanghai CBD

Question	Answer		
Have you ever made use of the PGIS?	Often	Sometimes	No
	30.71%	52.60%	16.69%
Are you satisfied with the PGIS?	Satisfied	Neutral	Unsatisfied
	32.27%	58.30%	9.43%

Table 12 demonstrates the information that drivers think is important in order to make more informed choices. Real time information about available spaces appears to be the most popular piece of information that drivers would like to have. The information about space availability of parking lots provided by PGIS is operated by the operator of each parking lot. In fact, the process is arbitrary. So the system may not accurately present the real time information. Maybe that is the reason many drivers stated that real time information is most important. Approximately 2 out of 3 drivers think it is important. Information about "Location of the car park" is the second most favorable demand of all drivers with a percentage of 41.51%. Information about "Parking Fee" and "Directions to the car park" is also very popular, 33.59% and 28.61%, respectively. Another 17.75% of the drivers stated that "Distance to the car park" was important information. Information about traffic conditions in the area of the facility was requested from 15.07% of the drivers.

Table 12. Drivers' responses to the importance of parking information

Information needed	Percentages
Real time information about available spaces	65.90%
Location of the car park	41.51%
Direction to the car park	28.61%
Distance to the car park	17.75%
Traffic conditions in the area of facility	15.07%
Parking Fee	33.59%
Opening Hours	8.69%
Others	6.65%
N/A	8.56%

4 Conclusions- Further Work

The following conclusions can be drawn from the analysis of this study. Parking behavior is mainly influenced by walking time, available parking spaces, parking fee and trip purpose. 69.10% drivers traveling into the Xujiahui Area know where to park before starting their trips. They may just go directly to their familiar area without a searching process. Unless the parking lot that they usually used is full, they will not change their travel patterns and keep on using their usual car parking lot.

Even when a PGIS was deployed, 22.61% of the drivers did indicate that parking was somewhat or very difficult. Most of the parking difficulty didn't result from the lack of parking spaces but the lack of parking information. Most of the drivers refer to the PGIS on their way to the parking lot and about 80% of the drivers have used this system, but only 32.27% drivers were satisfied with it. This indicates that the performance of the PGIS is not so good and needs improvement in the future. The findings of the survey revealed that "Real time information about available spaces", "Location of the car park", "Parking Fee" and "Direction to the car park" are the most important information that drivers need.

Future efforts should focus on further evaluation of the PGIS and on developing more advanced methods to improve its performance.

References

Hualiang (Harry) Teng and John C.Falcocchio.(2001)."Parking Difficulty and Parking Information Needs for Off-Street Spaces in the CBD."*80th Annual Meeting of the Transportation Research Board*, Washington D.C,

Hongzhi Guan, Lanhui Liu.(2003)."A Parking Behavior Model in Metropolis' Downtown:A Case Study on Xidan Areas of Beijing."*China Civil Engineering Journal*, 36(1),46-51.

I Tsopelas, D J Ling, and T J McCarthy.(2001)."Driver Attitudes to the Parking Guidance Information System in Manchester."*UTSG 33rd Annual Conference*, Vol.1, Oxford 3-5.

Teknomo, Kardi, Kazunori Hokao.(1997)."Parking Behavior in Central Business District-A Case Study of Surabaya, Indonesia."*EASTS Journal*, 2(2), 551-570.

EXPLORING DRIVING SIMULATOR AS A TRAFFIC SAFETY TEST TOOL AT SIGNALIZED INTERSECTIONS

M. Abdel-Aty[1], X. Yan[2], E. Radwan[3], X. Wang[4], and X. Chen[5]

[1 3 4]Department of Civil & Environmental Engineering, University of Central Florida, Orlando, FL 32816-2450, Email: mabdel@mail.ucf.edu.
[2]Civil and Environmental Engineering Department, University of Tennessee, Center for Transportation Research, 309 Conference Center Bldg, Knoxville, TN 37996-4133, Email: yxd22222@yahoo.com.
[5]School of Transportation Engineering, Tongji University, Shanghai, China 200092, Email: Chenxh@mail.tongji.edu.cn

Abstract: This paper presents a novel perspective of driving simulator application on traffic safety: exploring driving simulator as a traffic safety test tool at signalized intersections. In support of this concept, a signalized intersection with as many important features (including roadway geometries, traffic control devices, intersection surroundings, and buildings) was replicated into the UCF (University of Central Florida) driving simulator. A series of driving simulator experiments for the intersection were conducted to determine if the subjects' speed behaviors and traffic risk patterns in the driving simulator were similar to what were found at the real intersection. The experiment results showed that the advanced driving simulators can be used as a valid tool to assess traffic safety at signalized intersection.
Key words: *driving simulator, signalized intersection, simulator validation, geospecific database modeling*

1. Introduction

The driving simulator applications in traffic engineering can be categorized into two classes. One is to utilized designed roadway environments to test drivers' behaviors such as gap acceptance behavior, passing maneuver, crash avoidance study, and driving distraction due to mobile phones. The other focused on using virtual reality to evaluate the effect of highway design on driver's behavior in virtual reality in order to identify design drawback in traffic safety and search for improvement countermeasures. For example, Orit et al (2006) used a high-fidelity driving simulator to investigate the effects of improved versus unimproved intersections on driving performance of young and senior drivers. Bella (2005) evaluated the effectiveness of temporary traffic signs on driving behaviors on highways in virtual reality. In the second type of driving simulator studies, the fidelity requirement for simulated driving environments is very high. If the virtual driving environment fails to provide drivers a high realistic feeling, participants' behaviors may be not normal while using the simulator so as to result in inaccurate experimental conclusions (Fildes et al., 1997). Therefore, how to model geo-specific environments in driving simulators and the validity of driving behaviors in the virtual environments compose the most important bases for the driving simulator studies.

This paper presents a novel perspective of driving simulator application on traffic safety: exploring driving simulator as a traffic safety test tool at signalized

intersections. This study aimed at investigating if the driving simulator can be used as a valid tool to assess traffic safety at signalized intersection. In this study, a signalized intersection with as many important features (including roadway geometries, traffic control devices, intersection surroundings, and buildings) was replicated into a high-fidelity driving simulator. The specific intersection was selected because it was one of intersections with the highest crash frequencies in the Central Florida area. Then, a series of driving simulator experiments for the intersection were conducted to determine if the subjects' speed behaviors and traffic risk patterns in the driving simulator were similar to what were found at the real intersection (Yan et al., 2007). The validation method and results in this study would be applicable to other studies using similar driving simulator equipments.

2. Methodology

2.1 Apparatus/Equipment

The UCF (University of Central Florida) driving simulator was used in this research. The driving simulator has a motion base capable of operation with 6 degrees of freedom. It includes 5 channels (1 forward, 2 side views and 2 rear view mirrors) of image generation, an audio and vibration system, and steering wheel feedback. The simulated environment is projected at 180 degrees of field view and at a resolution of 1280×1024 pixels. The driving simulation system is composed of the following components:

- Simulator Cab: Saturn Sedan, automatic transmission, air conditioning, the left back mirror, and the back mirror inside the cab.
- Simview: The software provides the graphical display based on the computation.
- Scenario Editor: The software helps researchers to edit traffic scenarios.
- APIs for reading real-time data: APIs (Application Programmer Interface) can read the real-time data from Simview. The sampling frequency is 60Hz.

2.2 The selected Intersection

Alafaya Trail (SR434) and E. Colonial Drive (SR50) intersection located in Orange County was selected in this study. It is a major intersection (4X6) of state roads in Central Florida. Each approach at the intersection has two left turn lanes and one right turn lane. The intersection is skewed (81°) and the speed limits are 45 mph for southbound (434SB), eastbound (50EB), northbound approaches (434NB) and 50 mph for westbound approach (50WB).

2.3 Geospecific Database Modeling in the Simulator System

Replicating a real-world driving environment into a simulated 3-D virtual world is referred to as geo-specific database modeling. To provide drivers with a realistic driving experience, as many important features of the real intersection as possible were replicated in the driving simulator visual database. It consists of three essential components: the two intersecting roads (including traffic signs), the traffic lights, and the buildings (including important objectives close to the intersections, such as trees,

advertisement boards, electricity poles, and so on). The modeled road geometry is a near replica of the actual roadway network within 366 m (1200 ft) in each of the four approaches, achieved by merging roadway blueprints, autoCAD files, and hand measurements into the visual database. The traffic signs including speed limit signs were positioned in the visual database corresponding to the locations at the real intersection. The traffic lights were created for all signal cases which occurred at this particular intersection. Even the "walk/don't walk" flashing signs were built and integrated into the Traffic Control Device (TCD). The timing diagram and actuator sequencing from the actual intersection was used to build the TCD model. Since the study mainly paid attention to the isolated intersection safety problems but not road network issues, the signal coordination plan with other intersections was not considered in the simulation database. The further replication of the highway environment beyond the 366-m range of the intersection was extended to one to two miles for each approach, using a less precise, although more efficient method. Figure 2 illustrates the visual effect comparisons between virtual reality and the real intersection.

a. Snapshots for Chevron Gas Station & Shell Service Center

b. Snapshots at the southwest corner of the intersection

Figure 2: Geo-specific database modeling of the intersection

2.4 Speed Measurement in Field

For validating driver's speed behaviors in the simulator, free flow speeds at the real intersection were recorded for vehicles entering the intersection through each approach during the green phase, using a radar gun. These recordings for all the

approaches were taken on Tuesday dated 05/02/2006 from 9:30am to 5:00pm. Two observers were placed around 50 m (164 ft) downstream of the approach; the radar gun was pointed towards the opposing flow; speeds of the oncoming vehicles located upstream of the intersection were recorded. The vehicles were carefully selected such that they were under free flow conditions. Vehicles in queue were disregarded for data collection. There are 134 observations for 50WB approach, 104 observations for 50EB approach, and 91 observations for 434SB and 434NB each.

2.5 Intersection Crash History Analysis

Four years (1999-2002) of crashes at this intersection (SR434 & SR50) have been analyzed for the driving simulator's safety validation study. The crash information was obtained from the Florida Department of Transportation Crash Analysis Reporting System.

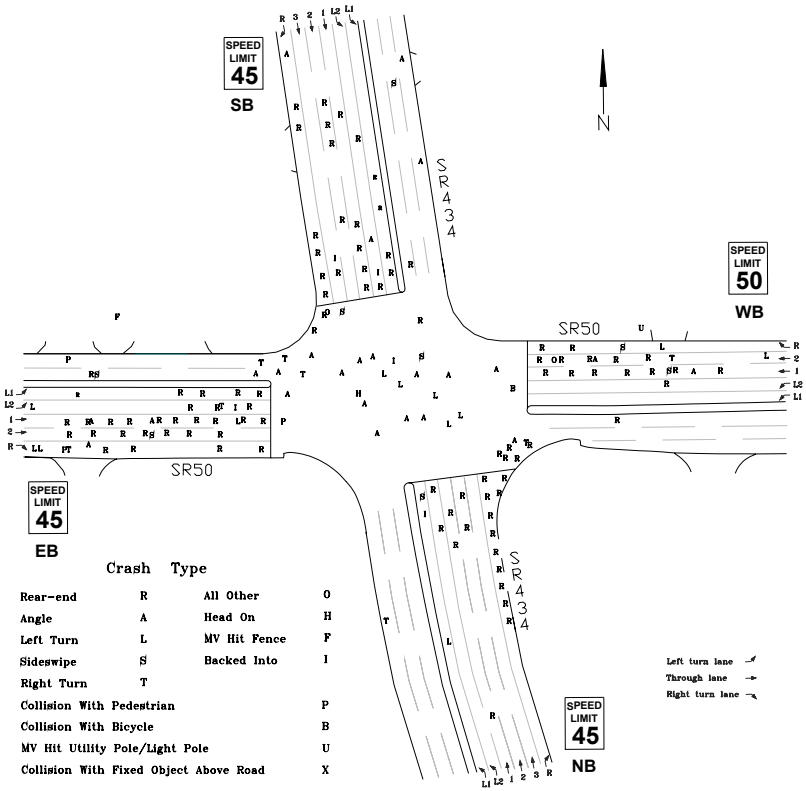


Figure 3: Crash Spot Diagram for Years 1999 through 2002

This intersection has experienced primarily rear-end crashes with a frequency of 95 and a relative frequency of 57.9%. The second most frequent type was angle (24 or 14.6%), followed by left turn (12 or 7.3%), sideswipe (10 or 6.1%), and right turn (8 or 4.9%). From the severity point of view there were 73 property-damage-only crashes, 90 injury crashes, and 1 fatal crash. Figure 3 shows all types of crashes that happened at this intersection for four years. In comparing the four approaches, there are two obvious trends for rear-end crashes. The first one is that the rear-end risk in the SR434 right-turn lanes is higher than that in SR50 right-turn lanes. According to the crash spot diagram, there is a total of 24 rear-end crashes occurring at SR434 right-turn lanes, 8 for 434SB and 16 for 434NB. During the same period, SR50 right-turn lanes have only 6 rear-end crashes, 4 for 50EB and 2 for 50WB. The second trend is that the rear-end risk in the SR50 through lanes is higher than that in the SR434 through lanes. The 50EB through lanes show the highest rear-end risk (17 crashes) while the 434NB through lanes show the lowest rear-end risk (6 crashes), compared to the other approaches. Another significant crash pattern at this intersection is that the through traffic in SR50 was involved in a higher angle crash rate (13 crashes) compared to that from SR434 (4 crashes). According to the crash report analysis, the typical contributing cause for the angle collisions at the intersection is "disregard traffic signal."

3. Driving Simulator Experiments and Result analyses

To assess if the driving simulator is a valid traffic safety tool at signalized intersections, this simulator validity study was conducted from two perspectives, a traffic parameter (speed) and a safety parameter (Crash history). For the speed validation, an experiment (#1) was design to investigate if the four-approach operation speeds are respectively measured in both simulator and field. For the safety validation, two experiments were designed to investigate if risk propensities observed in the further simulator experiment would be analogous to the traffic crash propensities found in crash history analysis. Specifically, one (experiment #2) is to verify if the rear-end crash risk at the 434NB right-turn lane is higher than that at the 50WB right-turn lane; the other (experiment #3) is to testify if the 50EB through lanes have higher rear-end risk and angle crash risk and than the 434NB through lanes.

Age and gender of the subjects are two independent variables (factors) considered for this experimental design. Since this study aimed at using the real traffic parameters and crash history at the intersection to validate the driving simulator experimental results, the age categorization followed the actual driver population using the quasi induced exposure method (Abdel-Aty et al., 1998; Stamatiadis and Deacon, 1997). Because very few crashes were involved by the older age group at the intersection, five age groups of interest are classified as Very Young (15 to 19), Young (20 to 24), Younger Middle-Aged (25 to 34), Middle Middle-Aged (35 to 44) and Older Middle-Aged (45+). Therefore, the experiment was a 5 (age) × 2 (gender) within-subject repeated measures design. Table 1 shows the sample size by age and gender for each specific scenario. Due to the sickness effect, the eventual number of subjects ranged from 58 to 62 in the specific scenarios. Note that the experiment #3 repeatedly utilized part of subjects in the experiment #1.

Table 1: Experiment Sample Size by Scenario

Age group	Experiment #1								Experiment #2							
					Experiment #3											
	50WB Through		434SB Through		50EB Through		434NB Through		434NB R-turn leading		50WB R-turn leading		434NB R-turn following		50WB R-turn following	
	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M
16 – 19	5	8	6	7	5	6	4	6	5	8	5	8	5	6	4	6
20 – 24	6	8	6	8	8	8	7	8	7	8	6	8	6	8	7	8
25 – 34	5	9	5	8	5	9	5	9	5	8	5	8	5	9	5	9
35 – 44	5	8	5	7	4	7	4	7	5	7	5	6	4	6	4	7
45 +	1	6	1	5	2	8	2	8	1	6	1	6	2	8	2	8
Subtotal	22	39	23	35	24	38	22	38	23	37	22	36	22	37	22	38
Total	61		58		62		60		60		58		59		60	

3.1 Experiment #1 – Speed Validation

3.1.1 Scenario design

There were four scenarios designed for the speed validation. Subjects started driving the simulator at 800 m upstream of the intersection and then cross through the intersections from the four intersection approaches, respectively.

3.1.2 Validation results

Table 2 shows the results of the Kolmogorov–Smirnov normality test of speed distribution. It was found that all speed distributions at four approaches for both field and simulator studies followed normal distributions at the 0.05 significance level.

Table 2: Kolmogorov – Smirnov Normality Test for Speed Distributions (km/h)

Approach	Statistical parameters	Field Study	Simulator Study
434NB	N	91	60
	Mean	70.46	70.32
	S.D.	10.19	13.67
	Kolmogorov-Smirnov Z	.808	.810
	P-value	.531	.528
434SB	N	91	58
	Mean	68.07	70.79
	S.D.	11.15	12.83
	Kolmogorov-Smirnov Z	.770	.304
	P-value	.594	1.000
50EB	N	104	62
	Mean	73.77	75.28
	S.D.	10.08	15.22
	Kolmogorov-Smirnov Z	.829	.633
	P-value	.497	.817
50WB	N	134	61
	Mean	72.57	76.59
	S.D.	10.13	14.31
	Kolmogorov-Smirnov Z	1.091	.619
	P-value	.185	.838

Table 3 shows the results of the F-test and the t-test for variance and mean comparisons, respectively. According to the F-test results, the speed variances in driving simulator were not significant different from those in field at the approaches of 434NB (p=.120) and 434SB (p=.219). However, the speed variances in driving simulator were statistically larger than those in field at the approaches of 50EB (p=.001) and 50WB (p=.004). Based on the variance type (equal or unequal), respective t–test statistic values were looked upon for mean comparison at each approach. From Table 3, there were no significant differences in mean speed between simulator and field at approaches 434NB (p=.940), 434SB (p=.173), and 50EB (p=.491). But the operation speed at the approach of 50WB was marginally faster in driving simulator than in field (p=.051).

Table 3: F-test and T-Test for Speed (km/h)

APPROACH	Levene's Test for Equality of Variances			t-test for Equality of Means						
	Variance	F	P	t	df	P	Mean Diff.	S.E.	95% C.I. of Diff.	
									Lower	Upper
434NB	Equal	2.446	.120	.076	149	.940	0.15	1.95	-3.70	3.99
	Unequal			.071	101.285	.943	0.15	2.06	-3.94	4.24
434SB	Equal	1.526	.219	-1.368	147	.173	-2.72	1.99	-6.65	1.21
	Unequal			-1.326	109.089	.188	-2.72	2.05	-6.78	1.35
50EB	Equal	12.367	.001	-.764	164	.446	-1.50	1.97	-5.38	2.38
	Unequal			-.692	93.339	.491	-1.50	2.17	-5.81	2.81
50WB	Equal	8.470	.004	-2.248	193	.026	-4.02	1.79	-7.55	-0.49
	Unequal			-1.982	88.396	.051	-4.02	2.03	-8.06	0.01

3.2 Experiment #2 – Safety Validation for Rear-End Risk at Right-Turn Lanes

3.2.1 Scenario design

For the safety validation of the rear-end crash, the right-turn movements at the 434NB approach were designed as test scenarios (the higher risk location) and those at the 50WB approach were designed as base scenarios (the lower risk location). Since the rear-end risk is related to both leading vehicles (e.g. sudden stopping) and following vehicles (e.g. following too closely), right-turn maneuvers as leading role and following role were both designed for the rear-end safety test. From the police crash report narrative, drivers' responses to signal change played a key role in most rear-end crashes. Therefore, the signal change was designed as a critical traffic event in the safety validation scenarios. As a leading role, there was no other vehicle in front of the simulator; when subjects were entering the right-turn lane and located at 100 m (328 ft) upstream from the stop line, the signal light changed from green to yellow to red. As a following role, subjects would follow a vehicle to enter the right turn lanes;

after the leading vehicle encroached into the right-turn lane, it gradually slowed down; and when it was located 60 m (197 ft) upstream from the stop line, the leading vehicle would make a sudden brake from 48.3 km/h (30 mph) to zero with a high deceleration rate 6.4 m/s^2 (21.06 ft/s^2 or 0.65 g); then, the subject driving the following simulator had to respond to the event quickly in order to avoid a rear-end crash happening.

3.2.2 Validation results -- as a leading vehicle

When subjects drove the simulator as a leading vehicle at the right-turn lanes of the 434NB and 50WB approaches, the independent variables were measured to evaluate the rear-end risk, including the simulator's average deceleration rate after the driver started braking, average speed distribution along right-turn lanes, and non-fully-stop rate at the stop line.

It was observed that average deceleration rate was higher for the 434NB approach (M=1.80 m/sec^2; SD=1.20 m/sec^2) than that for the 50WB approach (M=0.80 m/sec^2; SD=0.65 m/sec^2) [Equal variance; t(111)=5.563; p<.001]. Generally, a higher deceleration rate is more likely to lead to a rear-end crash. It was also observed that the mean speeds were consistently higher along the 50WB right turn lane than that along 434NB at locations 100m, 80m, 60m and 20m since the speed limit for 50WB (80.5 km/h) is higher than 434NB (72.4 km/h). However at the stop line, the mean speed at 434NB (M=18.38 km/h; SD=8.90 km/h) was higher than that at 50WB (M=14.68 km/h; SD=6.04 km/h). This difference was statistically significant based on the t-test [Unequal variance; t(104)=2.649; p=.009]. It means that when drivers make right turns in a situation such as pedestrians crossing the intersection, it is more likely to lead to an abrupt stop at 434NB than that at 50WB, to avoid collision with pedestrians.

Moreover, it was found that overall 69.75 % of the subjects did not stop at the stop line in the experiment, which displayed the general careless driving behavior of the subjects when they made right turns during the signal change. It was found that the non-stop rate for 434NB (81.67%) was higher than that for 50WB (57.63%), which was statistically significant based on the Chi-square test [$\chi^2(1,117) = 8.148$, p=.004]. For the 434NB right turn lane, the distance between the stop line and the edge of the SR50 was relatively smaller (see Figure 1). Therefore, it requires less time to make a right turn and drivers tend to quickly watch the traffic from other approaches and then turn without stopping. In contrast, the distance between the stop line and the edge of the SR434 is relatively larger. Hence, it requires longer time to make a right turn so that drivers tend to drive slowly or stop at this area. The drivers who did not stop fully at the stop line could make a sudden stop in emergency situations, such as yielding the right of way for pedestrians or conflicting traffic from the other approaches, so as to increase the risk of rear end collision with the following vehicles.

By taking average deceleration rate, non-fully-stop rate, and right-turn speed as surrogate measures for the rear-end risk, it can be concluded that the driving simulator experiment reflected that drivers' right-turn behaviors were more risky at the 434NB

approach than at the 50WB approach. Therefore, the driving behavior comparison between the two locations in the simulator is consistent with the crash history analysis.

3.2.3 Validation results -- as a following vehicle

When subjects drove the simulator as a following vehicle to make right turns, the leading vehicle's sudden brake may lead to the rear-end crash occurrence. It was observed that there were more rear-end crashes that occurred in the 434NB right turn lane than in the 50WB right turn lane (9 vs. 2). Based on the Z-proportion test, the rear-end probability (9/59=0.153) at the 434NB right turn lane was significantly higher than that (2/60=0.033) at the 50WB right turn lane (Z=2.24; p=.025). The results showed that the following speed at the 50WB right turn lane (M=46.13 km/h; SD=3.35 km/h) was slightly higher than that at the 434NB right turn lane (M=44.41 km/h; SD=4.39 km/h) [Equal variance; t (1,117)=2.411; p=0.017]. This result can be explained by the fact that the speed limit for the 50WB approach is higher than the 434NB approach. Furthermore, It was found that the average following distance at the 434NB right turn lane (M=30.19 m; SD=13.43 m) was significantly shorter than that at the 50WB right turn lane (M=35.58 m; SD=13.41 m) [Equal variance; t (1,117)= 2.191; p=0.030]. For the average deceleration rate, there is no significant difference found between the 434NB and 50WB right-turn lanes [Equal variance; t (1,117)=0.693; p=0.490].

For the intersection geometry design, the length of right-turn bay in 50WB (200 m) is considerably larger than that in 434NB (60 m). Such a long right-turn lane in 50WB may help driver prepare a turn maneuver much earlier than in 434NB, which explains why subjects kept relatively larger distances from the leading vehicles. Generally, a larger following distance can provide a driver enough reaction time to recognize a hazardous situation and make a stop decision. The shorter following distance in the 434NB right turn lane directly contributes to the higher probability of rear-end occurrence observed in the experiment. Using the driving simulator as a following vehicle, the experiment results displayed the similar rear-end crash trend at the intersection compared to the crash history analysis.

3.3 Experiment #3 – safety validation for rear-end and angle crash risk at through lanes

3.3.1 Scenario Design

For the safety validation related to through lanes, the main concern in the scenario design is the driver's stop/go decision at the onset of the yellow phase because wrong stop/go judgments may result in traffic crashes such as red-light running (angle crashes) or rear-end crashes. Upstream of signalized intersections, generally there is a length of roadway in advance of the intersection wherein drivers may be indecisive and respond differently to the onset of the yellow signal indication. This region of roadway is commonly referred to as the "dilemma zone". When vehicles are located in the dilemma zone, drivers who decide to proceed through the intersection at the onset of yellow may run a red light and potentially result in right-angle collisions. On the other hand, some drivers may stop abruptly while following

drivers may decide not to stop, which may contribute to the risk of a rear-end crash (Pline, 1999). The crash report analysis showed that both rear-end and angle crash rates in the 50EB through lanes are higher than those in the 434NB through lanes. Therefore, in scenarios AEBR and BNBR, subjects drive the simulator to go through the intersection and the phase will change from green to amber to red when subjects are approaching the intersection at 90 m (295 ft) away from the stop line. The subjects' driving response to the signal change in the base scenario BNBR was compared to that in the test scenario AEBR to find if the trends associated with rear-end and angle crash risks in the driving simulator were similar to those found in the crash report analysis.

3.3.2 Validation results

Results show the proportions of subjects' stop/go decisions at the onset of the yellow phase related to independent factors viz., intersection approach, driver gender, and driver age. At the onset of the yellow phase, drivers at the 50EB approach are more likely to cross the intersection compared to those drivers at the 434NB approach (37.1% vs. 13.3%). The Chi-square test showed that the p-value was 0.003 ($\chi_{1,122}^2 = 9.085$) and the drivers' stop/cross was statistically dependent on the two approaches at the 0.05 significance level. There was also a significant dependence of drivers' stop/cross decision on gender ($\chi_{1,122}^2 = 5.958$, P = 0.015). It appears that the male drivers were more likely to cross the intersection compared to the female drivers (32.9% vs. 13.0%). However, there was no statistically significant dependence of stop/go decision on driver age ($\chi_{4,122}^2 = 1.866$, P = 0.761).

Table 2 shows that the mean speed at the 50EB approach (73.8 km/h in the field study and 75.3 km/h in simulator) was higher than that at the 434NB approach (70.5 km/h in field study and 70.3 km/h in simulator), although the speed limits for both approaches were 72.4 km/h (45 mph). Note that the speed limit design for the SR50 is unbalanced and the speed limit for the 50WB approach is 50 mph. Moreover, speed limits in the field for most segments of the 50 highway are 80.5 km/h (50 mph), which may cause drivers to not fully reduce their traveling speeds to 72.4 km/h. Therefore, drivers at the 50EB approach are more likely to speed, which explains that drivers at the 50EB approach are less likely to stop at the intersection during signal change. Generally, when drivers encounter a yellow signal at 90 m away from the stop line of the intersection, they are more likely to fall into the dilemma and possibly to run a red light. Using crossing rate as a surrogate measure for angle collisions, it can be concluded that the vehicles from the 50EB approach were more likely to run a red light so as to result in a higher angle collision rate compared to the 434NB approach. Furthermore, according to the experiment results, there were 1 red light running observation of 60 subjects at the 434NB approach and 3 observations of 62 subjects at the 50EB approach. Since red-light running is a rare event, no conclusion could directly be drawn based on such a small sample size. On the other hand, the no-stop rate also can be considered as a surrogate measure for the rear-end collisions because within 90 m upstream of the stop line of the intersection, there is a potential conflict between the stopping drivers and crossing drivers during signal change. Therefore, a higher crossing rate at the 50EB through lanes may lead to more rear-end crashes

compared to the 434NB approach and thus this result also is also consistent with the crash history analysis.

4. Conclusions

Through carefully constructing geo-specific database for a signalized intersection in the driving simulator, this simulator validity study focused on investigating if the driving simulator can be developed as a test tool to assess traffic safety at signalized intersections. The validation research consisted of two aspects: speed measures and safety surrogate measures. Comparing speed data observed from the field to those in the simulator experiment, the study showed that both follow normal distributions and have equal means for each intersection approach at the 0.05 significance level, which validated the driving simulator in terms of absolute validation. Additionally, it was found that the speed variances were equal for the lower operation speed locations but unequal for the two locations with higher operation speed; there was a trend that the speed measured in the driving simulator showed a larger variability than the field. Based on overall comparisons of speed between simulation and real world, one can conclude that the driving simulator is a proper tool for traffic studies related to driving speed behaviors.

Furthermore, the driving simulator experiment can reflect the crash history trend of the intersection: drivers showed relatively higher rear-end and angle crash risks at locations with a high crash history record and lower risks at locations with a low crash history record. Driving simulators provide a safe and controlled virtual environment within which it is possible to reproduce dangerous driving conditions and situations in order to test driver behaviors. The validation results of this study suggested that through modeling the real intersection geometries and roadway environments in the simulator's 3-D virtual reality, driving simulator experiments may be applied to identify potential design problems, explain interaction between drivers and roadway surroundings, and most importantly, explore effective countermeasures to enhance the quality of traffic safety and operation.

References:

Abdel-Aty, M., Chen, C., and Schott, J. (1998). An Assessment of the Effect of Driver Age on Traffic Accident Involvement Using Log-Linear Models, *Accident Analysis & Prevention Journal*, Vol. 30, pp. 851-861.

Fildes, B., S. Godley, T. Triggs and J. Jarvis (1997). *Perceptual Countermeasures: Simulator Validation Study.* Department of Transport and Regional Development, The Federal Office of Road Safety, Roads and Traffic Authority of New South Wales, CR 169, Commonwealth of Australia.

Orit, S., Sherrilene, C., Burton, S., Ethan D., Roxanna, B., Patricia, B., Milapt, S., Dennis, M., and William, M. (2006). The Impact of Intersection Design on Simulated Driving Performance of Young and Senior Adults: Preliminary Results. *Topics in Geriatric Rehabilitation*. Vol. 22, pp. 27-35.

Pline, J. L., ed. (1999). Traffic Engineering Handbook, 5th ed. *Institute of Transportation Engineers*, Washington, D.C..

Stamatiadis, N. and Deacon, J. (1997). Quasi-Induced Exposure: Methodology and Insight, *Accident Analysis & Prevention* Vol 29, No. 1, pp 37-52.

Yan, X., Aty, Abdel-Aty, M., Radwan, E., Wang, X., and Chilakapati, P. (2007). Two Aspects of Driving Simulator Validation: Speed and Safety. Accepted by the Proceeding of TRB 85th Annual Meeting, Washington, D.C., January 2007.

IMPACT OF ETC ON TRAFFIC SAFETY AT TOLL PLAZA

Jiping DING, Fan YE and Jian LU

Transportation College, Southeast University, Nanjing, Jiangsu, 2 Si-Pai Lou, 210096, China; PH(86)025-8390776; FAX(86)025-83795352; email: dingjiping@hotmail.com

Abstract: Some researches and reports has indicated that ETC system does contribute to traffic accidents at toll plazas. Oppositely, others showed that ETC system does improve the traffic safety at toll plazas. There are still no definite conclusions on safety impact. For establishing a basis of evaluating the benefit of ETC, the impact of ETC on traffic safety at toll plaza was studied in the paper. Firstly, related literatures were reviewed. Then the safety conditions at a toll plaza and the contributing factors of ETC to safety are discussed. Based on traffic crash data collected, it was indicated that ETC may result in the increase of crash number and more severe crashes could happen due to the high passing speed at ETC gates.

Key words: *electronic toll collection (ETC), traffic safety, ITS, crash, evaluation*

1 Introduction

ETC is a system that automatically identifies a vehicle equipped with a valid encoded data tag or transponder as it moves through a toll lane and then posts a debit to the patron's account without requiring the vehicle to stop for the transaction. In China, with the increasing transportation needs and escalating constructions costs, transportation agencies are becoming more demanding in their pursuit to meet transportation needs through ETC and other ITS technologies.

The significant benefits of the ETC is improving throughput, decreasing emission, and reducing operational cost of toll plaza. So far, most ETC research studies have focused on the subject related to traffic operational benefits and little research has been done to evaluate the impact of an ETC on the safety performance. Virtually, in addition to the traffic performance of ETC, there is an urgent need for studying traffic safety issues of ETC in order to provide warrants for policy decision making, investing, transportation system design, and construction implementation, etc.

2 Literature Review

Some studies suggest that ETC systems can improve safety. For example, Oklahoma Turnpike Authority reported that there were no accidents in its PIKEPASS lane in the first year after the installation of the ETC system, as compared to 71 accidents in its regular lanes (Ort, 1995).

New York State Thruway Authority collected Accident data from the Thruway from 1992 through 1998 to analyze the impact of ETC implementation on traffic safety. The result show that with the increasing of ETC market share, the ETC related accidents increased, but this increase at all locations was lower than the increase in ETC use. The research also shows: at all study locations, the total number of accidents as a proportion of total transactions either decreased appreciably or remained the about the same. While this can not be attributed to ETC necessarily, it is fair to say that ETC did not result in an increase in total accidents.

Mohamed, A. A., M. Abdel-Aty, and J. G. Klodzinsks found that the average monthly crash rate at mainline toll plazas jumped to 7.5 crashes/month after the installation of the E-PASS system from 3.375 crashes/month before introducing the ETC system, ten percent of the total crashes were related to the ETC system because they either involved an E-Pass vehicle or took place in a dedicated ETC lane.

In 2002, a report about ITS benefits was published. In this report, it is pointed out: In Florida, driver uncertainty about congestion at ETC toll stations contributed to a 48% increase in accidents.

Summarize past research results and the following three different or contradictory conclusions were obtained:

(1) With the use of ETC at toll plaza, there is generally an improvement in safety that can be measured by reduction in number of crashes and quantified in terms of property, injury and life savings.

(2) After the introduction of ETC, with increasing ETC market share, the percent of total crashes that were ETC related increased. But, this increase was lower than the increase in ETC market share. The total number of crashes as a proportion of total transactions either decreased appreciably or remained about the same, which cannot be attributed to ETC necessarily. Thus, it is fair to say that ETC did not result in an increase in total crashes

(3) The implementation of ETC increased the potential of all crash types and crash severity due to an increase in speed variance and driver uncertainty about lane configuration so the drivers are paying price for reduced highway safety.

The contradictory findings from previous research studies regarding the safety impact of ETC make it difficult to give objective evaluation conclusions. Therefore , further investigation into the origins must be made to explain why they happened and what should be done to rectify the situation.

3 Safety Conditions at Toll Plaza

Based on toll plaza traffic accident reports of the Central Florida expressway system from 1999 to 2000, it can be found that traffic accidents are more like to occur at toll plaza, and Figure 1 presents the distribution of crashes at different locations. According to the reports, distribution of crash types at toll plaza can also be identified

and presented in Figure 2. From Figure 2, it can be found that rear end crashes are the most common crash types at toll plazas due to frequent stops of vehicles.

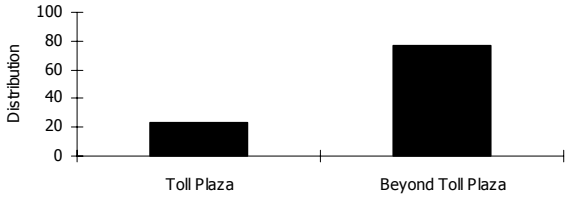


Figure 1. Location Distribution of Highway Crashes

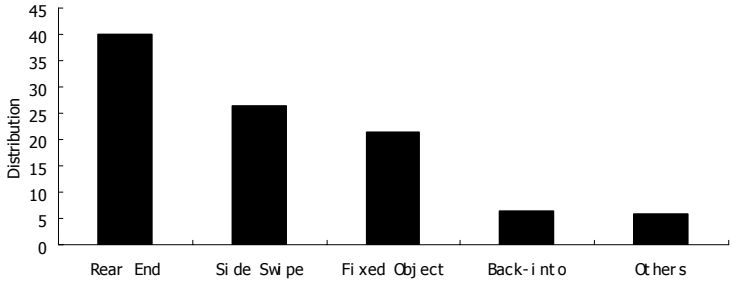


Figure 2. Distribution of Crash Types at Toll Plaza

There are six common accident types around toll plaza: sideswipe crash, rear-end crash, hit fixed object, back-into and pedestrian related crash(others). In general, crashes at toll plazas are more likely to occur due to at least two reasons: some drivers may reduce their speeds suddenly with more potential for rear-ended crashes, and there may be frequent lane changes of vehicles, resulting more potential for side swipe crashes. The main reasons for frequent lane changes may be contributed by drivers looking for an unoccupied toll lane or minimum queue length and selecting a specific lane type such as ETC, ACM, or manual lane.

4 Discussion on ETC Related Accidents

Reasons that contribute to traffic crashes at toll plazas are discussed as follows: First of all, sideswipe crashes are usually caused by merging movement. Drivers perform merging movements when approaching to the toll plaza and leaving from the toll plaza. This is mainly due to the reason that a driver must decide which lane to enter to a toll plaza within a relatively short time period.

For a toll collection station with ETC function, drivers have the opportunity to select ETC lanes or manual lanes. This could cause much more complicated driving maneuvers and traffic conflicts, resulting more crash potential.

Rear-end crashes most frequently occur at toll lanes with queuing conditions during peak hours. When vehicles approach a toll plaza and stop to pay tolls, vehicle queuing occurs at the toll plaza and extensive queues may form during peak hours.

Other possible factor that contributes to rear-end crashes may include the fact that when vehicles approach to the toll plaza, drivers need to determine which lanes to enter and possibly pay less attention to the vehicles stopping suddenly in a ETC lane, resulting rear end traffic conflicts.

Lack of driver's knowledge about ETC lanes may also result in the rear-end conflicts. For the vehicles without ETC capability may mistakenly enter an ETC lane and then realize that the lane does not accept conventional payment methods. Vehicles that stop in the ETC lane while trying to figure out how to pay the required toll may create the potential for a severe rear-end collision with a vehicles with ETC capability. Usually, when a vehicle with ETC capability is moving to an ETC lane, the drive does not expect to stop at the ETC lane.

Hit-plaza crashes are usually caused by vehicles that are moving at high speeds. Before the implementation of ETC systems at a toll station, speed variance of vehicles are usually low due to the fact that all vehicles had to stop at the toll plaza. However, with installation of the ETC system, because ETC users do not have to stop and usually do not slow down much either, the speed variance may increase dramatically, resulting more potential for traffic conflicts and vehicle crashes. Meanwhile, the crash severity also gets worse if speed variance gets larger.

Hit-plaza's barrier is the main proportion of hit-plaza crashes. When an ETC user exceeds the speed limits of ETC lane (The speed limit in China which in is 30 km per hour) or follows the front vehicle too closely, a barrier related crash could occur because the barrier cannot lift in time.

Most pedestrian related crashes are related to ETC system. With the ETC capability, ETC vehicles may increase the hazard for any person crossing a plaza lane. Pedestrians at toll plazas are either employees at the toll plaza or drivers who exit their vehicles for any reason.

Based on above discussions, it could be concluded that the main causes of ETC related crashes is the drivers' unfamiliarity with the ETC system and the plaza configuration. Also, increasing speed variance may be a contributing factor for the severity of crashes. In the other word, we still cannot contribute the decrease of safety benefits to ETC system itself because we do not know what will be after increase drivers' familiarity and optimize the plaza configuration.

5 Field Data Collection and Analysis

To further evaluate whether ETC systems have impact on traffic safety of toll collection stations, field data collection was performed in this research. One of the best application locations of ETC systems in China is the Xiamen Bridge. Thus, for the purpose of safety evaluation of ETC systems, Xiamen Bridge was used for field data collection. Crash data at the toll plaza of Xiamen Bridge were collected. The toll plaza has one ETC lane and six manual toll collection (MTC) lanes and its layout is presented in Figure 3.

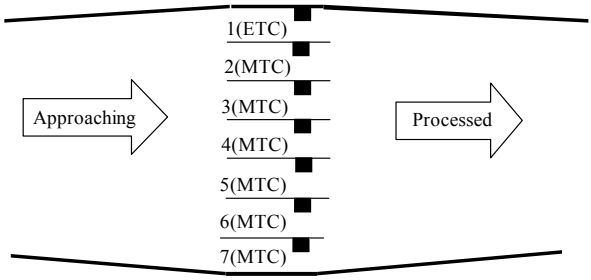


Figure 3. The layout for the toll plaza of Xiamen Bridge

In this research, we divide crashes at the toll plaza in two main categories: hit-barrier crashes and hit-structure crashes. In order to have fair comparisons between ETC lane and MTC lanes, crash data in last thirty four months were summarized and the unit of annual crashes per lane was used in the analysis. Figure 4 presents the summarized crash data for ETC and MTC lanes. It is indicated from the figure that there were more hit-barrier crashes involving ETC users as compared to MTC users. However, less hit-structure crashes involving ETC users were found as compared to MTC users.

Based on the original crash data patterns for both ETC and MTC cases, it was found that the main reason contributing to hit-barrier and hit-structure crashes was speeding of vehicles. Since the crash data did not show significant decreasing trend although drivers got more familiar with ETC lane and its configurations, it is hard to conclude whether the factor of drivers' familiarity had significant impacts on ETC safety performance. Definitely, more field data from more sites are needed to make such conclusions regarding the contributing factors to ETC safety impacts.

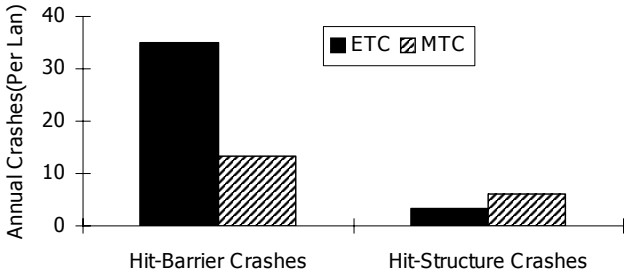


Figure 4. Annual Crashes (Per Lane) of ETC and MTC at Xiamen Bridge in China

6 Conclusion

As discussed in the paper, use of ETC may result in not only the increase in total crash number or crash rate but also the number of personal injuries and the collision severity at toll plazas. The main causes of ETC related crashes could contribute to the factors such as drivers' unfamiliarity with new ETC systems plaza configurations and the increased speed variance. However, there are no sufficient evidences to make conclusions regarding the safety impacts of ETC. Further research studies should be carried on to obtain more field data so that the safety performance of ETC systems could be better understood.

References

Abdelwahab, H. T. and M. A. Abdel-Aty, "Traffic Safety Analysis for Toll Plazas Using Artificial Neural Networks and Logit Models", The 80th Transportation Research Board Annual Meeting, National Research Council, Washington, D.C., 2001.

Gillen, D, J. Li, J. Dahlgren, and E. Chang, Assessing the Benefits and Costs of ITS Projects: Volume 2 An Application to Electronic Toll Collection, California PATH Research Report UCB-ITS-PRR-99-10, March, 1999.

Mitretek Systems, ITS Benefits: Data Needs 2002 Update, Washington,DC, 2002

Mohamed, A. A., M. Abdel-Aty, and J. G. Klodzinski, "Safety Considerations in Designing Electronic Toll Plazas: Case Study", ITE Journal, March 2001, pp. 20 – 24.

Menta, V. K., H. E. Strate, D. A. Boss, and A. J. Saracena, "Electronic Toll Collection and Safety at the Holland Tunnel", Proceedings, Conference on Traffic Congestion and Traffic Safety in the 21st Century, 1997, pp. 236 – 242.

Vollmer Associates LLP, E-Z Pass Evaluation Report, Prepared for：New York State Thruway Authority, August, 2000.

Piecewise Optimum Delay Estimation for Intersection Control

Chun Shao[1] and Ping Yi[2]

[1] Ph.D. Candidate, Department of Civil Engineering, University of Akron, Akron, OH, 44325 Email: cs57@uakron.edu
[2] Corresponding Author, Associate Professor, Department of Civil Engineering, University of Akron, Akron, OH 44325 Email: pyi@uakron.edu

Abstract
An adaptive control strategy called Piecewise Optimum Delay Estimation (PODE) has been developed and tested in this paper. The limitations in the existing adaptive control strategies are reviewed and used as the basis for improvement. By using real time field data for traffic demand estimation and through piecewise optimization, PODE dynamically searches movement combinations for phasing and timing decisions that minimize system wide delay. Preliminary results show that PODE is able to considerably reduce delay over actuated control. Its performance consistency, especially in congested and oversaturated conditions, has demonstrated PODE's competitive strength among other adaptive control strategies.

Introduction

Improvements in traffic signal control have been characterized by the innovative design of control devices and the development and testing of control logic in the last century. Mechanical control devices have largely been replaced by electronic systems and the actuated control logic is widely used today in addition to the fixed time control. Recent progress of microchip processor and computer networking technology makes the implementation of adaptive control strategy practical. Among the adaptive control strategies, Optimized Policies for Adaptive Control (OPAC) (1) and Real-time, Hierarchical, Optimized, Distributed, and Effective System (RHODES) (2) are widely published in which the optimization makes use of real-time arrival data at the intersection. OPAC was first presented by N. H. Gartner in 1983 (1), and it uses a simplified Dynamic Programming (DP) method and Rolling Horizon approach to optimize signal control for isolated intersection. OPAC was later expanded to include a coordination/synchronization strategy suitable for arterials and networks control (3). Similarly, RHODES is a hierarchical strategy introduced by Head and Mirchandani in 1992 (2). RHODES has different algorithms for intersection control and network control. At the intersection control level, RHODES applies a procedure called Controlled Optimization of Phases (COP) (4) which optimizes phasing and timing by using a DP method. For network control, RHODES uses the REALBAND model (5) based on platoon flow prediction.

According to previous studies (6), the reduction of vehicle delay in OPAC varies from 3.9% to 15.94% compared with the actuated control logic in three field tests. The software simulation performed for RHODES (7) shows that the average vehicle delay decreases in the range of 50% for low traffic volumes to 30% for high volumes against actuated control.

However, current adaptive control logics have their limitations. Those include uncertain traffic flow prediction and difficulty in estimating arrival time. Motivated by the need for improvement, we developed a novel adaptive control logic called Piecewise Optimum Delay Estimation (PODE) to address the above limitations. PODE is a real-time data dependent, flexible optimum interval, vehicle behavior estimation based, and auto-adjusted adaptive control logic.

The following sections will first discuss the limitations in current practice of adaptive control and areas for improvement. Then, the PODE control logic is introduced. Three main functions used in PODE, initial queue size, vehicle arrivals and releases, are explained in detail. Preliminary testing results show that PODE is able to notably and consistently reduce vehicle delay against actuated control. Improvements over other control logics have also been found in the tests.

LIMITATIONS IN EXISTING ADAPTIVE CONTROL STRATEGIES

Uncertainty in predicting future traffic flows

Adaptive control strategies rely mainly on the prediction of traffic flow information (8). The longer the prediction horizon, the less accurate it will be. Though long-term optimization is ideal for reaching the global optimums, the real-time data based adaptive control strategy relies on a number of short-term optimizations to reduce uncertainty in traffic demand estimation and improve accuracy in the computation. For example, the OPAC strategy relies on data from the past 50 to 100 seconds (9), so its effectiveness is largely dependent on the accuracy of its estimation based on historical data. No matter how the estimation model is built, there will always be differences between the prediction and the field condition. Hence, a desirable adaptive control strategy should reduce reliance on prediction, especially the one from historical data, as much as possible.

Difficulty in estimation of arrival time

A reliable estimation model must be developed to provide real-time traffic information for adaptive control. Vehicle arrival information is usually obtained from detectors placed upstream of the intersection, and the objective of estimation is to obtain the vehicle travel time between the upstream detector and the intersection stopline (1, 9). The geometric layout of the travel time estimation is shown in Figure 1.

RHODES (10) split travel time into two parts, the time between upstream detector d_U and detector d_A, and the time from d_A to intersection A. Detector d_A is located several hundred feet upstream of intersection A in order to provide long enough "reaction time" for the system to adjust the signals. In most conditions, however, the travel time from d_A to intersection A cannot be easily determined because it is largely affected by the existing queue length and the signal status at intersection A. Give the following notations:

- i – Index for vehicles arriving at intersection A
- μ_i – Estimated arrival time of vehicle i to intersection A
- δ_i – Estimated arrival time of vehicle i at detector d_A
- τ_i – Estimated travel time between d_A and intersection A

We can obtain this relationship from the above notations:

$$\mu_i = \delta_i + \tau_i \tag{1}$$

In equation 1, δ_i is affected by the detected speed and distance between d_U and d_A. Head (10) stated that δ_i can be predicted with acceptable errors. However, the travel time τ_i is not merely to divide the distance between d_A and intersection A by the approaching speed of the vehicle. Different queue sizes remaining in the approach will change τ_i. For instance, as depicted in Figure 2, there are two cases which have the approaching vehicle with same speed. Due to the different queue size of each case, the travel time τ_i is different and this leads to different values of μ_i, the estimated vehicle arrival time to the intersection.

PODE OPTIMIZATION LOGIC

Piecewise Optimum Delay Estimation (PODE) is a piecewise optimizing strategy that optimizes signal operations in short-term intervals based on field detection. In piecewise optimization, all the possible movement combinations will be treated as candidate phases and each interval can only have one phase. The preferred phase and the length of interval are dynamically determined in PODE by the algorithm according to real-time traffic conditions. Upstream detectors in each intersection approach are used to record vehicles demand information in real time. The detected vehicle's speed and position will be estimated and tracked until it is released from the intersection.

Initial queue size

$\lambda_{j,0}$ is the initial queue size at the beginning of optimization interval j. It is also the final queue size at the end of last optimization interval which represented by $\lambda_{j-1,c_{j-1}}$.

When the system is first initialized, the queue size $\lambda_{0,0}$ is set to zero. As the optimization progresses, the queue size is tracked interval by interval.

Individual Vehicle Status Estimation

Before introducing the estimation of vehicle arrival $\alpha_{j,t}$ and release $\gamma_{j,t}$, we need to depict the individual vehicle status, which is essential to the arrival and release calculation. Every vehicle pass through the upstream detector d_U will be recorded in the system and assigned a unique ID. Its status, such as speed and position, will be estimated and tracked. The position of a vehicle is the relative distance to the intersection stop line. Once a vehicle is detected, the position value is initialized to the distance between the intersection and the upstream detector. The initial speed of vehicle is the one obtained from the detector. At any time t in optimization interval j, each vehicle has its own position and speed which can be represented by the following notation:

i - Vehicle index
$v_{i,j,t}$ - Speed of vehicle i at time t in interval j
$x_{i,j,t}$ - Position of vehicle i at time t in interval j
$v'_{i,j,t}$ - Acceleration/Deceleration rate of vehicle i at time t in interval j

From the definition, we can find that

$$v_{i,j,t} = v_{i,j,t-1} + v'_{i,j,t-1} \tag{2}$$

and

$$x_{i,j,t} = x_{i,j,t-1} - v_{i,j,t-1} \tag{3}$$

With equation 2 and 3, if we can find the vehicle acceleration/deceleration rate at any time, the status of vehicle can be determined by its initial speed and position. In PODE, we use the car-following model (11) to calculate $v'_{i,j,t}$. According to the space limitation, detail discussion will be depicted in other papers.

Vehicle arrivals and releases

Because of the influence of the remaining vehicles in the intersection (not cleared in the previous optimization interval), we estimate the arrival time at the queue tail instead of the stopline. This will not affect the optimization in low volume conditions because the queue tail is at the stopline if there are no waiting vehicles. From the discussion of previous section, we can obtain each vehicle's speed at any time. Since the system in PODE is looked as a pipeline, any vehicle is considered joining the queue when its speed is down to 0.

Vehicle releases $\gamma_{j,t}$ are dependent on the queue size $\lambda_{j,t}$, the arrivals $\alpha_{j,t}$, and the phase p_j. The basic vehicle releasing rate r can be measured from the field.

Collectively, the calculation of vehicle releases can be written as

$$\gamma_{j,t} = \begin{cases} 0 & \text{lane} \notin p_j \\ 0 & \text{lane} \in p_j \text{ but } t \in \text{start - up delay} \\ \max\left\{ r, \lambda_{j,t} + \alpha_{j,t} \right\} & \text{lane} \in p_j \text{ and } t \notin \text{start - up delay} \end{cases} \quad (4)$$

Optimization Process

Based on the above discussions, we can find the system delay D_j of the selected lane in optimization interval j. Since the objective of PODE algorithm is to find the minimum system wide delay, the measure of performance, M_j, is the system delay D_j divided by the optimization interval c_j. This is because of the flexibility of interval length that we need to compare system delay under the same condition. The objective function for M_j can be written as follows:

$$\min(M_j) = \min\left(\frac{D_j}{c_j} \right) \quad (5)$$

By varying p_j and c_j, the measure of performance M_j is calculated sequentially for all possible combinations. In each iteration, the current M_j is compared with the previously recorded minimum value which is replaced if the newly computed is smaller. At the end of the optimization, the set of p_j and c_j corresponding to the smallest value of M_j is the optimal solution.

The basic steps in PODE are as follows:

1. Collect and record vehicles' information
2. Adjust the initial queue $\lambda_{j,0}$ for selected lanes
3. Design possible combination of phase p_j and length c_j
4. Estimate the vehicle arrivals $\alpha_{j,t}$ with given p_j and c_j
5. Estimate the vehicle releases $\gamma_{j,t}$ with given p_j and c_j
6. Calculate the system delay D_j and measurement M_j
7. Compare with the current minimum system delay and replace with the smaller one if applicable
8. Repeat step 3-7 until all the feasible combinations of p_j and c_j have been calculated

SIMULATION AND EVALUATION

Software simulations are performed to compare PODE with different control logics. The platform of simulation is VISSIM, which is a microscopic, behavior-based multi-purpose traffic simulation program widely used in traffic analysis (12).With the help of the development package, we build the signal control logic with Visual C++ .NET to interface VISSIM.

The simulation is performed for a four legged isolated intersection. Each approach has two lanes for the through/right turn movements plus a left turn pocket. During simulation, we changed the traffic demand of the whole intersection from 1600 veh/hr to near 7000 veh/hr at the intervals of 400 veh/hr. The high volume represents congested situation because the theoretical and practical capacity is reported in the range of 6000 to 6400 veh/hr (13). Seventy simulation runs were made with different traffic load distributions on each approach.

We choose the actuated control logic as the basis for comparison. The setup for phasing and timing in actuated control follows the standard of National Electrical Manufacturers Association (NEMA). The parameters of actuated control are adjusted at different traffic demand levels and load distributions to get the best performance.

The simulation results are shown in Figure 5 based on seventy runs. We can see from this figure that PODE consistently outperforms the actuated control logic, and the vehicle delay from PODE is notably reduced in high volume situations. In the low to medium traffic demand range, from 1600 veh/hr to 4800 veh/hr, the saving in vehicle delay by PODE against the best performance of actuated control is within 10 seconds/veh. In high traffic demand, from 5200 veh/hr to 6400 veh/hr (actuated control cannot handle higher demand), actuated control begins to increasingly show large "jumps" in delay, which means its performance is unstable with different traffic load distributions. In comparison, although vehicle delay in PODE also goes up with the increase in the traffic demand, the change is much smoother and the incrementals are much smaller than those from actuated control. In addition, PODE can handle a larger volume (exceeding 6800 veh/hr) whereas actuated control breaks down (continuous queue growth) when volume reaches 6400 veh/hr. Another noticeable advantage of PODE is that there is no need to manually adjust the system parameters during the seventy simulation runs, whereas in the actuated control logic the location of the detectors and the maximum green have to be changed in order to obtain its best performance shown in Figure 4.

Table 1 shows the percentage of reduction in vehicle delay. The range of reduction varies from 15% to 50%. With the increase of traffic demand, the percentage of reduction also increases which indicates that the PODE works very well in heavy to oversaturated traffic situations.

Compared with other control logics, PODE has also showed that it is competitive. According to OPAC field tests (6) and simulation results (14), the reduction percentage is from 10% to 20% against actuated control. There is very little reported work of RHODES in isolated intersection control. Sen and Head (4) presented a chart to show the difference of delay between RHODES and actuated control for an isolated intersection. From this chart, the average delay from RHODES increased sharply from 15 sec/veh as the traffic demand increased from 375 veh/hr/lane. In the case of PODE, however, the delay is around 15 sec/veh to 20

sec/veh at the same level of traffic demand and geometric layout. The reductions are encouraging. It should be noted that the authors of this paper did not implement OPAC and RHODES. To do that would require a thorough understanding of the details of those models, which is beyond the scope of this paper. As the performance measures of those models used in the discussion are obtained from other people's work, cautions must be exercised should they be cited in future references. Nevertheless, we believe the competitive strength of PODE, especially in heavy and oversaturated traffic, has been shown in the above testing and evaluation process.

CONCLUSION AND FUTURE WORK

A novel adaptive control logic called PODE has been introduced and tested. The preliminary test performed in software simulation has shown that the algorithm in PODE is very effective for isolated intersection control. Based on our own work, PODE performs consistently better, and its reduction in vehicle delay has been found to reach nearly 50% in congested traffic conditions compared with actuated control. According to simulation results, PODE can further improve vehicle delay over other adaptive control strategies, such as OPAC and RHODES. The control logic used in PODE requires input from multiple detectors to obtain real-time data. With the help of advanced technologies in vehicle detection, such as video processing and computer networking, such a requirement should not become an obstacle to testing PODE in the field. Work is already underway to implement PODE in hardware at the University of Akron Traffic Lab, where PODE as a component model will be combined with the Bandwidth Optimization and Segmentation Analysis (BOSA) system to be applied to network wide traffic control.

Reference

1. Gartner, N.H. OPAC: A Demand-Responsive Strategy for Traffic Signal Control, *Transportation Research Record 906*, 1983
2. Head, K.L., Mirchandani, P.B. and Sheppard, D., Hierarchical Framework for Real-Time Traffic Control, *Transportation Research Record 1360*, 1992
3. Gartner, N.H., Pooran, F.J. and Andrews, C.M., Implementation of the OPAC Adaptive Control Strategy in a Traffic Signal Network, *2001 IEEE Intelligent Transportation Systems Conference Proceedings*, 2001
4. Sen, S. and Head, K.L., Controlled Optimization of Phases at an Intersection, *Transportation Science Vol. 31 No. 1*, 1997
5. Dell'Olmo, P. and Mirchandani, P.B., REALBAND: An Approach for Real-Time Coordination of Traffic Flows on Networks, *Transportation Research Record 1494*, 1995
6. Gartner, N.H., Tarnoff, P.J., and Andrews, C.M., Evaluation of Optimized Policies for Adaptive Control Strategy, *Transportation Research Record 1324*, 1991

7. Mirchandani, P. and Head, L., RHODES: A Real-Time Traffic Signal Control System: Architecture, Algorithms, and Analysis, *Transportation Research Record, Part B*, 2001
8. Klein, L.A., *Sensor Technologies and Data Requirements for ITS*, ISBN 158053077X, Artech House Publisher, Boston, London, Jan 1, 2001
9. U.S. Department of Transportation/Office of the Secretary of Transportation, *Demand-Responsive Decentralized Urban Traffic Control Part II: Network Extensions*, 1983
10. Head, K.L., Event-Based Short-Term Traffic Flow Prediction Model, *Transportation Research Record 1510*, 1995
11. Gerlough D.L and Huber M.J., Traffic Flow Theory, *Transportation Research Board Special Report 165*, 1975
12. VISSIM home page, PTV America, Inc., http://www.ptvamerica.com/vissim.html, accessed July, 2006
13. TRB, *Highway Capacity Manual*, National Research Council, Washington, D. C., 2000
14. Li, H.L., Zhang, L., and Gartner, N.H., Comparative Evaluation Of Three Adaptive Control Strategies: Opac, Tacos, And FLC, *Transportation Research Board*, 2006

TABLE 1 Delay Reduction in PODE Against Actuated Control

Traffic Demand (veh/hr)	Average Delay (sec/veh)		Reduction (seconds)	Reduction Percentage
	Actuated Control	PODE		
1600	13.94	10.19	3.75	26.91%
2000	14.73	11.52	3.21	21.78%
2400	16.30	12.68	3.62	22.23%
2800	18.51	14.07	4.44	23.98%
3200	21.78	16.21	5.57	25.58%
3600	24.85	18.15	6.70	26.98%
4000	27.14	20.36	6.78	24.97%
4400	30.45	22.41	8.04	26.41%
4800	33.19	25.25	7.93	23.91%
5200	39.29	27.89	11.39	29.00%
5600	48.03	32.48	15.55	32.37%
6000	61.14	37.60	23.54	38.50%
6400	88.78	49.47	39.31	44.28%
6800	N/A	68.09	N/A	N/A
Overall Average Reduction Percentage				28.05%

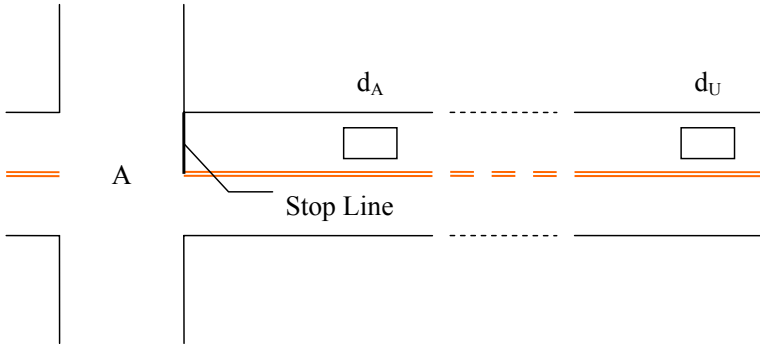
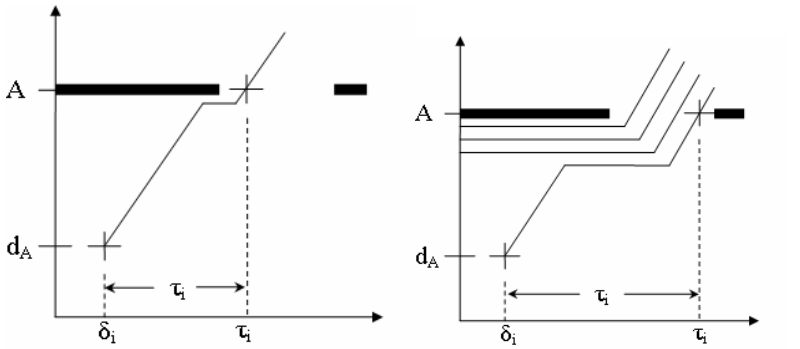


FIGURE 1 Geometric Layout for Vehicle Arrival Time Estimation



(a) Vehicle arrives without queue in the lane (b) Vehicle arrives with queue in the lane

FIGURE 2 Different Travel Time between d_A and A with Different Queue Sizes

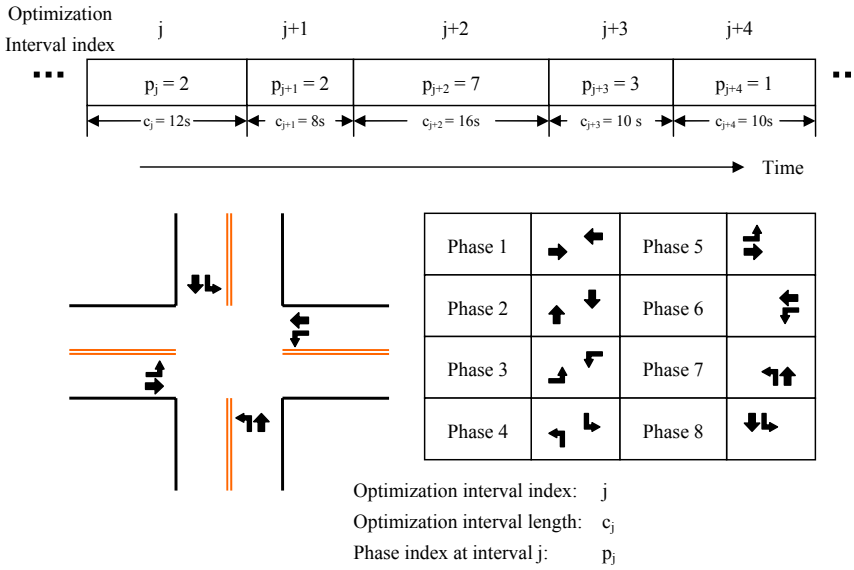


FIGURE 3 Sample Optimization Result

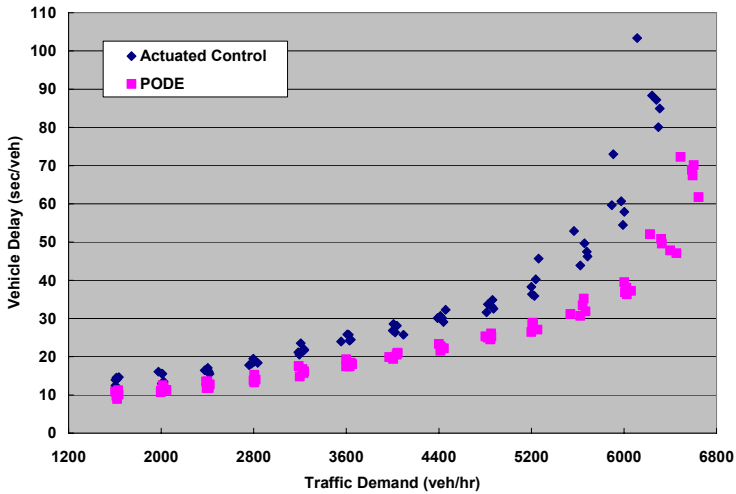


FIGURE 4 Vehicle Delay Comparison between Actuated Control and PODE

Performance Evaluation of Stone and RAP Interlayers in Soil-Cement Pavements under Accelerated Loading

Zhong Wu[1], Zhongjie Zhang[2] and Louay N. Mohammad[3]

[1] Louisiana Transportation Research Center, 4101 Gourrier Ave, Baton Rouge, LA 70808; PH (225) 767-9163; FAX (225) 767-9108; email: zhongwu@ltrc.lsu.edu
[2] Louisiana Transportation Research Center, 4101 Gourrier Ave, Baton Rouge, LA 70808; PH (225) 767-9162; FAX (225) 767-9108; email: doczhang@dotd.la.gov
[3] Department of Civil and Environmental Engineering and Louisiana Transportation Research Center, Louisiana State University, 4101 Gourrier Ave, Baton Rouge, LA 70808; PH (225) 767-9126; FAX (225) 767-9108; email: louaym@lsu.edu

Abstract: This paper presents the finding from a full-scale accelerated pavement testing study on three flexible pavement test lanes. The three lanes generally varied in their base (or interlayer) and sub-base materials, but had similar hot mix asphalt top layers. The objectives of this study were: (1) to determine the effectiveness of using an untreated reclaimed asphalt pavement (RAP) interlayer in lieu of stone interlayer in a soil-cement asphalt pavement structure; and (2) to investigate the performance of soil cement sub-base layers by varying layer thickness and cement content. The accelerated test was terminated due to severe rutting developed on each of test lanes. The field performance was monitored using Falling Weight Deflectometer (FWD) and DYNAFLECT tests as well as rutting and surface cracking measurements. This study concluded that an untreated RAP interlayer in a soil-cement pavement structure was as effective as the crushed stone interlayer. In addition, the rutting performance of a soil cement layer was found influenced more by layer thickness than by cement content.
Key words: *Flexible Pavement, soil cement, untreated RAP, Accelerated loading, Rutting, FWD, DYNAFLECT, VESYS*

1 Introduction

Soil cement base is a composite material compacted with pulverized soil, portland cement, water, and possibly admixtures. Soil cement mixture hardens to form a rigid base material that enhances the strength characteristics of a pavement structure. Due to cement hydration, however, shrinkage cracking in the soil cement layer is a major cause of reflective cracking found in asphalt pavements. The reflective cracks, in turn, accelerate pavement deterioration and decrease pavement life. Because of problems associated with reflective cracks of soil cement bases, the Louisiana Department of Transportation and Development (LADOTD) adopted a stone "interlayer" base course design as a standard option for soil cement pavements in Louisiana in 2004 (Rasoulian, 2004). To relieve the reflective cracks, the stone interlayer pavement system consists of a layer of crushed stone on top of a cement-stabilized base course. The tensile stresses developed within the base can be absorbed by the stone particles, thus, minimizing reflective cracking. A recent study conducted at the Louisiana Transportation Research Center (LTRC) reported that, for low-volume roads, the initial cost of the stone interlayer pavement is slightly more than that of an in-place soil cement base course; but a 30-year life-cycle cost analysis indicated that the stone interlayer construction would reduce the annual cost per lane-mile by 28 percent due to the improved pavement performance (Rasoulian, 2004).

The use of reclaimed asphalt pavements (RAP) is not new to the highway industry. RAP usage has been reported as early as in the 1930s (Taylor, 1977). However, it was not until the mid to late 1970s that hot mix asphalt (HMA) recycling became popular (Roberts et al, 1991).

In current practice, RAP is typically used as an aggregate in cold-mix asphalt, as base course treated with non-bituminous materials (such as fly ash), or as a fill or embankment material (Collins and Ciesielski, 1994). RAP can be also used in a flexible pavement construction as a granular base course material without any treatments. However, the acceptance of untreated RAP as a granular pavement base material has been limited because of the lack of laboratory testing information and field performance data (Grag and Thompson, 1996). In addition, there is very limited published literature on the use of untreated RAP as a base material. Garg and Thompson conducted a field demonstration project where the crushed RAP was considered as pavement base course and compared to a dense-graded aggregate base course (Grag and Thompson, 1996). Based on the field construction experience as well as field and laboratory testing results, they reported that crushed RAP can successfully be used as a conventional flexible pavement base material. However, this study lacks the long-term pavement performance data to support such conclusion.

Considering that Louisiana does not have a natural source of high-quality stone aggregates and must rely on imported material, and also that the costs of stone base courses have increased with the steady rise in transportation costs, the potential utilization of RAP material as the interlayer in soil cement pavement construction would save the state millions of dollars annually in addition to improving the life of the pavements.

2 Objectives

The primary objective of this study was to determine the effectiveness of using an untreated RAP interlayer in lieu of stone interlayer in a soil-cement asphalt pavement structure under accelerated loading. The secondary objective was to investigate the performance of soil cement sub-base courses by varying layer thickness and cement content. The third one was to predict the accelerated pavement rutting using the VESYS computer program.

3 Accelerated Loading Facility (ALF) Experimental Design

Figure 1 shows a picture of the Louisiana ALF. It is used to simulate dual tires of a single truck axle with the load adjustable from 43.4 kN to 93.5 kN. With a computer-controlled load trolly, the load axle moves repetitively in one direction at an approximate speed of 17 km per hour.

Figure 1. Louisiana ALF

3.1 Pavement structures

Three test lanes were constructed in this study, Figure 2. Each lane had similar HMA top layers paved in two courses. The bases (or interlayers) and sub-bases for Lanes 1, 2, and 3 were: 89 mm untreated RAP and 254 mm, 5 percent cement stabilized soil; 89 mm untreated RAP and 152 mm, 10 percent cement stabilized soil; and 89 mm crushed stone and 152 mm, 10 percent cement stabilized soil, respectively. All three lanes had silty clay embankment. Generally, lanes 1 and 2 have different soil cement sub-base courses, whereas, lanes 2 and 3 have different interlayer courses.

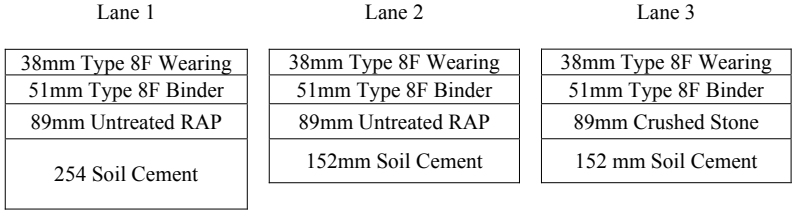


Figure 2. Pavement structures of ALF test lanes

3.2 Pavement materials

3.2.1 Hot mix asphalt (HMA) mixture

Conventional wearing course mixture (Type 8F), which is designed for high volume road applications in Louisiana, was selected for both the wearing and binder course mixtures of test lanes in this study. Specifically, an elastomeric polymer-modified asphalt cement, classified as PG 76-22, was specified for this project. The binder met the LADOTD specification for PG 76-22M (LADOTD, 2000). The final aggregate blend included 43.2 percent No. 67 and 17.3 percent No. 78 crushed siliceous limestone, 21.1 percent No. 11 screenings, and 14.4 percent coarse siliceous sand. The optimum asphalt binder content of 4.0 percent was determined from compacted mixtures with the Marshall Hammer at 75 blows per face at 4.0 percent air voids. Table 1 presents the Job Mix Formula.

Table 1. Job mix formula

Sieve (mm)	% Passing		Properties	Average	Specification
	Type 8F	Specification			
25	100	100	% AC	4.0	/
19	97	91-100	Gmb	2.430	/
12.5	84	78-89	Gmm	2.592	/
9.5	66	62-74	%Gmm	96.0	/
4.75	39	33-45	%Voids	4.0	3-5
2.0	26	20-31	% VMA	13.4	13 min
0.43	13	7-15	% VFA	70	/
0.08	6	1-9	Stability, lbs	2588	2000 min
0.075	4.0	2.0-5.5	Flow, 1/100in	10	6-15

3.2.2 Base, subbase and embankment materials

The stone base selected in Lane 3 consisted of 100% crushed limestone, which complies with the gradation specified by the LADOTD specification (LADOTD, 2000). The optimum moisture content for the crushed stone base was 5.9 percent with a maximum dry density of 21.8 kN/m3 (138.7 lbs/ft3). The RAP base materials used in Lanes 1 and 2 were milled from existing Type 8 asphalt wearing course mixtures constructed for a previous ALF experiment. After milling, it was stockpiled adjacent to the work area to be used in the construction. This RAP material was checked and complied with the gradation specified by LADOTD (LADOTD, 2000). The optimum moisture content was 8.6 percent with a maximum dry density of 18.4 kN/m3. Table 2 presents the gradations for both stone and RAP bases.

Table 2. Gradations for stone and RAP bases

Sieve, mm	RAP Base		Stone Base	
	% Passing	Specification	% Passing	Specification
63.4	100	100	100	
50.8	96.6		100	
38	96.0		100	100
25.4	92.7		94.3	
19	89.1		83.8	50-100
12.5	80.8		72.2	
9.5	71.4		65.6	
4.75	51.8	35-75	52.7	35-65
2.36	36.5		33.7	
1.18	34.0		30.6	
0.6	19.3		20.3	
0.43	13.9		18.5	10-32
0.3	9.8		17.1	
0.15	3.1		15.3	
0.075	0.45		12.9	3-15
W_{opt} (%)	8.6		5.9	
$\gamma_{max,}$ kN/m^3	18.4		21.8	

An A-4 soil was used in both the soil cement sub-base and the embankment subgrade. Table 3 presents the basic soil properties. Specifically, the cement content used in Lane 1's subbase was 5 percent by weight, while 10 percent cement content was used for both lanes 2 and 3.

Table 3. Soil properties

Passing # 200 (%)	Clay (%)	Silt (%)	Organic (%)	LL(%)	PI	W_{opt} (%)	γ (kN/m^3)	Classification	
								USCS	AASHTO
93	23	70	4.7	28	22	16.5	17	CL -ML	A-4

3.3 ALF loading

The applied load was 43.4 kN for the first 200,000 cycles. The load was then increased to 53.6 kN after 200,000 cycles and 63.9 kN after 525,000 cycles. An average rut depth of 12.5 mm at the surface and/or 50 percent of the trafficked area with cracking more than 5 m/m2 was selected as the initial failure criterion. Figure 3 presents the ALF loading history together with the corresponding 80-kN Equivalent Single Axial Load (ESAL) values (converted based on the fourth power law).

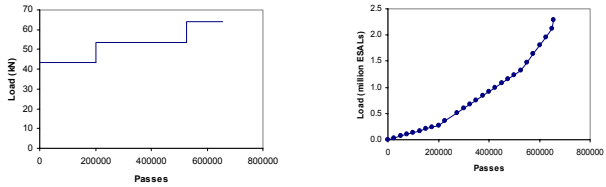


Figure 3. ALF loading history

3.4 Field measurements

Non destructive test (DYNAFLECT and FWD) as well as the rutting and cracking survey were performed at the end of each 25,000 load repetitions. The effective loading area of ALF testing is about 12m (40 ft) long, in which deflection measurements and distress survey were taken at 8 stations over a length of 9 meters.

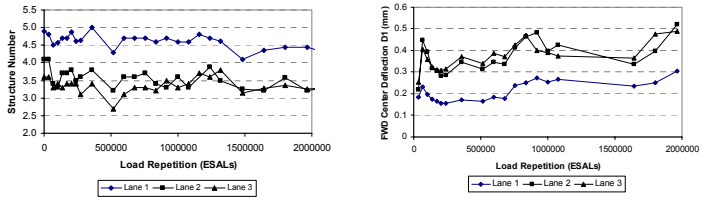
A Dynatest 8002 model FWD device was used with nine sensors spaced at 0, 200, 300, 450, 600, 900, 1,200, 1,500 and 1,800 mm (0, 8, 12, 18, 24, 36, 48, 60 and 72 in.), respectively. The DYNAFLECT used in this study is a trailer mounted device which induces a dynamic load of 4.45 kN (1,000 lb) at a frequency of 8 Hz on the pavement and measures the resulting deflections by using five geophones spaced under the trailer at approximately 300 mm (1 ft) intervals from the application of the load. A structure number (SN) and subgrade resilient modulus of existing pavement structures can be determined using a nomograph chart method based on a temperature-corrected Dynaflect center deflection and a spreadability value. More details can be referred elsewhere (Kinchen and Temple, 1980).

4 Discussion of Results

4.1 DYNAFLECT and FWD test results

Figure 4(a) presents the average structure number (SN) values predicted from DYNAFLECT test results for the three test lanes evaluated under different loading passes (in ESAL values). Lane 1, with a RAP interlayer and 254-mm thick cement soil sub-base, possesses the highest SN values among the three tested lanes, with an initial SN value of 4.9 and an ending value of 4.4 after 2 million ESALs loading. By comparing Lane 1 and Lane 2's pavement structures, the SN results indicate that a pavement structure with a thicker cement soil subbase generally has a higher structure capacity than a thinner pavement structure, even though the thinner one had higher cement content. On the other hand, the SN values for Lane 2 and Lane 3 were very similar during the whole loading period. These observations imply that Lane 1 should perform better than both Lanes 2 and 3, whereas, Lane 2 and Lane 3 may perform similarly.

Figure 4(b) presents the average FWD center deflection test results for the three test lanes evaluated. The center deflection, measured directly under the FWD loading plate, was used as an indicator of the overall capacity of a pavement structure in this study. High surface deflection indicates weak loading capacity. Figure 4(b) shows that the center deflections of Lane 1 were much lower than both Lane 2 and Lane 3, but the values for Lanes 2 and 3 are generally similar to each other. Therefore, FWD results further confirms the observations from DYNAFLECT's SN results, which indicates that in terms of overall loading capacity, Lane 1 is the strongest while Lane 2 and Lane 3 are similar.



(a) (b)

Figure 4. (a) Predicted Structure Number results; (b) Average FWD center deflection test results on ALF test lanes

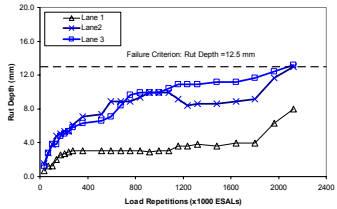


Figure 5. Average rut depth measurements

4.2 Rut measurements of ALF test lanes

The accelerated pavement loading was terminated at the end of 650,000 loading repetitions (about 2.1 million ESALs), when the average rut depth measured on both Lanes 2 and 3 reached the rutting failure criteria of 12.5 mm. None of the test lanes exhibited any surface cracking. As shown in Figure 5, the final average rut depths for Lanes 1, 2 and 3 were 8.0, 13.0 and 13.2 mm, respectively.

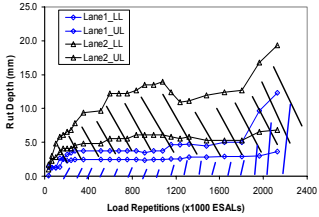
The average rut depth and standard deviation for each test lane were evaluated through the computation of upper limit (UL) and lower limit (LL) of 95 percent confidence level by the following equations (Mohammad et al, 2000):

$$UL = d_{ave} + t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \qquad (1) \qquad LL = d_{ave} - t_{\alpha/2} \cdot \frac{s}{\sqrt{n}} \qquad (2)$$

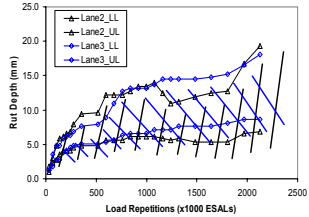
Where: UL – Upper limit of rut depth, mm,
LL – Lower limit of rut depth, mm,
d_{ave} – Average rut depth, mm,
tα/2 – t-value at confidence level of (100-α), α = 5,
s – Standard deviation, and
n – Measuring sample number.

Figure 6(a) shows the corresponding rut depth measurements at 95 percent confidence levels for Lane 1 and Lane 2, whereas, Figure 6(b) presents the corresponding rut depth measurements at 95 percent confidence levels for Lane 2 and Lane 3. At 95 percent confidence, the rut depths of Lane 1 were significantly smaller than those in Lane 2 as indicated by hatched area showed in Figure 6(a). This indicates that the thicker soil-cement layer in Lane 1 performed better than the thinner soil cement layer in Lane 2, even though Lane 1 possesses lower cement content. On the other hand, the rut depths of Lane 2 and Lane

3 were similar to each other as the majority of hatched areas overlapped, Figure 6(b). This observation is consistent with both DYNAFLECT and FWD analysis results.



(a) Rut Depths at 95% Confidence Levels between Lane 1 & Lane 2



(b) Rut Depths at 95% Confidence Levels between Lane 2 & Lane 3

Figure 6. Statistic analyses of rut depth measurements at ALF test lanes

The rutting analysis on ALF test lanes showed that the untreated RAP layer performs as effectively as the stone interlayer based on the rutting comparison between Lane 2 and Lane 3. On the other hand, the rutting performance of a soil cement layer was observed to be more sensitive to the layer thickness than the cement content of the layer (between Lane 1 and Lane 2).

4.3 FWD deflection analysis

Figure 7 presents the measured rut depths and FWD center deflections at the center station of the loading area for Lane 1 under increasing ALF load repetitions. It is noted that similar observations were made on both lanes 2 and 3.The center deflections (D1) were first normalized to a standard load of 40-kN (9,000 lbs) and then temperature-corrected to 20^{o}C (Lukanen et al, 2000).

As shown in Figure 7, the surface deflection generally shows an increasing tendency due to trafficking, but also fluctuates because of the seasoning effect. The temperature correction of surface deflection only adjusts the temperature effects on the deformation of the HMA layer (which in this study is 89 mm thick), while the seasoning effect, such as moisture content, contributes to the seasonal variations of the strengths for the base and subgrade materials. In this study, high surface deflections were generally measured during the period between April and October, in which the measured deflections in June or July were the highest. High surface deflections indicate that a pavement structure is in its weak condition during that period. This season effect links very well to the weather condition of the ALF test location. Between April and October, ALF test pavements received the majority of a year's rainfall, which in turn gradually weakens both the subgrade and the soil-cement subbase. After summer, the moisture content gradually reduces due to evaporation, and thus the strengths of subgrade and soil-cement layer will be regained.

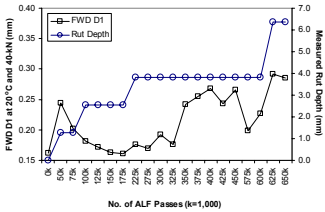


Figure 7. FWD deflection versus and rut depth on Lane 1

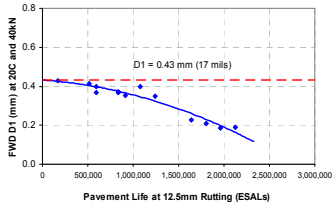


Figure 8. Relationship between FWD deflection and pavement rutting life

4.3.3 FWD deflection in predicting pavement rutting life

An attempt was made in this study to predict the pavement rutting life directly using the FWD surface deflection measurements. A flexible pavement structure is assumed to reach its rutting life when the measured rut depth is 12.5 mm (0.5 in.). It was noted that the FWD tests performed on ALF test lanes at the end of 50,000 ALF passes were in the summer of 2001, when pavement structures were weakest in terms of load-carrying capacity due to high surface temperatures and base moisture contents. Also, it can be reasonably assumed that after 50,000 passes the initial consolidation of pavement materials has been finished. Therefore, the center deflection data (D1s) at the end of 50,000 passes was selected to represent the pavement load-carrying capacity in the rutting analysis. Figure 8 presents the relationship between the center deflection and the pavement rutting life. A fairly strong relationship was observed between the FWD deflection (D1) and the pavement life at 12.5 mm rutting. As the D1 increases, the pavement rutting life decreases. The observed threshold limit for FWD D1 at 20 ^{o}C and 40 kN is 0.43 mm (17 mils). Below this value, pavement can withstand some trafficking before reaching its rutting life, whereas, above this value, the pavement would develop a rut depth of 12.5 mm immediately. This observation is promising because it provides a potential for pavement engineers to directly use FWD measured deflection data in predicting pavement rutting performance and making decisions when pavement maintenance is needed.

4.4 Pavement rutting prediction

The VESYS 5W computer program – a windows version of VESYS 5 (Kenis et al, 1982; Zhou and Scullion, 2002) – was employed in this study to predict the rutting development of ALF test lanes. The basic assumption used in the VESYS rutting prediction model is that the permanent strain developed in each pavement layer is a linear function of layer deflection and the logarithm of the number of load repetitions as expressed in Equation 3:

$$\Delta\varepsilon_p(N) = \varepsilon_e \times \mu \times N^{-\alpha} \qquad (3)$$

Where: $\Delta\varepsilon_p$ – Permanent strain per load repetition,
ε_e – Dynamic resilient strain,
α, μ – Material permanent deformation parameters,
N – Number of load repetitions.

The primary inputs required for rutting prediction include load repetitions with different load amplitudes, material elastic modulus information, and α- and μ- values of each pavement layer. To determine the permanent deformation parameters (e.g. α, μ) of pavement materials, a repeated-load permanent deformation test was conducted on each pavement material used on those test lanes. The detailed discussion on repeated-load permanent deformation tests can be found elsewhere (Mohammad et al, 2005). The test results of the permanent deformation parameters for each material are listed as follows:

Type 8F HMA:	$\mu = 0.071$, $\alpha = 0.688$ (at temperature of 25°C)
	$\mu = 0.245$, $\alpha = 0.716$ (at temperature of 40°C)
RAP:	$\mu = 0.102$, $\alpha = 0.613$
Stone:	$\mu = 0.020$, $\alpha = 0.905$
Cement Soil (10%):	$\mu = 0.012$, $\alpha = 0.821$
Embankment Soil:	$\mu = 0.005$, $\alpha = 0.470$

Since the soil cement mixture with a 5-percent cement content (by weight) used in Lane 1 was not tested in this study, the VESYS rutting prediction was limited only for Lane 2 and Lane 3. Figure 9 shows the predicted versus the measured rut depths for Lanes 2 and 3.

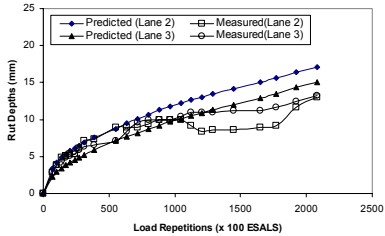


Figure 9. Predicted vs. measured rut depth on Lanes 2 & 3

As indicated in Figure 9, the predicted rutting development curves correlated reasonably well to the measured ones up to 0.6-million and 1.25-million ESALs for Lane 2 and Lane 3, respectively. After that, the predicted versus the measured curves depart from each other with the predicted rut depths higher than the measured ones. This can be explained by the fact that the rutting model used in the VESYS program can only consider the primary rutting development in a pavement structure. After the initial primary rutting period, the rutting development in a pavement structure enters into its secondary stage with a lower rutting cumulative rate. As load is continuously applied, it eventually reaches into its third stage (tertiary flow) and then pavement fails with a rapid rutting accumulation. The three-stage rutting development can be viewed in Figure 9, where the measured rutting for both Lanes 2 and 3 clearly indicates the pavements failed at the third rutting stage. On the other hand, the predicted rutting development in Lane 2 was generally higher than that in Lane 3. This is because the lab-determined permanent deformation parameters (μ- and α-values) used for the stone interlayer of Lane 3 resulted in a smaller primary rutting cumulative rate than those for the RAP interlayer of Lane 2. More research on predicting the secondary rutting and tertiary flow are needed for the VESYS program. In summary, the VESYS 5W program proved to be a useful tool for predicting the pavement rutting in this study. With the laboratory measured material parameters, it can be used to predict the primary rutting up to 10 mm of a pavement structure.

5 Summary and Conclusions

Three test lanes were constructed at the Louisiana Pavement Research Facility to evaluate the performance of soil-cement pavements with different interlayers and soil-cement sub-base courses. Lane 1 had a 89-mm reclaimed asphalt pavement (RAP) interlayer and a 254-mm, 5 percent cement soil-cement sub-base. The interlayers and sub-bases for Lanes 2 and 3 were 89-mm RAP and 152-mm, 10 percent cement stabilized soil, and 89-mm crushed stone and 152 mm, 10 percent cement stabilized soil, respectively. DYNAFLECT and FWD tests were conducted periodically to monitor the pavement performance. The accelerated

loading was accomplished at the end of 650,000 load passes. Significant rutting was observed on all three test lanes. The following observations and conclusions were made from this study:

- Both DYNAFLCET and FWD test results indicated that Lane 1 had a better loading capacity compared to Lanes 2 and 3, whereas Lane 2 and Lane 3 displayed similar loading capacities.
- Lane 1 developed significantly lower rut depths than both Lanes 2 and 3. Lanes 2 and 3 showed similar rut developments. At a 95 percent confidence interval, the rut development in Lane 1 was significantly lower than Lane 2. On the other hand, the rut development in Lane 2 was essentially similar to that in Lane 3 at 95 percent confidence.
- FWD surface deflection was found to possess a good, direct relationship with the pavement rutting performance.
- A good agreement was observed between the predicted ruts from the VESYS program and the field measured ones, up to a rut depth of 10 mm.
- The usage of the untreated RAP materials as the interlayer in a soil-cement pavement structure was as effective as the crushed stone interlayer.
- The rutting performance of a soil cement layer was found to be influenced more by layer thickness than by cement content.

Acknowledgement

This study was supported by the Louisiana Transportation Research Center and the Louisiana Department of Transportation and Development. The authors would like to express thanks to all those who provided valuable help in this study.

References

Collins, R.J. and Ciesielski, S.K.(1994). "Recycling and Use of Waste Materials and By-Products in Highway Construction." National Cooperative Highway Research Program Synthesis of Highway Practice 199, pp. 43-46.

Grag, N., and Thompson, M.R. (1996). "Lincoln Avenue-Reclaimed Asphalt Pavement Base Project." Transportation Research Record 1547, Washington D.C., pp.89-95.

Kenis, W., Sherwood J. A., and McMahon, T.F. (1982). "Verification and Application of the VESYS Structural Subsystem." *The Fifth International Conference on the Structural Design of Asphalt Pavements*. The Delft University of Technology.

Kinchen, R.W. and Temple, W.H. (1980). *Asphalt concrete overlays of rigid and flexible pavements*. Report No. FHWA/LA-80/147, Louisiana Department of Transportation and Development, Baton Rouge, Louisiana.

LADOTD. (2000). "Louisiana Standard Specifications for Roads and Bridges," Louisiana Department of Transportation and Development, Baton Rouge, 2000 Edition.

Lukanen, E.O., Stubstad, R. and Briggs, R. (2000). "Temperature Predictions and Adjustment Factors For Asphalt Pavement." Publication NO. FHWA-RD-98-085.

Mohammad L.N., Huang, B., Roborts, F. and Rasoulian, M. (2000). "Accelerated Loading Performance and Laboratory Characterization of Crumb Rubber Asphalt Pavements." International Journal of Road Materials and Pavement Design, Vol. 1, pp467-493.

Mohammad L.N., Wu, Z., Xu, Q. and Rasoulian, M. (2005). "Accelerated Pavement Testing of Untreated RAP Layer in Flexible Pavement." Proceedings of the Fourth International Conference on Maintenance and Rehabilitation of Pavements and Technological Control, 8-19th August, University of Ulster, Northern Ireland, UK

Rasoulian, M. (2004). **"**Research Pays Off - Stone Interlayer Pavement System: Extending the Service Life of Low-Volume Roads." Transportation Research News (TR News) 233, July/August, bimonthly magazine, Transportation Research Board.

Taylor, N.H. (1977) "Life Expectancy of Recycled Asphalt Paving," Recycling of Bituminous Pavements, Editor, L. E. Wood, ASTM STP 662, American Society for Testing Materials, Philadelphia, PA, pp. 3-15.

Zhou, F. and Scullion, T. (2002). "VESYS5 Rutting Model Calibrations with Local Accelerated Pavement Test Data and Associated Implementation." *Report No.FHWA/TX-03/9-1502-01-2*. TX.

ASPHALT PAVEMENT STATISTICAL TEMPERATURE PREDICTION MODELS DEVELOPED FROM MEASURED DATA IN CHINA

Lu JIA[1], Lijun SUN[2], Yonghua YU[3]

[1]Key Lab of MOE, Transportation Engineering School, Tongji University. Siping Road 1239, Shanghai, P.R. China, 200092; email: I-am-Robin@163.com
[2]Transportation Engineering School, Tongji University. Siping Road 1239, Shanghai, P.R. China, 200092
[3]Quality Inspection and Supervision Station of Liaoning Transportation; Shashan Street 128#, Shenyang, P.R.China, 110005

Abstract: The temperature distributions in asphalt pavement have a great effect upon the bearing capacity and performance of asphalt pavement. The Transportation Engineering School of Tongji University started comparative research about 5 years ago. Thermal sensors were installed in several pavement sections in different regions to monitor the temperature fluctuation inside the pavement structure. Based on data from this research program and related weather information, the application of SHRP and LTPP pavement temperature prediction model in China is verified. Then statistical temperature prediction models established through regression analysis are introduced in detail. Compared with the other existing prediction models, the models show comprehensive applicability and excellent accuracy.
Key words: *temperature distribution, statistical methods, temperature prediction model, linear regression*

1 Background

The temperature profile in asphalt pavement structure has a significant effect on asphalt pavement performance because of the temperature susceptibility of the asphalt mixture. If the temperature distribution in pavement could be accurately predicted any time, the extreme high and low temperature in pavement service life can be determined and the proper PG grading for the asphalt binder can be selected.

The pioneering research about temperature distribution in asphalt pavement structure was carried out by Barber depending on the heat diffusion theory (Barber 1957). Based on heat and energy transfer theory, Solaimanian and Kennedy proposed a model using calculated maximum solar radiation and measured high air temperature to predict the extreme pavement temperature (Solaimanian and Kennedy 1993), which was employed in the Strategic Highway Research Program. Regression based models using data from the Seasonal Monitoring Program of Long Term Pavement Performance Research (LTPP-SMP) was developed to relate the seven-day average high air temperature to the seven-day average high pavement temperature by Lukanen et al (Lukanen et al 1998). In order to validate the Superpave pavement temperature

algorithm and its limitation, Bosscher developed a refined statistical model for estimation of low and high pavement temperatures from meteorological data (Bosscher et al 1998).

In China, without long term monitoring data about asphalt pavement temperature and related weather information, the systemic and continuous research about temperature distribution in asphalt pavement had not been carried out until recent years. In order to obtain an accurate and convenient temperature prediction model for asphalt pavement, the School of Transportation Engineering in Tongji University started research about 5 years ago. It is a good opportunity to verify those existing temperature prediction models such as SHRP and LTPP at different reliability levels. Then statistical prediction models are developed through regression analysis between the measured pavement temperature data and meteorological data from weather stations in different regions in China, which includes the monthly historical average air temperature as the regional correction coefficient. The suggested models were successfully validated utilizing data from the test sections in this research.

2 Introduction of Asphalt Pavement Temperature Monitoring Program in China

Funded by the Western Transportation Construction Projects (Grant No. 200131800095) and the National Science Fund for Distinguished Young Scholars of China (Grant No. 50325825), Tongji University conducted specific research for an asphalt pavement temperature profile. In this research, several test sections are selected in different cities around the country. The detailed information about those test sections is listed in Table 1 below.

Table 1. Detailed information about test sections

City	Location		Monitoring Phase	Air Temperature	Solar Radiation
	Lat.(^{o}N)	Long.(^{o}E)			
Shanghai	31.25	121.47	2002.7-8	Available	Available
			2003.11-12	Available	Available
Urumchi	43.77	87.50	2003.3-4	Available	Available
Tangshan	39.62	118.18	2004.7-8	Available	Available
Shijiazhuang	38.07	114.47	2005.1-6	Available	N.A.
Xi'ning	36.62	101.77	2003.1-3	Available	N.A.
Wuhan	36.74	101.77	2005.8-11	Available	Available

In our research, the PT100 platinum resistance thermo sensors and EN880 data logger are used to monitor the temperature in the asphalt pavement structure. The EN880 data logger can automatically store the data of 12 channels. For asphalt pavement temperature monitoring, the interval time is set at 600s (10 min), which means the data logger would receive 6 pavement temperature readings for each channel per hour.

To install the thermo sensors into the pavement structures, a sample on the test sections was first cored . Then the sample was drilled at different depths. After that, the thermo sensors were embedded in the sample at different depths, and were connected to the data logger by cables. Finally, the sample was put back into the pavement structure and fixed with asphalt sand.

The typical variation of pavement temperature and air temperature are shown in figure 1. During most of the time, the asphalt pavement surface temperatures are almost synchronous with the air temperature. The daily maximum air temperature and the pavement surface temperature happen at almost the same time. With the increasing depth, there would be obvious lags between the pavement temperatures and the air temperatures.

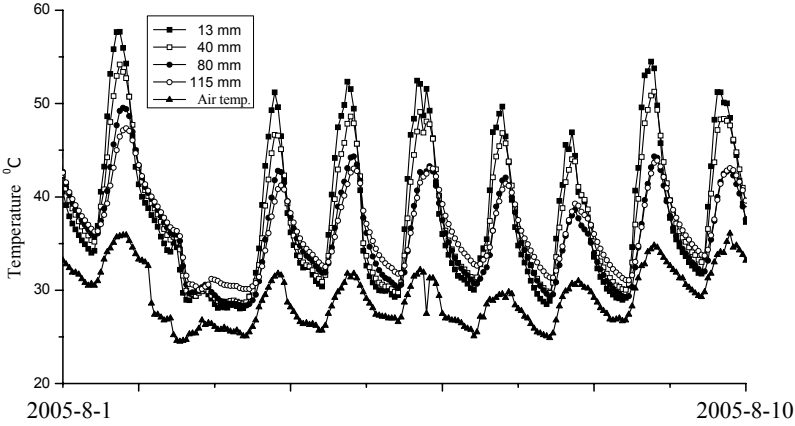


Figure 1. Asphalt pavement temperature and air temperature

3 Verify Temperature Prediction Models in China

3.1 The SHRP Model

(1)High Pavement Temperature Model

In SHRP-A-648A, "Weather Database for the Superpave Mix Design System" an equation which described the relationship between air temperature and pavement surface temperature was deduced as follows:

$$T_{surf} = T_{air} - 0.00618Lat^2 + 0.2289Lat + 24.4 \quad (1)$$

Where T_{surf} is the pavement surface temperature expressed in □, T_{air} is the 7-day highest average air temperature in □, and Lat is the latitude in degrees.

Below the pavement surface, the temperature can be predicted as below:

$$T(d) = T_{surf}\left(1 - 0.0063d + 0.007d^2 - 0.0004d^3\right) \quad (2)$$

Where *T(d)* is the pavement temperature at d in inches, *T(d)* and T_{surf} are the pavement temperature in □.

In the Superpave program the following formula was developed to predict the pavement temperature at a depth of 20 mm:

$$T_{20mm} = (T_{air} - 0.00618Lat^2 + 0.2289Lat + 42.2)(0.9545) - 17.78 \quad (3)$$

Where T_{20mm} is the pavement temperature at 20mm in □.

The predicted model considering reliability is calculated as follows:

$$T_{pav} = T(d) + z\mathrm{g}S_{air} \quad (4)$$

Where T_{pav} is the high asphalt pavement temperature with reliability,, in □; *T(d)* is the high asphalt pavement temperature at calculated depth, in □; S_{air} is the standard deviation of the high 7-day mean air temperature, in □; *z* is the standard normal dist, *z*=0 for 50% reliability, *z*=2.055 for 98%r reliability.

(2) Low Pavement Temperature Model

For low pavement temperature, the SHRP model assumed the surface temperature to be equal to the minimal air temperature. Then an equation was developed for the change in temperature with depth based on the low air temperature as follows:

$$T(d) = T_{surf} + 0.051d - 0.000063d^2 \quad (5)$$

Where T_{surf} is the pavement surface temperature which is equal to the T_{air}, in□, *d* is the depth in mm.

The predicted model considering reliability can be described as follows:

$$T_{pav} = T(d) - z\mathrm{g}S_{air} \quad (6)$$

Where T_{pav} is the low asphalt pavement temperature with reliability, in□; *T(d)* is the low asphalt pavement temperature at calculated depth, in℃; S_{air} and *z* are the same parameters as indicated in equation 4.

3.2 The LTTP Model

(1) LTPP High Temperature Model

The LTPP high temperature model is an empirical model developed from LTPP seasonal monitoring data. This model relates high pavement temperature to air temperature, latitude, and depth. The LTPP high pavement temperature model is:

$$T_{pav} = 54.32 + 0.78T_{air} - 0.0025Lat^2 - 15.14\lg(H + 25) + z(9 + 0.61S_{air}{}^2)^{0.5} \quad (7)$$

Where T_{pav} is the high AC pavement temperature below surface, in □, T_{air} is the high air temperature, in □, *Lat* is the Latitude of the section, in degrees, *H* is the depth to surface, in mm, S_{air} is the standard deviation of the high 7-day mean air temperature, *z* is the standard normal dist., *z* = 0 for 50% reliability, *z*=2.055 for 98%r reliability.

(2) LTPP Low Temperature Model

An equation developed by LTPP for the change in temperature with depth for low temperature is:

$$T_{pav} = -1.56 + 0.72T_{air} - 0.004Lat^2 + 6.26\lg(H+25) - z(4.4 + 0.52S_{air}{}^2)^{0.5} \quad (8)$$

Where T_{pav} is the low AC pavement temperature below surface, in □, T_{air} is the low air temperature, in □. *Lat, H,* S_{air} and *z* are the same parameters in equation 7.

3.3 Verify the Temperature Prediction Models in China

(1) High Pavement Temperature Model

According to the temperature data from test sections in different regions of China, the application of SHRP and LTPP temperature prediction models in China are fully investigated.

In this study, the measured pavement temperature data of Shanghai City from Jul. to Aug. 2002 and the data of Wuhan City from Jul. to Aug. 2005 are compared with the predicted pavement temperature, depending on the SHRP and LTPP high temperature models at different depths,The results are listed in tables 2~3.

Table 2. Measured and predicted high pavement temperature in Shanghai

2002.7-8 T_{air}=28.6ºC	Measured Temp.	SHRP Model		LTPP Model	
		50%	98%	50%	98%
T_{surf}	54.9	54.1	60.3	53.0	60.8
T_{20mm}	51.4	50.8	57.0	49.1	56.9
T_{40mm}	47.9	48.1	54.3	46.7	54.5
T_{60mm}	46.1	45.8	52.0	45.0	52.8
T_{80mm}	45.0	43.9	50.1	43.6	51.4
T_{100mm}	42.8	42.3	48.5	42.4	50.2
T_{120mm}	39.7	40.9	47.1	41.4	49.2
T_{140mm}	34.9	39.6	45.8	40.6	48.4

The summer of 2002 was cool ,and the 7-day highest average air temperature in Shanghai happened from Aug. 18th to 26th, and the Tair was only 28.6□. The SHRP model and LTPP model at 50% reliability provided the acceptable predicted results within 2□, but over-estimated the temperature by 5.0□ at 140mm depth. The SHRP model and LTPP model at 98% reliability overestimated the pavement temperature by 6.0~9.0□.

Table 3. Measured and predicted high pavement temperature in Wuhan

2005.7-8 T_{air}=28.6ºC	Measured Temp.	SHRP Model		LTPP Model	
		50%	98%	50%	98%
T_{surf}	61.6	57.3	63.7	55.5	63.4
T_{20mm}	59.1	53.9	60.3	51.7	59.6
T_{40mm}	55.8	51.1	57.5	49.2	57.1
T_{60mm}	55.0	48.7	55.1	47.5	55.4
T_{80mm}	50.8	46.7	53.1	46.1	54.0
T_{100mm}	48.9	45.0	51.4	44.9	52.8
T_{120mm}	46.7	43.5	49.9	44.0	51.9

The 2005 summer in Wuhan was a more typical summer. The 7-day highest average air temperature in Wuhan happened from Aug. 4th to 11th. The T_{air} was as high as 31.7□. The SHRP model and LTPP model at 50% reliability both underestimated the pavement temperature in the whole asphalt layer by 3-4□, but the SHRP Model at 98% reliability more closely predicted pavement temperature; the predicted errors are within 2~3 □. For the LTPP model with 98% reliability, the predicted temperature from pavement surface to 60mm depth fit well with the measured temperature and the predicted errors are within 1~2□. But from 80mm to 120mm deep, the predicted results are higher than the measured temperature by 4~6□.

(2) Low Pavement Temperature Model

The measured pavement temperature data of Shanghai City from Nov. to Dec. 2003 and the data of Xi'ning City from Dec. 2003 to Mar.2004 are compared with the predicted pavement temperature, depending on the SHRP and LTPP low temperature models at different depths. The results are listed in tables 4~5.

From the data shown in tables 4&5, the SHRP and LTPP low temperature models do not closely agree with measured values. For the test sections at Shanghai, the measured temperature at pavement surface is similar to the SHRP predicted results, which is close to the minimum air temperature -2.2□. Along with the depth, the SHRP model at 50% reliability will underestimate the pavement temperature by 1-3□, at 98% reliability it will underestimate the temperature by 9-10□. But the LTPP model at 50% reliability will overestimate the pavement temperature by 4-7□ at any depth, at 98% reliability it will underestimate the temperature by 4-6□.

For the test section in Xi'ning, the minimum air temperature is extremely low - as -19.2□. The SHRP model at 50% reliability does provide closely predicted results from 123mm to 223mm, and the estimated errors are no more than 1□. But it will underestimate the pavement temperature by 2-3□ from 12mm to 77mm. The SHRP model at 98% reliability will underestimate the temperature by 7-8□. The LTPP model at 50% reliability will overestimate the pavement temperature by 5-8□, and at 98% reliability it will provide closer predicted results, with errors within 2□.

Table 4. Measured and predicted low pavement temperature in Shanghai

2003.11-12 T_{air}=-2.2ºC	Measured Temp.	SHRP Model		LTPP Model	
		50%	98%	50%	98%
T_{surf}	-2.1	-2.2	-11.7	1.7	-6.4
T_{20mm}	-0.9	-2.1	-11.5	1.9	-6.2
T_{40mm}	-0.5	-2.0	-11.4	2.1	-6.0
T_{60mm}	-0.6	-1.9	-11.3	2.3	-5.8
T_{80mm}	0.4	-1.8	-11.2	2.4	-5.7
T_{100mm}	1.4	-1.7	-11.1	2.6	-5.5
T_{120mm}	2.0	-1.6	-11.0	2.8	-5.3
T_{140mm}	1.8	-1.5	-10.9	2.9	-5.2

Table 5. Measured and predicted low pavement temperature in Xi'ning

2003.11-2004.3 T_{air}=-19.2ºC	Measured Temp.	SHRP Model		LTPP Model	
		50%	98%	50%	98%
T_{12mm}	-16.4	-18.6	-24.7	-10.9	-17.1
T_{27mm}	-15.0	-17.9	-24.0	-10.0	-16.2
T_{40mm}	-14.3	-17.3	-23.4	-9.4	-15.6
T_{65mm}	-13.2	-16.2	-22.3	-8.5	-14.7
T_{77mm}	-12.0	-15.6	-21.7	-8.2	-14.4
T_{123mm}	-13.6	-13.9	-20.0	-7.1	-13.3
T_{150mm}	-13.0	-13.0	-19.2	-6.7	-12.9
T_{175mm}	-12.7	-12.2	-18.3	-6.3	-12.5
T_{200mm}	-10.3	-11.5	-17.7	-6.0	-12.2

In summary, the SHRP and LTPP pavement temperature prediction models, including the high temperature model and low temperature models, cannot be appropriately used in China directly. One reason for the discrepancies is that the climates in China are quite different from those in North America, and there are great climatic differences between the test sections in China, even if their latitude, which is considered in the SHRP or LTPP model, is very similar. In order to overcome those problems and to establish a more general prediction model with higher accuracy in China, the regression analysis is carried out as follows.

4 Develop Pavement Temperature Prediction Model in China

It is well known that air temperature is the most significant factor in pavement temperature distribution. Besides air temperature, the solar radiation, only happening in day time, is another main factor to cause the temperature increase in pavement structure. Compared with air temperature and solar radiation, the effects of the cloud conditions, the relative humidity, the wind velocity, and the precipitation to pavement

temperature distribution are not so significant. Therefore the air temperature and solar radiation are the primary factors considered in our prediction model development.

According to the research carried out by Lijun Sun, Lu Jia, et al (Lijun Sun, Lu Jia, et al 2004), the pavement temperature is most sensitive to the average temperature for the previous 5 hours, T_{a5} and the square of solar radiation for the previous 5 hours, Q_5^2. Consequently, T_{a5} and Q_5^2, which are used to reflect the accumulation efforts of environmental factors to pavement temperature in a short phase, are introduced to the prediction models. Moreover, the nonlinear factor, such as air temperature multiplied by depth and solar radiation multiplied by depth ($H \times T_a$, $H \times Q$), are considered to represent that the environmental factors' influences on pavement would vary with depth. Besides that, the three order multinomial function is employed in order to accurately simulate the distribution of pavement temperature with depth. Furthermore, the monthly historical average air temperature is introduced to prediction models as a regional correction coefficient to reflect the regional deference in pavement temperature. The finial asphalt pavement temperature prediction models are described as below:

$$T_p = p_1 + \left(p_2 T_{a5} + p_3 Q_5^2\right) + H\left(p_4 T_a + p_5 Q\right) + \left(p_6 H + p_7 H^2 + p_8 H^3\right) + p_9 T_m \qquad (9)$$

Where: T_p is the pavement temperature at H cm, □; T_a is the air temperature at this time, □; Q is the solar radiation at this time, kW/m^2; T_{a5} is the average air temperature for previous 5 hours, □; Q_5 is the average solar radiation for previous 5 hours, kW/m^2; H is the depth of the predicted point, cm; $p_1 \sim p_8$ are the undetermined regression coefficients for the prediction model; T_m is the monthly historical average air temperature for the past 20 years, □.

Employing the linear regression method, the measured pavement temperature data and weather information from different test sections in China are used to develop the specific prediction models, and the results are listed in table 6.

According to the results listed in Table 7 it can be concluded that the prediction models have very high accuracy, and the coefficient of determination R^2 and the adjusted R^2 are 0.98, and the max standard error in the prediction model is 2.8□.

Table 6 Regression analysis results for pavement temperature prediction model

p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	R^2	R_{adj}^2	Std. Error of the Estimate
-1.23	0.90	24.9	-0.01	-0.80	-0.66	0.11	-0.004	0.47	0.98	0.98	2.80

In order to prove the validation of modified pavement temperature prediction models in different locations, the measured pavement temperature data in Wuhan City, which are not included in the regression analysis when developing the modified

prediction model, are used to be compared with the prediction results from the modified prediction model with the solar radiation, air temperature and the history air temperature data of Wuhan City. As shown in figure 2, the measured data and predicted data are scattered at both sides of the line Y=X. Moreover, the correlation coefficient between the measured pavement temperature data and predicted temperature of Wuhan is 0.933. It can be concluded that the modified pavement temperature prediction model can be used in different regions with high accuracy.

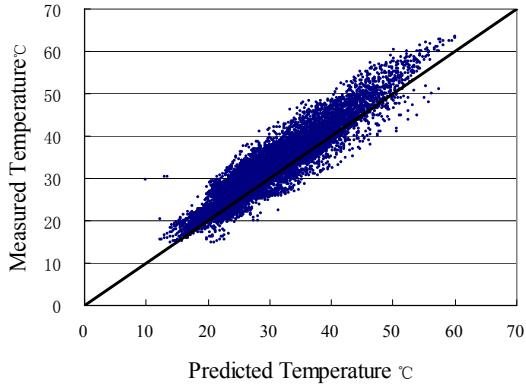


Figure 2 Measured and Predicted Pavement Temperature Using Equation 9

5 Conclusions

In this paper, the pavement temperature monitoring program in China is first introduced in detail . Then depending on the measured pavement temperature and weather information, the application of SHRP and LTPP pavement temperature prediction models in China are fully investigated. The following conclusions have been drawn:

1. Neither the SHRP nor LTPP high temperature models can provide satisfying prediction results for asphalt pavement in China. In some cases, the SHRP and the LTPP high temperature models will underestimate high pavement temperatures. In other cases, they may give a close approximation of pavement temperatures at the pavement surface, but they will overestimate the temperature at the bottom of an asphalt layer.

2. The low temperature models for SHRP slightly overestimate low pavement temperatures, even if temperatures at the pavement surface predicted by SHRP are close to the measured temperature. But the LTPP model underestimated the pavement low temperature. The temperature differences between measured temperature and LTPP predicted results are as much as 8□.

3. Because of the different climatic and geographical conditions between China and North America, the SHRP and LTPP pavement temperature prediction models cannot be directly applied in China.

4. Two general prediction models, which include the air temperature and the historical monthly temperature or the solar radiation, are established by the regression analysis between the measured pavement temperature data and meteorological data in China. The regression analysis results prove that both models show great agreement with the measured data and can be used in different regions in China with high accuracy.

6 Acknowledgements

This research in China was funded by the Western Transportation Construction Projects of Ministry of Communication of P.R China (Grant No. 200131800095) and the National Science Fund for Distinguished Young Scholars of China (Grant No. 50325825).

References

Barber, E.S.(1957) "*Calculation of Maximum Pavement Temperature from Weather Reports."* Highway Research Board Bulletin., No.168.1-8

Bosscher, P. J., H. U. Bahia,et al.(1998) "*Relationship between Pavement Temperature and Weather Data."* Transportation Research Record., No.1609, 1-11

Kennedy, T.W., M. Solaimanian.(1993) "*Predicting Maximum Pavement Surface Temperature Using Maximum Air Temperature and Hourly Solar Radiation*." Transportation Research Record., No.1417, 1-11.

Lijun Sun. Lu Jia. and Jian Qin.(2006), "*Temperature Field Prediction for Asphalt Pavements"* Proceedings of the 10th International Conference on Asphalt Pavements. Quebec, Canada.

Lukanen, E.O., C. Han, and E. L. Skok, Jr. (1998) "*Probabilistic Method of Asphalt Binder Selection Based on Pavement Temperature."* Transportation Research Record., No.1609, 12-20.

RELATIONSHIP BETWEEN FRACTAL AND ROAD PERFORMANCE OF ASPHATL MIXTURE

Rui-hua YANG, Zhi-hong XU, Chao ZHANG

Dept. of Road and Airport Engineering, School of Transportation Engineering, Tongji University, Shanghai 200092,China; PH(086)021-65984218; FAX(086)021-65985891; E-mail:yangruihua1995@163.com

Abstract: The fractal model of grain size distribution, aggregate volume and aggregate pore are deduced by using the theory related to fractal geometry. The results of the road performance tests showed that the fractal dimension of aggregate volume is strongly related to air voids, dynamic stability, TSR (retained tensile strength), structure depth and the failure strains of a low temperature bending test. The influences of the fractal dimension of aggregate volume and asphalt aggregate ratio on road performances were appraised by using grey relation analysis method. The correlation model of pavement performance and fractal dimension of aggregate volume is deduced by using the correlation analysis method.
Key Word: road engineering; asphalt mixture; fractal; pavement performance

Preface

Fractal geometry has the potential to provide us with a new way to understand and analyze such natural spatial phenomena, which are not smooth, but rough and fragmented to self-similarity over a range of scales [1]. A considerable number of applications using fractal geometry have been studied in material. Fractal dimension measures the degree of irregularity based on self-similarity and provides a quantitative index to describe the self-similarity of complex phenomena. It should be a quantitative index, which can guide material design [2].

Dense-gradation asphalt mixture is designed by dense gradation principle. It is a kind of asphalt mixture that is widely used in China. The objectives of this study are: (1) to deduce the fractal models of aggregate, and (2) to constitute the quantitative relationship between fractal dimension and pavement performance of dense-gradation mixture.

1. Fractal structure of asphalt mixture aggregate

Asphalt mixture aggregates have several hierarchical fractals. The surface of the rock is self-similar fractal [3]. When the grains with different sizes are mixed together, grain size distributions become a mathematical fractal. So, the mass distribution functions and solid volume structure become fractals [4].

1.1 Fractal of grain size distribution

For a grain with diameter, suppose that the function of the grain distribution is $F(r)$, so the accumulate percentages of the grains whose diameters > r is:

$$F(r) = \frac{n(r)}{n_0} \qquad (1)$$

作者简介：杨瑞华（1977-），女，广西桂林人，工学博士研究生，E-mail：yangruihua1995◎163.com

where $F(r)$ is the function of grain distribution, $n(r)$ is the quantity of grains whose diameters $> r$, n_0 is the gross quantity of all grains in the system.

In fractal geometry, for points or line segments, the fractal model can be presented as:

$$n(r) = A_0 (r / r_{\max})^{-D} \qquad (2)$$

where A_0 is constant, D is the fractal dimension of grain size, $r_{\max}$ is maximum particle diameter.

Boundary conditions are $F(r_{\max}) = 1$ and $F(r_{\min}) = 0$, so $F(r) = \dfrac{r_{\min}^{-D} - r^{-D}}{r_{\min}^{-D} - r_{\max}^{-D}}$ (3)

And the mass distribution function of grain size can be defined as follows:

$$P(r) = \frac{m(r)}{m_0} \qquad (4)$$

where $m(r)$ is the gross mass of grains which diameters $> r$ and m_0 is the gross mass of all of the grains.

From function (4), we can get: $\mathrm{d}m(r) = m_0 \cdot \mathrm{d}P(r)$ (5)

According to the mathematical relationship of volume and mass, $\mathrm{d}m(r)$ can be also presented as follows:

$$\mathrm{d}m(r) = \rho \cdot V(r) \cdot \mathrm{d}n(r) \qquad (6)$$

where ρ is grain density; $V(r)$ is the volume of grains which diameters in the range（r，$r+\mathrm{d}r$）, and $\mathrm{d}n(r)$ is the quantity of grains which diameters in the rang（r，$r+\mathrm{d}r$）

Suppose the volume of grain is: $V(r) = K_V \cdot r^3$ (7)

where K_v is bulk form factor of grain.

From function (1) and function (3), function (8) is deduced as follows:

$$\mathrm{d}n(r) = n_0 \cdot \mathrm{d}F(r) = \frac{n_0 \cdot D \cdot r^{-1-D}}{r_{\min}^{-D} - r_{\max}^{-D}} \mathrm{d}r \qquad (8)$$

So, $$\mathrm{d}P(r) = \frac{1}{m_0} \cdot \mathrm{d}m(r) = \frac{1}{m_0} \rho \cdot K_V \cdot r^3 \cdot \frac{n_0 \cdot D \cdot r^{-1-D}}{r_{\min}^{-D} - r_{\max}^{-D}} \mathrm{d}r \qquad (9)$$

Integral calculus function (9) and make use of boundary condition $P(r_{\max})=1$ and $P(r_{\min})= 0$, the fractal model of grain size distribution can be represented in the form:

$$P(r)=\frac{r_{\min}^{3-D}-r^{3-D}}{r_{\min}^{3-D}-r_{\max}^{3-D}} \qquad (10)$$

The slop of the graph of $lgP(r)$ versus $lg(r/r_{max})$ is 3-D.

1. 2 Fractal of aggregate volume

A single grain hasa finite volume and its solid volume isn't fractal. But when the grains with different diameters are mixed together, the solid volume of aggregate is fractal because all the voids cannot be filled in completely [2].

In fractal geometry, for solid volume, the fractal model can be represented in the form:

$$V(r)=V_0(r/r_{max})^{3-Dv} \qquad (11)$$

where $V(r)$ is fractal volume, V_0 is whole volume and D_v is fractal dimension.

The aggregate fractal volume which $r \in (r, r+\mathrm{d}r)$ is:

$$\mathrm{d}V=(r/r_{max})^{3-Dv}\,\mathrm{d}V_0 \qquad (12)$$

According to the definition of volume: $\mathrm{d}V_0=M_0/\rho\times\mathrm{d}P$ (13)

where ρ is density of aggregate, M_0 is the gross mass of aggregate and P is passing percent.

Integral calculus dV, $$V=\frac{3-D}{6-D-D_V}\bullet\frac{M_0}{\rho}\bullet\frac{r_{\min}^{6-D-D_V}-r_{\max}^{6-D-D_V}}{r_{\min}^{3-D}-r_{\max}^{3-D}}\bullet r_{\max}^{D_V-3} \qquad (14)$$

where V is the fractal volume of aggregate, D is the fractal dimension of grain size distribution, D_v is fractal dimension of aggregate volume.

A fractal model of aggregate pore can be represented in the form:

$$V_{oid}=\frac{\frac{M_0}{\rho}-V}{\frac{M_0}{\rho}}=1-\frac{3-D}{6-D-D_V}\bullet\frac{r_{\min}^{6-D-D_V}-r_{\max}^{6-D-D_V}}{r_{\min}^{3-D}-r_{\max}^{3-D}}\bullet r_{\max}^{D_V-3} \qquad (15)$$

the aggregate pore can also be presented as $V_{oid}=1-\frac{\gamma}{\rho}$ (16)

where γ is the tightened density of aggregate and ρ is the bulk density of aggregate which be computed by using the bulk gravity of coarse aggregate and the apparent gravity of fine aggregate.

Therefore, when we demonstrate the tightened density and bulk density of aggregate, then we can get the fractal dimension of aggregate volume by using model (15) and model (16). It is easy to see that the fractal dimension of the aggregate volume is closely related to densities, shapes, gradation, maximum particle size and minimum particle size. So the fractal dimension D_v is a very important parameter with rich information.

2 Project of test and result

In this study, 10 asphalt mixtures (10 aggregate types with maximum particle size 19mm) will be developed，each with the same type of asphalt binder (AH-70). The gradation of the subjected mixtures and the asphalt-aggregate ratios (P_B) are shown in

Tab.1.

γ and ρ of 10 gradations aggregate are shown in Tab.2. V_{oid} can be obtained by using function (16), then D_V of 10 gradations aggregate will be obtained by using function (15) . D_V of 10 gradation aggregates are shown in Tab. 2.

Tab.1 Proportioning of asphalt mixtures

Mixture type		Screen size/mm											P_B /%
		0.075	0.15	0.3	0.6	1.18	2.36	4.75	9.5	13.2	16	19	
1	Passing Percent /%	3	5	7	11	15	24	33	59	74	89	100	2.52
2		4	6	9	13	18	29	40	63	76	90	100	3.03
3		4	7	10	15	21	32	45	67	79	91	100	3.44
4		5	8	11	17	23	36	50	69	80	91	100	3.79
5		4	6	8	12	16	25	34	60	74	89	100	3.55
6		6	9	13	18	24	37	51	70	81	92	100	5.31
7		7	10	12	17	2	33	45	66	78	91	100	5.86
8		8	11	14	19	24	36	49	69	80	91	100	6.45
9		7	9	11	15	19	28	38	62	76	89	100	5.91
10		8	11	13	18	22	32	43	65	78	90	100	6.75

Note: P_B is asphalt-aggregate ratio

Tab.2 Fractal dimensions of grain size distribution and aggregate volume of 10 aggregates

Mixture type	ρ/g.cm^{-3}	γ/g.cm^{-3}	V_{oid}/%	D	D_v
1	2.6932	1.818	32.5	2.3854	2.6904
2	2.6977	1.8113	32.9	2.4181	2.6895
3	2.7014	1.8059	33.1	2.4406	2.6983
4	2.7045	1.8013	33.4	2.4572	2.7048
5	2.6937	1.8171	32.5	2.437	2.6969
6	2.7050	1.8004	33.4	2.5074	2.7236
7	2.7008	1.8064	33.1	2.5208	2.7285
8	2.7039	1.8019	33.4	2.5375	2.7345
9	2.6963	1.8131	32.8	2.5205	2.7284
10	2.6999	1.8077	33	2.5437	2.7367

Note: ρ is the bulk densirty of aggregate and γ is tighten density of aggregate**.**

The volumetric property test was performed by using specimens compacted with 75 blows by Marshall Compactor. Then a high temperature rutting test, a freeze-thaw splitting test, a structure depth test and a low temperature bending test were performed. The results of the serials of tests are shown in Tab. 3.

Tab.3 The result of the tests

Mixture type	D_V	P_B /%	Dynamic stability /repetition .mm^{-1}	TSR/Mpa	Structure depth /mm	Failure strain /με	Air void /%
1	2.6904	2.52	1785	60.76	1.01	1995	9.6
2	2.6895	3.03	2510	57.64	0.95	2277	8.8
3	2.6983	3.44	2325	52.7	0.83	2138	8
4	2.7048	3.79	1865	56.3	0.72	2408	7.3
5	2.6969	3.55	1627	63.64	0.97	2485	6.6
6	2.7236	5.31	1034	79.37	0.69	2372	2.5
7	2.7285	5.86	731	87.19	0.62	2918	2.09
8	2.7345	6.45	750	93.88	0.55	2992	1.22
9	2.7284	5.91	904	86.98	0.79	2806	2.26
10	2.7367	6.75	247	94.53	0.52	3149	0.94

Note: The density of asphalt is 0.983g/cm^3

3 Analysis of test results

3.1 Analysis of testing results in grey relation method

The Grey relation reflects the uncertain relationship of different objects. By using the grey relation analysis method the relationship between every kind of affecting factor and the development of the system will be found. Therefore, which factor is the key factor will be known [5].

A Grey relation value reflects the degree of similarity or dissimilarity among the developing trends of different factors in the system. In the dynamic development of the system, if the changing trends of 2 factors are consistent, the grey relation value of the 2 factors is great. The Grey relation value not only signifies the contribution of comparative rank to standard rank, but also responds to the development of system dynamically [6].

Suppose the standard rank is $X_0 = \{x_0(1), x_0(2), \text{K K}, x_0(n)\}$ and comparative rank is $X_i = \{x_i(1), x_i(2), \text{KK}, x_i(n)\}$, $i = 1,2,\text{K}\ n$, the traditional formula of grey relation is :

$$\gamma(x_0(k), x_i(k)) = \frac{\Delta_{\min} + \zeta \cdot \Delta_{\max}}{\Delta_{0i}(k) + \zeta \cdot \Delta_{\max}} \quad (17)$$

Where $k = 1,2,\text{K}, n$ $i = 1,2,\text{K}\ m$; $x_0(k)$: reference rank ; $x_i(k)$: comparative rank ; $\zeta \in [0,1]$ (usually $\zeta = 0.5$) ; $\Delta(k) = |x_0(k) - x_i(k)|$, $k = 1,2,\text{K}, n$ $i = 1,2,\text{K}\ n$; $\Delta_{\min} = \min_i \min_k |x_0(k) - x_i(k)|$;

$\Delta_{\max} = \max_i \max_k |x_0(k) - x_i(k)|$

Then, the Grey relation value is: $\Gamma_{0i} = \frac{1}{k}\sum_{i=1}^{k}\gamma_{0i}(k)$ (18)

With the grey relation analysis method, we studied the affection of D_v and asphalt aggregate ratio (P_B) on the results of tests, taking air voids, dynamic stability, TSR, structure depth and failure strain as standard ranks and taking D_v and asphalt aggregate ratio as comparative ranks. The grey relation values of standard ranks and comparative ranks are shown in table 4. From table 4, it is easy to see that D_v *greatly* affects the air voids, dynamic stability, structure depth and the result of low temperature bending test ,, but asphalt-aggregate ratio greatly affects TSR . So, the fractal dimension of aggregate volume is a very important parameter which is related to performance of the asphalt mixture directly.

Tab.4 Grey relation value of standard ranks and comparative ranks

Standard rank	Air void		Dynamic stability		TSR		Structure depth		Failure strain	
Comparative rank	D_V	P_B	D_V	P_B	D_V	P_B	D_V	P_B /%	D_V	P_B
Grey relation value	0.793	0.657	0.769	0.610	0.494	0.698	0.742	0.520	0.624	0.479

3.2 Relationship between the fractal and performance of dense-gradation asphalt mixture aggregate

3.2.1 Relationship between air voids and fractal dimension of dense-gradation asphalt mixture aggregate

The fractal dimension of aggregate volume D_v gives an index of the space-filling capacity of aggregate. As the D_v increases, the space-filling capacity of aggregate will increase. The air voids of asphalt mixture should be related to D_v and asphalt-aggregate ratio. Take D_v and asphalt aggregate ratio as explanatory variables and take air voids as response variables. Model the relationship between the two explanatory variables and response variable by using multiple linear regression analysis. The regression model can be represented in the form:

$$V_a = 125.55 - 1.67 \times P_B - 41.58 \times D_v \quad R^2 = 0.98 \quad (19)$$

where V_a is air voids of asphalt mixture, P_B is asphalt aggregate ratio, D_v is fractal dimension of aggregate volume.

D_v and P_B have a strong correlation with air voids. The weighting of D_v is more than the asphalt aggregate ratio, so D_v plays a key role in determining volumetric property. From the regression model (19), we know that the air voids of mixture would decrease along with the increase in D_v. Increase in the fractal dimension of aggregate volume signifies an increase in the space-filling capacity of aggregate, so the air voids of mixture are descendant. The air voids of mixture will decrease along with the increase in asphalt-aggregate ratio. An

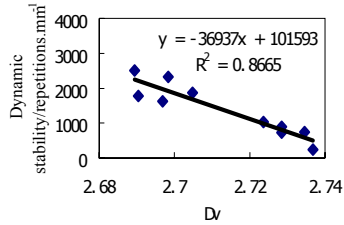
increase in the asphalt content signifies that more air voids of aggregate will be filled by asphalt, so the air voids of mixture are on a downtrend.

3.2.2 Relationship between high temperature stability and fractal dimension of dense-gradation asphalt mixture aggregate

The rutting of asphalt pavements is one of the initial damages receiving much concern. It is very important to ensure that the hot asphalt mixture has enough rutting resistance [7].

Correlation analysis of dynamic stability and D_v is shown in figure (1).

The figure (1) indicates that R^2=0.87 and D_v has a strong correlation with dynamic stability. Increase in the fractal dimension of aggregate volume signifies an increase in the space-filling capacity of the aggregate. In this condition, it is difficult to form a skeleton for aggregate and dynamic stability is on a descending trend. According to the correlation model figure (1), $D_v \nless 2.729$ if the dynamic stability $\geqslant$ 800 repetitions.mm^{-1}.



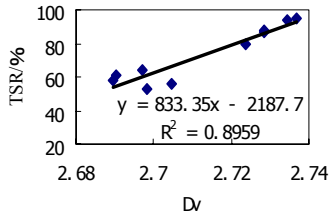


Fig.1 Correlation of dynamic stability and D_v

Fig.2Correlation of TSR and *Dv*

3.2.3 Relationship between water stability and fractal dimension of dense-grading asphalt mixture aggregate

Water damage has become the central issue of the damage in asphalt pavement. It is important to consider the water stability during the course of asphalt mixture designing [7]. A correlation analysis of TSR (retained tensile strength) and D_v are shown in fig (2).

Figure (2) indicates that R^2=0.90 and D_v has a strong correlation with TSR. TSR of mixture would increase along with the increase in D_v. Increase in the fractal dimension of aggregate volume signifies increase in the space-filling capacity of aggregate. In this condition, water tightness of mixture will be improved. According to the correlation model

figure (2), D_v ≮ 2.715 if the TSR ≥ 75%.

3.2.4 Relationship between anti-skidding performance and fractal dimension of dense-grading asphalt mixture aggregate

To satisfy the running safety of vehicles, the asphalt pavement should have enough anti-skidding performance[8]. Correlation analysis of structure depth and D_v is shown in fig (3).

The figure (3) indicates that R^2=0.8118 and D_v has strong correlation with structure depth. Structure depth of mixture would decrease along with the increase in D_v. According to the correlation model figure (3), D_v ≯ 2.739 if structure depth ≥ 0.55 mm.

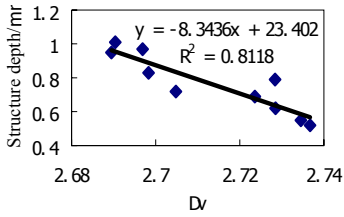


Fig.3Correlation of structure depth and D_v

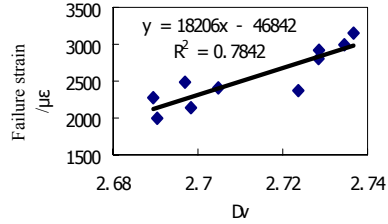


Fig.4 Correlation of failure strain and D_v

3.2.5 Relationship between low temperature anti-cracking performance and fractal dimension of dense-grading asphalt mixture aggregate

In a cold region or a region where temperatures can decrease abruptly, the asphalt pavement will crack due to temperature stress. So in this region, the anti-cracking performance of HMA is a very important factor which should be considered in design[8]. Correlation analysis of failure strain and D_v is shown in fig (4).

The figure (4) indicates that R^2=0.7842 and D_v have a strong correlation with failure strain of a low temperature bend test. Failure strain will increase along with the increase in D_v. An increase in the fractal dimension of aggregate volume signifies an increase in the space-filling capacity of aggregate and a decrease in the air voids of mixture gradually. The criticality strain energy density increases along with the increase in air voids of mixture, and it will get minimum value when V_a=10% [9]. Because V_a<10% in this research, the low temperature anti-cracking performance of HMA increases as D_v increases.

4 Discussion

Because of the limited time, some questions need further research. We offer some discussion below:

(1) The correlation analysis of pavement performance and fractal dimension of aggregate volume has certain limitations, because only one type asphalt was used in the

research. In fact, asphalt is the key factor affecting pavement performance. The performances of asphalt mixture with different grade asphalts or different modifiers are not the same.

(2) Although we used 10 mixtures in this research, the maximum particle size of the mixtures are the same (19mm). We need to do more to develop the relationship among fractal and performances of other kinds of mixtures.

5 Conclusions

The fractal models of aggregate are deduced by using the theory related to fractal geometry. Through research on the relationship between fractal characteristics and performance of dense-gradation asphalt mixture, some initial conclusions were drawn:

(1) Model (10) (fractal model of grain size distribution), model (14) (fractal model of aggregate volume) and model (15) (fractal model of aggregate pore) for mixture aggregate are deduced.

(2) The fractal dimension of aggregate volume gives an index of the space-filling capacity of aggregate and it is a very important parameter with rich information. Analysis of the testing results in the grey relation method indicated that the fractal dimension of aggregate volume is related to the pavement performance closely.

(3) The air voids, high temperature stability and anti-skidding performance of asphalt mixture decrease and the water stability and low temperature anti-cracking performance increase along with the increase in the fractal dimension of aggregate volume. The performances of the asphalt mixture with different fractal dimension of aggregate volume can be estimated according to the correlation analysis of pavement performance and fractal dimension of aggregate volume.

***References*:**

[1] Liu S.D, Liu S S. An introduction to fractal and fractal dimension [M]. Beijing: Aero logical Press,1993 (in Chinese)

[2] Li Guoqiang, Den Xuejun. The effect of gradation aggregate [J].Concrete, 1995,(1): 3-7 (in Chinese)

[3] Tang Ming. Research on fractal characteristics of concrete void [J]. Concrete, 2000, (8):3-5 (in Chinese)

[4]Xia Chun, Liu Haowu. Research on Fractal Characteristics of the Size –Distribution of Concrete Aggregates [J].Journal of Southwest Communication University, 2002,2(37):186-189 (in Chinese)

[5] Deng Julong. Grey Forecasting and Decision [M]. Wuhan: Huazhong University of Science and Technology Press, 2002. (in Chinese)

[6]Yi Desheng, Guo Ping. Grey Theory and Methods, Beijing: Petroleum Industry Publishing House, 1992 (in Chinese)
[7] Sha Qinglin. Study on Expressway Bitumen Pavement Initial Damage and Preventing [M]. Beijing: People's Communications Press, 2001 (in Chinese)
[8] LV Weimin. Design principle and methods for asphalt mixture [M]. Shanghai: Tongji University Press, 2001. (in Chinese)
[9] Hao Peiwen. Research on mixture proportion design of the asphalt concrete [D].XI'AN: XI'AN HIGHWAY UNIVERSITY Ph D Theses, 1996 (in Chinese)

About the Author:

XU Zhi-hong: professor, Address: school of Transportation Engineering, Tongji Univ. Shanghai, China, E-mail: xuzhh1939@163.com
YANG Rui-hua , engineering doctoral student, Tongji University, E-mail: yangruihua1995@163.com
ZHANG Chao, engineering doctoral student, Tongji University

DCP FOR LOW VOLUME ROAD DESIGN/CONSTRUCTION

Shie-Shin Wu

IES Consultants, 106 Ivywood Ln., Cary, NC 27518; PH (919)-852-0185; email; shinwu@bellsouth.net

Abstract: Due to its low impact on overall highway system in case of failure, low volume road design and construction acceptance are very often ignored by highway agencies. The negligence more often than not is due to lack of staff to handle these "minor" projects. From a macro highway system point of view, impact of low volume road failure is not significant. But for users who depend on this system to support their daily life, a failure is detrimental. There is a need to develop a simple, easy to follow, and reliable procedure to design and accept low volume roads. Data collected using a Dynamic Cone Penetrometer (DCP) on several construction projects show that such a procedure is possible. A DCP base low volume road design and construction acceptance procedure is proposed. This procedure is simple enough that a reliable road can be designed and constructed under the supervision of a junior engineer.

(**Key words:** *DCP, low-volume road, design, construction acceptance*)

1 Introduction

Proper design and construction of low volume roads are often neglected. Because failure of such projects has low impact on overall highway system performance, when the agency in charge is overwhelmed by projects, it is logical to pay less attention to these minor projects. But, for road users, the one road that serves them day in and day out is the most important road, regardless of daily traffic volume. The logic behind the highway agency's practice is understandable, but unacceptable. There is an urgent need to develop a design and construction procedure that is less demand on engineer's expertise and time consumption.

A recently completed research project to investigate the possibility of using Dynamic Cone Penetrometer (DCP) for construction acceptance in Ohio found that DCP is a viable device in construction quality assurance (Wu and Sargand, 2006). Data collected from ten (10) construction projects in Ohio during the construction season of 2003 and 2004 indicated that this device can also be a useful tool in low volume road design and construction.

2 The Dynamic Cone Penetrometer (DCP)

The DCP was developed in South Africa for evaluation of in-situ pavement strength or stiffness in the 1960s. Dr. D. J. van Vuuren designed the original DCP with a 30° cone (van Vuuren, 1969). The DCP was used to investigate road pavement (Kleyn, 1975). In 1982, Kleyn described another DCP design, which used a 60° cone tip, 8 kg (17.6 lb) hammer, and 575 mm (22.6 in) free fall (Kleyn, 1982). This design was then gradually adopted by countries around the globe. In 2003, the ASTM D6951-

03 Standard Test Method for Use of the Dynamic Cone Penetrometer in Shallow Pavement Applications described using a DCP with this latest design (ASTM, 2003).

DCP testing consists of using the DCP's free-falling hammer to strike the cone, causing the cone to penetrate the base or subgrade soil, and then measuring the rate of penetration per blow (penetration rate, PR, in mm/blow). This measurement denotes the stiffness of the tested material, with a smaller PR number indicating a stiffer material. In other words, the PR is a measurement of the *penetrability* of the subgrade soil.

In the early stages of DCP development, researchers tended to concentrate their efforts on correlating DCP readings to commonly accepted strength or stiffness parameters, such as CBR (US Army Corp of Engineer, North Carolina Department of Transportation, et al.), resilient modulus (Kansas Department of Transportation, Transport and Road Research Lab, U.S. Army Corps of Engineers, Danish Road Laboratory, et al.), unconfined compressive strength (De Villiers,1980) UCS values and Young's modulus (Texas Department of Transportation).

To assess in-situ test methods, Abu-Farsakh et al. (Abu-Farsakh, Alshibli, Nazzal, and Seyman, 2004) developed the correlations between the DCP data and Static Plate Load (SPL) test, Falling Weight Deflectometer (FWD) test, and CBR test data collected in the field. They found good correlations between the PR and both the initial modulus and the reloading stiffness of the SPL test. They also found that the values calculated using DCP results are more consistent and correct than values calculated based on data from either a Geogauge or a Light Falling Weight Deflectometer (LFWD). This study showed that the DCP readings correlate better with CBR values than data gathered using the other two devices.

All these studies concluded that the DCP is an effective tool for evaluation of layers strength and can take deeper measurements than the other devices. Therefore, DCP can be used to profile in-situ CBR values or the modulus of the base and subgrade.

The purpose of such correlation was to prove the validity of the DCP as a soil stiffness measurement device. Converting DCP data to a commonly accepted parameter also enabled the incorporation of DCP data into an established pavement design procedure. Such conversion, which helped users understand and accept DCP, does have historical value. Development of the relationship between DSN800 and pavement performance by Kleyn (Kleyn and Savage, 1982) pioneered the application of DCP measurement to pavement design. It is now time to accept DCP into the mainstream.

It is important to understand that no matter how sound the correlation, estimation errors are unavoidable. If we agree to accept the use of the DCP to measure the stiffness of unbound material (and in some cases, bound material such as a thin AC layer), then it is logical to accept the use of DCP readings, that is, PR (in mm/blow), in practical application. This approach makes field operation and implementation much easier. Therefore in this study the researchers used PR values gathered using DCP sounding and developed acceptance criteria in terms of DCP measurements instead of older measures.

3 Field Test Results

After grading operations were completed and the subgrade was finished, DCP tests were performed. Testing could be stopped when penetration depth reached 1 m or upon refusal. During the two-year period of this study, ten projects were tested. The number of samples taken depended on the length of finished subgrade available for testing when the DCP testing team was on site. Table 1 summarizes the projects tested.

Table 1. Summary of Tested Projects

Project	Sample ID	Core	Tested on	No of Samples
Chestnut	1 to 23		5cm AC on 30cm CT	23
Chestnut	24 to 26		natural soil	3
DEL23	1 to 6		30cm CT	6
DEL23CT	1 to 33	core hole	30cm CT	31
ERI02	A1 to A6	core hole	30cm CT	6
ERI02	B1 to B6	core hole	natural soil	5
ERI02	C1 to C6	core hole	30cm LCT	4
HAM126	1b to 6b	core hole	natural soil	6
Livingston	1 to 6		30cm LT	6
LOG33	A1 to A6	core hole	30cm CT	6
SR35	1654 to 1698		natural soil	10
US30	1 to 21		natural soil	18
US50	E1 to E17		10cm NJ OGGB and 15cm Ohio304 DGGB	17
US50	W1 to W17		10cm Iowa OGGB and 15cm Ohio304 DGGB	17

3.1 Nature Soil

The ODOT construction specification considers the soil layer from the top of the subgrade to a depth of 300 mm to be the subgrade. According to the specification, requirements for subgrade construction, such as material and quality of compaction, are more stringent than the soil below (Ohio DOT, 2002).

DCP tests were performed on five projects in which the subgrade was constructed with natural soil. Most of these test results show a clear layer interface at around 300 mm below the subgrade surface. This is exactly the bottom of the subgrade layer (Figure 1). These results indicate that subgrade layer construction achieved the goal of the specification. But many samples indicated that the high quality layer does not reach the design depth (300 mm), i.e., they fail to meet the specification requirement.

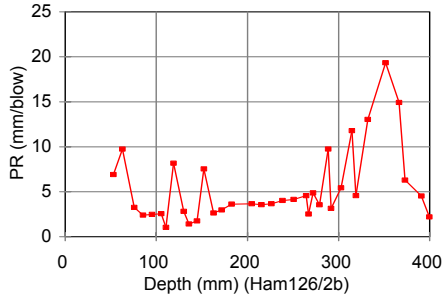


Figure 1. A Good Subgrade Layer

3.2 Stabilized Soil

Test results indicate that, except for the Chestnut project, the average PRs are quite close. Student's t-test independent samples testing was used to test the null hypothesis that the difference of the two means is zero. The results showed that PRs from the Chestnut project are significantly different from PRs from all other projects. For the rest of the project pairs, the LCT data is significantly different from only one set of CT data. The student's t-test results indicated that data sets from the different treatment methods are not significantly different at the 90 percent level. In other words, despite the different material used to stabilize soils in these projects (cement or lime or a combination of the two), DCP test results taken from these projects can be considered to have been taken from the same population; that is, the data sets can be pooled into one.

Samples from some of the test locations showed a homogeneously low-PR (stiff) layer approximately 300 mm thick. Figure 2 is an example of this good construction quality measured at the Livingston project.

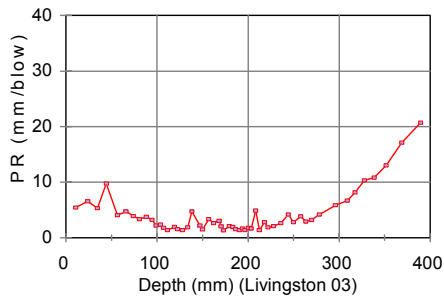


Figure 2. A Good Stabilized Soil Layer

Many other samples showed a stiff layer much less than 300 mm thick followed by a gradually increasing PR. Figure 3 is a plot typical of these cases. This plot shows

a good quality stabilized layer to a depth of 150 mm followed by a gradually increasing PR to a depth of 300 mm, which indicates a poorly constructed stabilized subgrade. The depth of this stronger layer reaches is defined as *effective thickness*.

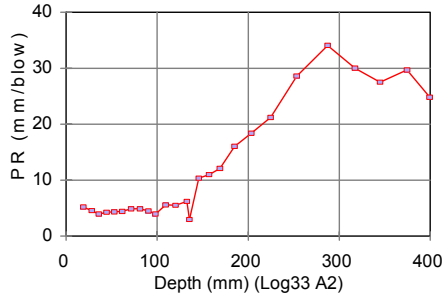


Figure 3. A Poorly Constructed Stabilized Soil

The average PR of the pooled treated soil (excluding the Chestnut project) is 3.8 mm/blow, and a standard deviation of 2.5 mm/blow results in a high coefficient of variation (66 percent). Figure 4 shows that the 95th percentile is 8 mm/blow. This study's DCP data indicated that regardless of types of treatment, with 95 percent certainty, one can expect the treated soil to have a PR less than or equal to 8 mm/blow.

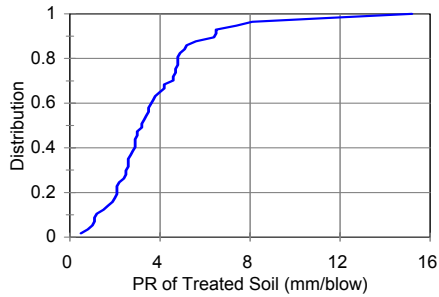


Figure 4. Distribution of Stabilized Soil

Figure 5 is a plot of the distribution of the effective thickness. This figure shows 80 percent of the samples taken; the effective thickness is less than the designed thickness, which is 300 mm. In other word, most of the stabilized soil did not achieve the design target.

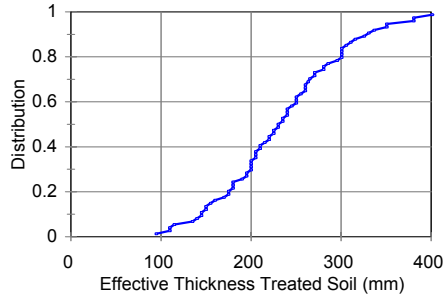


Figure 5. Distribution of Effective Thickness

3.3 Gravel Base

To corroborate the U.S. Route 50 data, DCP test data for the eastbound and westbound Ohio 304 Dense Graded Gravel Base (DGGB) were compared using the Student's t-test. Results showed that these two groups of data are not significantly different at the 95 percent confidence level (t = -2.0, df = 32). Therefore, the two sets of data are grouped in one. The average PR for DGGB is 5.35 mm/blow. Figure 6 is a plot of the PR distribution for the Ohio 304 DGGB. The 95th percentile PR is 8 mm/blow.

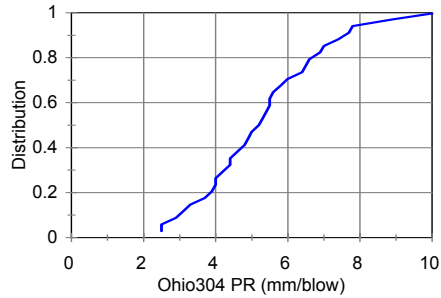


Figure 6. Distribution of DGGB

3.4 AC Surface

Part of the Chestnut project was constructed with a 50 mm AC surface on 300 mm cement-stabilized soil. The rest was constructed on untreated natural soil. The DCP can penetrate through a thin layer of asphalt surface with little effort. In this project, most of the PR values for the AC surface course were concentrated in the range of 2 to 7 mm/blow. The average PR for the AC surface course is 5.2 mm/blow and a standard deviation of 2.9 mm/blow. Figure 7 is a scatter plot of the stiffness of the surface course versus that of the CT soil. It is interesting to point out that a few AC

surface course PR readings are far outside the normal range of AC surface course readings. These outlying data points coincide with a weak substrate. Weaker (less stiff) AC can be the result of poor compaction.

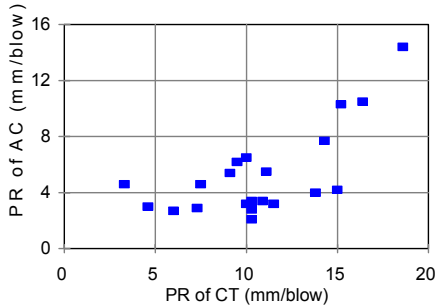


Figure 7. PR of Asphalt Course vs. Cement Treated Subgrade

4 Low Volume Road Design/Construction Procedure

The experience from this research project showed that DCP can be used to penetrate a thin layer of asphalt surface course with little effort. Data show that PR reading on AC is very consistent. Therefore, using DCP to evaluate *in situ* thin asphalt layer is practical. Data from gravel base, stabilized and natural subgrade indicated that DCP not only can provide construction personnel with surface condition but also the condition of the *in situ* base and subgrade. DCP is a low cost capital investment and a simple device. The limited data collected from this study suggests that development of a DCP base low volume road design/construction procedure is possible.

A road is considered low volume if 1) low daily traffic 2) very few heavy truck traffic and therefore, 3) pavement failure is primary due to environmental effect. In this case, pavement structure, which includes the surface layer and, if exist, the base and/or stabilized soil, is primary for dust control and water protection and secondary for load distribution. Design of this type of road can be done by a catalog design procedure. According to projected traffic, especially heavy trucks, a pavement structure can be selected from the catalog. Some information (or knowledge) of local soil is helpful but not necessary. This is all it is needed for pavement design. The real issue in this procedure is how to control construction quality?

During construction, field personnel must verify the quality of every layer constructed, from the soil and up. There is no need for special quality control equipment except one manual DCP. Layer strength standard is list in the following sections. A substandard layer must be improved and meet the standard before the subsequent layer construction is allowed. AC quality control will rely on hot mix plan to produce AC material that meets mix design requirement. Field personnel will use DCP to assure the *in situ* quality.

4.1 Natural Soil

Data collected showed that the minimum PR of a thin AC layer can be achieved only when subgrade PR is less than 12 mm/blow. So for project when AC is paved on top of natural soil, the top 300mm soil must meet this requirement. Or else, apply undercut, lime and/or cement stabilization etc. techniques to improve subgrade strength.

4.2 Stabilized Soil

This study found that the 95th percentile PR for stabilized soil is 8 mm/blow. This result indicated that under the current Ohio soil stabilization specification, a properly constructed stabilized subgrade is expected to be 8 mm/blow or less.

4.3 Gravel Base

The 95th percentile PR of the DGGB is 8 mm/blow. This value is derived from Ohio data. Highway agencies must develop their own standard base PR on local material.

4.4 Asphalt Surface

The acceptable PR for a thin AC surface course on a low-volume road shall be set at 7 mm/blow.

4.5 Advantage of the Procedure

Advantages of adopting this design/construction procedure are:

1. do not require higher level engineer and technician
2. do not require expensive equipment to perform the needed tests
3. uniformity of the often neglected project can be improved
4. location of inferior material can be identified, hence can be corrected prior to subsequent construction.

5 Conclusions and Recommendations

The low column road pavement design/construction procedure proposed in the paper is a practical, workable and simple procedure. It can be applied to any remote region. The acceptance standard is based on data collected from Ohio. Like all engineering experience formula, highway agencies shall verify these numbers and adjust according to local condition.

The conclusions are based on 10 projects tested over a two-year period. The suggested standards are reasonably achievable under the current specification and practice. They may not be the optimal target values for these parameters. More DCP test data is needed to formulate a good standard. It is strongly recommended that highway agencies interesting in this procedure shell implement DCP testing on construction projects. Data collected can be utilized to improve the standard recommended by this study.

Acknowledgement

The author wants to express his thanks to Ohio Department of Transportation for sponsoring the initial DCP project and the staff in Ohio Research Institution for Transportation and the Environment for collecting DCP data.

References

Wu, ShieShin and Sargand, Shad, *Use of Dynamic Cone Penetrometer in Subgrade and Base Acceptance,* State Job Number 14817(0), Ohio Research Institute for Transportation and the Environment, Ohio University, April 2006

Kleyn, E. G. *The Use of the Dynamic Cone Penetrometer (DCP),* Transvaal Roads Department, South Africa, July, 1975.

Kleyn, E. G., Maree, J. H., and Savage, P. F. "The Application of a Portable Pavement Dynamic Cone Penetrometer to Determine in situ Bearing Properties of Road Pavement Layers and Subgrades in South Africa." The European Symposium on Penetration Testing, Amsterdam, Netherlands, May, 1982.

van Vuuren, D. J. "Rapid Determination of CBR with the Portable Dynamic Cone Penetrometer," *The Rhodesian Engineer*, September, 1969.

ASTM, *Standard Test Method for Use of the Dynamic Cone Penetrometer in Shallow Pavement Applications,* Publication number D6951-03, ASTM, Washington, DC, 2003.

Abu-Farsakh, M. Y., Alshibli, K., Nazzal, M. D., and Seyman, E. *Assessment of In-situ Testing Technology for Construction Control of Base Course and Embankments,* LTRC Project No. 02-1GT, Louisiana Department of Transportation and Development, May, 2004.

De Villiers, P. J. *Dynamic Cone Penetrometer Correlation with Unconfined Compressive Strength.* University of Pretoria, Pretoria, South Africa, 1980

Ohio DOT, *The 2002 Construction and Material Specification,* Office of Construction Administration, Ohio DOT, Columbus, OH, 2002.

LABORATORY STUDY ON ULTRAVIOLET AGING BEHAVIOR OF ASPHALT BINDER AND MIXTURE

TAN Yiqiu[1]，WANG Jiani[2]，XUE Zhongjun[3]，FENG Zhongliang[4] and XU Huining[5]

[1]School of Traffic Science and Engineering, Harbin Institute of Technology, Harbin, Heilongjiang, China, yiqiutan@163.com
[2]School of Traffic Science and Engineering, Harbin Institute of Technology., Harbin, Heilongjiang, China, wangjiani3328@126.com
[3]School of Traffic Science and Engineering, Harbin Institute of Technology, Harbin, Heilongjiang, China, xuezhongjun@126.com
[4]Jiangsu Transportation Research Institute, Nanjing, Jiangsu, China
[5]School of Traffic Science and Engineering, Harbin Institute of Technology., Harbin, Heilongjiang, China

Abstract: As a type of polymer material, bitumen ages easily under ultraviolet (UV) radiation conditions. However, the effect of UV aging is still not very clear. This paper discusses the behavior of UV aging on asphalt performance of five binders and mixtures by experiment. The results show that binders' physical aging behavior is highly consistent with its rheological properties; the low-temperature flow deforming ability of asphalt degraded remarkably after UV aging. There are close correlations between the mixture's low-temperature deforming ability and the binder's stiffness, and the mixture moisture stability evaluation index (i.e. TSR) and binder's penetration, respectively. In addition, the paper points out that binders' aging behaviors can be well explained with its chemical character: aromatics and resin are the two decisive chemical components of a binder for its anti-ultraviolet aging ability.
Key words: *age, ultraviolet aging, rheological properties, anti-ultraviolet aging*

1 Introduction

Because of the difference in air composition and length of a solar radiation path between plateaus and low latitude plains, plateaus have a unique climate and environmental features due to their more intensive radiation strength, longer daylight time, and especially the much higher percentage of UV radiation content, which, ranging from 20% to 25% of the total solar light energy, is five times that in plains (XIAO, 2003). Such intensive UV radiation aging causes serious degradation on asphalt performance and attenuates the pavement durability. But the current asphalt performance evaluation system gives little consideration on UV radiation aging.

UV aging and thermal aging are two quite different types of aging, and the different influences of UV aging were compared with thermal aging (YE et al. 2006). However, still very little research has been reported on asphalt's UV aging behavior.

This paper studied the UV aging behavior of asphalt binders and mixture caused by UV radiation. Therefore, it is necessary and important to carry out corresponding studies specifically on UV aging. This study investigates the influence of UV aging on asphalt performance.

2 Experimental Methodology

2.1 Artificial Aging Procedures

The rolling thin-film oven (RTFO) was used in this work for thermal aging; the UV radiation equipment mentioned in (Tan, 2003). The UV aging test was at 73℃ for 9 hours, so that the radiation energy is equivalent to the total energy received outdoors in 5 months in Lasa. The asphalt mixture UV aging test used similar equipment and procedures: the mixture was tested in a "semi-compacted state' at 73℃ with air-blowing for 48 hours.

2.2 Basic Physical Properties of Binders

The binders used in this study included three basic binders (denoted as KL, LH, and XT, respectively) from different sources with the same penetration grade and two polymer-modified asphalt binders(K-SBS、K-SBR). Each binder listed in the following text was assigned an identification code which provides information of the source and physical state of the material. E.g. KL-O means an original asphalt sample from KL; K-SBS-T means a thermal aged K-SBS sample. Binders' basic physical properties are summarized in table 1 and figure 1.

Table 1: Physical Indices of Original & Aged Binders

Binder variety		KL	LH	XT	K-SBS	K-SBR
Penetration (25°C; 0.1mm)	Original bitumen (Po)	79	82	81	96	99
	Thermal aged (Pt)	60	74	54	78	82
	UV aged (Pu)	46	51	49	69	68
Ductility (10°C;cm)	Original bitumen (Do)	52.7	74.6	45.3	167.5	165
	Thermal aged (Dt)	36	36.2	29.8	153.5	123.5
	UV aged (Du)	11.8	30	23.1	139	113.5
Softening point (°C)	Original bitumen (So)	46.3	42.6	45.9	61.9	71.7
	Thermal aged (St)	48.1	47.3	50	60.7	66
	UV aged (Su)	56.7	48.0	54.2	64.9	67.3

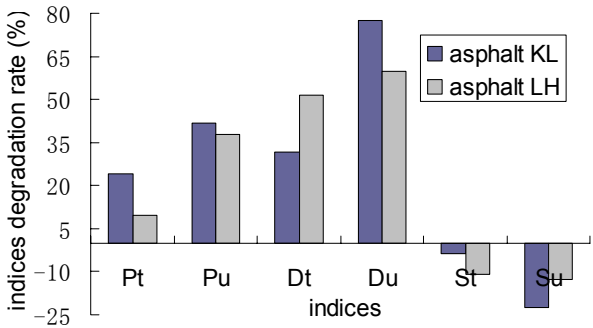


Figure 1 Physical Indices Degradation Rates of Two Aged Binders

2.3 Bituminous Binder Experiments

The DSR test was conduced with a Gemini-150 rheometer in a parallel-plate configuration with a gap of 2000μm at 0℃; The stress was 1.5×104Pa at a frequency of 0.1rad/s. The AASHTO MP5 method was adopted in the BBR test; the temperature was -12°C. The ASTM D-4124 method of Corbett was used to separate the original and aged bitumen samples for chemical component study. The rheological indices and chemical components before and after aging were shown in table 2, table3, and table 4.

Table 2 Loss Modulus (MPa) & Phase Angle (°)

Binder variety	KL		LH		XT		K-SBS		K-SBR	
	δ	***G″***	***δ***	***G″***	***δ***	***G″***	***δ***	***G″***	***δ***	***G″***
Original binder	20.3	1.19	13.1	0.97	13.2	0.96	30.8	1.30	32.1	1.32
Thermal aging	18.9	1.03	12.7	0.99	11.2	0.85	30.3	1.30	31.7	1.37
UV aging	13.8	0.8	12.0	0.92	10.0	0.74	28.4	1.21	18.9	1.18

Table 3 Stiffness (S) & m Values (m)

Binder variety	KL		LH		XT		K-SBS		K-SBR	
	S	***m***	***S***	***m***	***S***	***m***	***S***	***m***	***S***	***m***
Original binder	45.8	0.54	120	0.5	76.6	0.39	90.3*	0.42*	119*	0.44*
Thermal aging	62.5	0.9	146	0.45	112	0.43	99*	0.41*	96.1*	0.48*
UV aging	70.3	0.47	152	0.45	141	0.39	30.8	0.49	52.3	0.39

Remark: * test temperature is -18°C

Table 4 Chemical Components of Original and Aged Binders

Chemical components	KL-O	KL-T	KL-U	LH-O	LH-T	LH-U	XT-O	XT-T	XT-U
Saturates (symbol : S)	31.5	30.0	29.9	35.9	34.7	34.6	23	19	18.7
Aromatic (symbol : Ar)	33.6	24.5	23.2	35.1	30.9	30.7	45.3	43.5	44.6
Resin (symbol : R)	33.5	43.6	44.9	24.3	27.4	28.1	20.5	25.8	23.7
Asphaltene (symbol: At)	1.5	1.8	2.1	4.7	6.9	6.6	10.3	11.7	13

2.4 Asphalt Mixture Experiment

To exclude influences from unrelated factors, the crucial mixture indices such as gradation and air void were maintained in all tests. An AC-13Ⅱ gradation was adopted for these mixtures. The asphalt mixtures' low-temperature flow deforming ability was evaluated by indirect creep tests. The prism specimen size was 250mm×30mm×35mm; the fixed stress in this test was about 10% of the failure load; it lasted for 20~30mins at 0℃. The AASHTO T-283 test was adopted in this experiment. The moisture stability of the asphalt mixtures was evaluated on the tensile strength ratio. The experimental results are listed in table 5 and table 6, respectively.

Table 5 Losing of Creep Rate of UV Aged Asphalt Mixtures

Asphalt variety		KL	LH	XT	K-SBS	K-SBR
Creep Rate (1 /s /MPa)	Thermal aging	2.00E-06	1.29E-06	2.86E-06	7.79E-06	3.50E-06
	UV aging	2.89E-06	7.82E-07	1.81E-06	6.53E-06	2.04E-06
Losing of creep rate (%)		-44.5	39.5	36.8	16.2	41.9

Table 6 Tensile strength ratio (TSR) of UV Aged Asphalt Mixtures

Asphalt variety	KL	LH	XT	K-SBS	K-SBR
TSR of un-UV aging (%)	83.8	81.5	87.3	93.1	76.7
TSR of UV aging (%)	76.8	90.7	79.5	86.1	82.7

3 Results and Discussion

3.1 Influences on Binder Properties

(1) Changes of physical properties

Table 2 presents the physical indices of the original, thermal and UV aged binders. After UV aging, the values of penetration and ductility decreased while softening points increased. Fig. 1 shows the changes of the physical indexes of bitumen KL and LH. KL is more UV radiation sensitive because larger physical indices changes can be seen, showing a poor ability of anti-ultraviolet. On the

contrary, LH is not very sensitive to UV, as there were only small changes by UV aging.

(2) Changes of rheological properties

This study applied the DSR and BBR tests to investigate the influence on rheological properties by UV radiation. Rheological properties can be well reflected by the loss modulus G″ and phase angle δ. The energy dissipated as flowing deformation to resist loading is expressed as the loss modulus G″, and δ is an index for the ratio of viscosity and elasticity. As degradation of rheological properties, lower values of G″ and δ are undesired. Reductions of these two indices reveal that UV aging changes the viscoelastic character of asphalt, improves elasticity, and decreases viscosity (in Table 3). As a result, the low-temperature flow property drops.

Two BBR indices, the stiffness, S, and the slope m-value, m, at 60 sec, were used to evaluate the low temperature performance. The results are shown in table 4, where the effects of UV aging are clearly exhibited by the variations of S and m. It can be observed that S increased and m decreased during UV aging. The variation tendencies of S and m imply that rheological properties fall after UV aging.

The above experiment results displayed that UV aging gave noticeable influences on the binders' physical and rheological properties. Because of the decrease in rheological properties, low-temperature flow deforming ability falls significantly.

(3) Influences comparison of UV and thermal aging

To compare the influences of thermal and UV aging on the properties of bitumen, the penetration and loss modulus of binders were listed in order as shown in Figure 2, where it can be seen that the changes of the physical and rheological properties after UV aging are not consistent with those after thermal. It clearly indicates some noticeable differences in their aging properties.

(4) Relationship between UV aging behaviors and chemical components

The transformation trend between components can be briefly expressed as: aromatic transforms to resin, and resin transforms to asphaltene at the same time (Siddiqui, 1999). As a product of the aging reaction, the asphaltene content is much higher after UV aging than that after thermal (in table 5), indicating some difference in the reaction extent. The asphaltene content of KL and XT after UV aging was higher than that after thermal aging, with an increase of 40% and 26.2%, respectively. Yet, LH shows an opposite aging effect. The variation trend of the asphaltene amount reflects asphalts' sensitivities to thermal and UV radiation, consistent with the changes of physical and rheological properties. So it is feasible to estimate the extent of the aging reaction by observing the conversion of asphaltene's content.

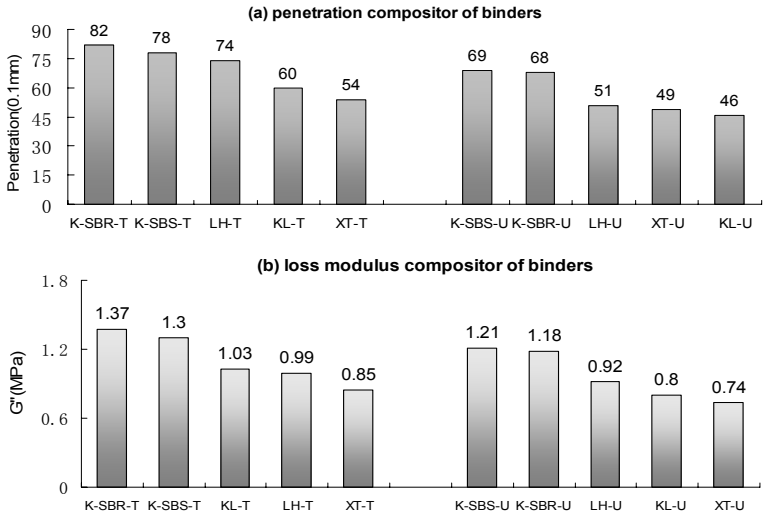


Figure 2 Physical and Rheological Indices Compositor of Thermal and UV Aged Bitumen

3.2 Influences on Asphalt Mixture Performance

(1) Influences on low-temperature performance and moisture stability

As low-temperature performance is mainly binder dependent (Julie,1996), this paper studies influences of UV aging on low-temperature performance of an asphalt mixture with the creep test. From the test results shown in table 6, a falling trend of the creep rate can be seen on most of the asphalt mixtures, displaying a degradation of the anti-cracking ability.

Moisture stability is another aspect of asphalt performance influenced greatly by UV aging due to the changes in consistency and viscosity. The results of the AASHTO T-283 test show a decrease in TSR (and thus the moisture stability) in four asphalt mixtures (in table 5) and the decrease also reveals a strong binder dependency. The moisture stability is consistent with the binder aging properties.

(2) Relationship Between Aged Asphalt Mixture and Binder

To investigate the relationship of the aging behaviors of asphalt binder and mixture, a regression curve was established in figure 3. The curve shows a noticeable linear correlation between S and the creep rate loss, indicating a clear relationship between the low-temperature rheological performances of mixture and binder after UV aging; the curve of the modified binder is very similar to that of a basic binder.

The correlation analysis of the binder's physical-chemical properties and TSR was conducted. As a conditional viscosity index, penetration can also reflect the binder's consistency. Since moisture stability is adhesion determined to some extent, changing of penetration affects the TSR directly; this is shown clearly in the regression curve in figure 4.

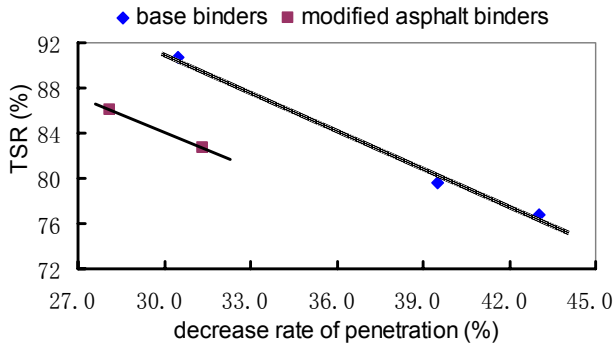


Figure 3 TSR of Mixtures vs. Decreased Rate of Penetration

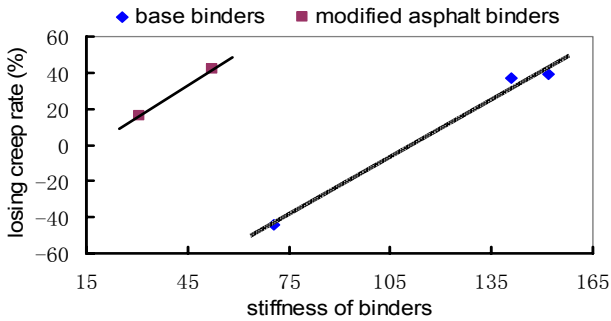


Figure 4 Losing Creep Rate of Mixtures vs. Stiffness of Binders

4 Conclusions

This paper presents the study of the influence of UV aging on binder and asphalt mixture performance by experiments, comparing sensitivities to UV radiation and thermal and observing changing of asphalt mixture performance during UV aging. Based on these investigations, some conclusions can be drawn:

(1) UV radiation aging has obvious effects on the performance of both bitumen and asphalt mixture, especially bitumen's low-temperature flow deforming ability and thus the mixtures' anti-cracking ability.

(2) The transformations taking place in chemical components during UV aging are related to the changes of the properties. Most bitumen has different sensitivities to thermal and UV radiation. The evaluation method based on thermal aged asphalt does not truly reflect influences of UV aging, and there will be some limitations on the substitution of UV by thermal aging.

(3) For the same asphalt, the binder's physical aging behavior is highly consistent with its rheological properties. In addition, there are close correlations between the mixture's low-temperature deforming ability and binder's stiffness; and the moisture stability evaluation index, TSR, and the penetration, respectively.

5 Acknowledgments

The authors gratefully acknowledge the Ministry of Education of P.R.China for the financial support of this project. The authors also thank Xinjiang Transportation and Communication Department for their laboratory assistance.

References

Huang, Baoshan, Louay N. Mohammad, G. Wije Wathugala (2004), "Application of a Temperature Dependent Viscoplastic Hierarchical Single Surface Model for Asphalt Mixtures," ASCE Journal of Materials in Civil Engineering 16:2 147~154.

Julie E. Kliewer, Huayang Zeng, Ted S. Vinson (1996), "Aging and low-temperature cracking of asphalt concrete mixture," journal of cold regions engineering. p.134~148.

M.N. Siddiqui, M.F. Ali(1999), "Studies on the aging behavior of the Arabian asphalts," Fuel.78 1005–1015.

Tan, Yiqiu, Feng Zhongliang, Zhou Xiingye etc. (2003), "The Evaluation Method Study of Anti-ultraviolet Aging Ability of Asphalt," Science paper Online

XIAO, Yide, WANG Guangyong; LI Xiaogang,(2003), Corrosion Behavior of Atmospheric Environment and Corrosion Feature of Materials in Our Western Area. Journal of Chinese Society For Corrosion and Protection. Vol.23 (4) p.248-255

Ye, Fen, Jin Yang, Peng Huang (2006) Performance of Modified Asphalt Aging under Ultraviolet Radiation ASCE Pavement Mechanics and Performance p.102.

Reasonable Spacing Model of Chevrons

Shunying ZHU[1], Wuxiong PENG[2], Qinmei SUN,
Yuan XU, Juxiang GUAN, Hongli DONG

[1] Professor, Dept. of Traffic Engineering, Wuhan University of Technology, 1040 Heping AVE, Wuchang, Wuhan, Hubei, China. 430063. E-mail: zhusy2001@163.com.
[2] Master Student, Dept. of Traffic Engineering, Wuhan University of Technology, 1040 Heping AVE, Wuchang, Wuhan, Hubei, China. 430063. E-mail: pengwuxiong@163.com.

Abstract: In order to improve the shortage in calculating principles and methods of chevron spacing, and to increase the dependability for four visible chevrons and the accuracy of road curve radius perception, the authors have concentrated on six basic curves under different design speeds according to the Design Specification for Highway Alignment in China. They have taken into account the five factors including the angle and distance of acute cone vision, scattering angle and distance of the vehicle lamp, chevron legible distance, driver's eye focus location, and deceleration distance when a vehicle approaches the curve at a high speed. The goal is that four chevrons can be seen at the same time at the edge of the median of a dual two-lane highway. The five characteristic points of the basic curves are set as the chevron setting jumping-off point. The model of chevron spacing is established according to the curve's integral characteristics and by the tangent off-set method. Then with the minimum as the reasonable spacing, and based on the reasonable setting space of the basic type of curve under different design speeds, the ultimate chevron spacing model is achieved. In addition, this paper discusses assumptions about problems about curve inducement in multi-lane, and dependability for four visible chevrons at the same time, and so on.
Key words: *Chevron alignment sign; Spacing; Safety*

1 Background and research objectives

Because of landform, geology and project price limits, road alignment always uses low technique guidelines at local sections, so it induces sharp changes of curves. But driver's visual capability and speeds tend to be adaptive and inertial. It is difficult for drivers to discover the sharpness of the curve and change speed in time, which results in alignment problems, and certain local sections have become notorious for traffic accidents. Chevrons can improve the drivers' visual information, make them communicate sufficiently with the road line type, perceive the curve direction, curvature and operational clues, foretell or alarm drivers about the curve variety in front of them on the road so that they can take measures earlier and improve driving

security. Therefore it is very important to place chevrons reasonably. However, until now there has been little specific prescription about how to set up chevrons and how many should be set up. The Manual on Uniform Traffic Control Devices, PART 2.SIGNS, Chapter 2C.Warning Signs, 2C.10 gives guidance: spacing of Chevron Alignment signs: they should be such that the road user always has at least two in view, until the change in alignment eliminates the need for the signs. China Road Traffic Signs and Markings (GB5768-1999) provides the prescription: the number of the chevrons should be decided according to the radius, curve length and deflection angle. For the curve section where the deflection angle is small, one chevron can be placed at the midpoint of the curve. For the curve section with a big deflection angle and a long curve, the number of the chevrons can be decided according to the demand. On the ramp of a small radius curve, the chevrons should guarantee that the driver can see at least three in view continuously along the curve.

Zwahlen and Park (1994) performed a study to determine the optimum number of chevrons that a driver needs to be able to accurately estimate the sharpness of the curve. The young drivers, seated in a black booth, saw a slide of a standard curve with 12 chevrons equally spaced around the curve in 2s. Then they turned 90 degrees to view the test curves. They found that their judgment accuracy increased as the number of chevrons increased to four, but there was no fundamental difference between four and eight chevrons. So they concluded that "four equally spaced delineation devices, such as chevrons, within a total visual field of about 11 degrees provided adequate curve radius estimation cues for unfamiliar drivers." They also concluded that having four delineation devices instead of three, while it improves perceptual accuracy slightly, would be better because if a chevron were missing, the driver would still have three chevrons visible, which would allow higher accuracy levels than with just two visible.

Zador et al. (1987) placed chevrons in such a way that three were always in view on rural curves. Coaxial cables were used to collect speed and lateral placement data 100 ft after the beginning of the curve. The study found that the presence of delineation modifications significantly influenced both vehicle speeds and lateral placement, and the chevrons resulted in a small reduction in vehicle speeds and a movement of vehicles away from the centerlines. Jennings and Demetsky (1983) also tested chevrons for effectiveness. They found that chevron signs reduce speed and speed variance, and promote more desirable lateral placement. Chevrons were spaced at two times MUTCD recommended delineator spacing.

But Zwahlen's (1993) study was not the same as others'. The author found that some curves had increased speed and some decreased after chevrons were used. Overall there was a slight speed reduction, but only affecting speed when drivers approached. Also, the study states that the center speed was not significantly influenced by the presence of delineation devices. Niessner (1983) states in his report that he found chevrons designed with a yellow background and black legend were

quite visible and the line of the chevrons, consisting of three in view, accurately delineated the curve. The data collected for this study also show that chevrons significantly reduced the total fatal accident rate.

Based on six assumptions, Rose and Carlso (2005) give the spacing between two, three and four chevrons on two-lane rural roads that can be seen at the same time according to the common smallest radius single curve. The first chevron is set at the point of spiral to curve (SC). The start position must guarantee that the vehicle has enough distance to decrease the speed to 25mph before it reaches the point of spiral to curve (SC), with deceleration being $11.2ft/s^2$. But no specific model was provided. For the single curve with four chevrons in view, Mao Hui (2001) gives the calculation model. The vehicle start point is decided by the least values of the perceived distance and sight distance; the first chevron is set in front of the driver, two are equally located on the advancing direction; the others are placed on the residual circular curve at an average distance, and on the transition curve by twice the average distance.

Carlson (2004) summarized former research in "Simplifying Delineator and Chevron Applications". The report indicates: among the accidents on a rural road, about 40% have crashed off the road. The accident rate has strong correlation with curve curvature; drivers underestimate curvature when the curve angle is less than 90; and drivers overestimate the curvature when the exit curve angle is between 8 and 15.

According to the achievements mentioned above, there is little research about the specific calculated model of chevron spacing. Although Mao Hui gave out the specific calculated model for a two-lane road, she did not take the viewing distance of human beings and irradiation distance of the vehicle headlight into account. So the space is incorrect when the radius increases gradually. At the same time, the principle in the upward references can only guarantee that the drivers see a certain number of chevrons before they enter the curve. But in the curve, the visible chevron number can not be guaranteed, which reduces the security. Rose and Carlso set the first chevron at the curve jumping-off point in their research, which was not found in other research. If the first chevron is not placed at the curve jumping-off point, then the driver can not correctly perceive the jumping-off point of the curve, which will cause danger too. In Carlson's research report, there is a summary of former experiences and a theoretical model which all indicate: the setting space of the chevron increases obviously as the radius increases. But the research in the following chapters denies it.

Based on the previous research, this paper considers various restraining factors from the point of driver's vision physiology, which will increase security before the driver enters the curve as well as after entering the curve.

The rest of this paper is organized as follows. Chapter 2 introduces the placing principle of the chevrons. Chapter 3 gives the data source for the model and the assumptions. Chapter 4 states the model process. Chapter 5 gets the specific spacing model. Chapter 6 analyses and discusses the model principle and results.

2 Principle

The former principle of placing chevrons considered the period of the car driving before the curve, which has two obvious defects: First, the position of the first chevron is uncertain, maybe on the curve or on the line before the curve; Second, the distance between the driver and the fourth chevron is beyond the distance of headlights or the driver's eyesight.

The first defect is obviously negative for the driver's safety. The point at which the driver can judge correctly should be the point of curvature (ZH). But if placed according to the former method, it could cause the driver's wrong estimation of the starting point of curvature, which leads to swerving earlier or later, and could result in a traffic accident. The second is as dangerous as the first . If the distance between the driver and the fourth chevron is beyond the headlights or the driver's eyesight, then the safe number of chevrons present in driver's eyes can't be guaranteed; the accuracy of radius perception decreases, and an accident could occur.

Elisabeth R. Rose and Paul J. Carlson (2005) investigated the spacing of chevrons by the previous principle, but the study also indicated the position of the first chevron should be at the point of curvature (PC), the ZH; in addition to the range of headlights and eyesight, the placement of the last chevron in the driver's view should also take some longitudinal factors into account, such as the distance of headlights and eyesight. So the position of the first four chevrons can be confirmed, but that only improves one aspect of the former method, and could only increase the driver's safety before the curve. If the spacing in the curve is the same as the spacing before curve, probably the driver can't see an adequate number of chevrons, and the safety can't be guaranteed when driving on the curve.

In foreign criteria, the principle of setting chevrons is that the road user always has an adequate number of chevrons (regardless of the number) in view, until the change in alignment eliminates the need for the signs. According to it, countless points i need to be set on the curve, and then calculation begins with every point as the first point in turn. Obviously, the spacing S_i is not the same. Considering the comfort of driving on the curve, the spacing must obey some rule. At the same time, drivers should at least have four in view together. Therefore choose $S = \min\{S_i\}$ to be placed equally on the curve. But it is impossible to calculate every point. This paper clarifies this method. First, construct a basic curve, select the five character points on the curve as ZH、HY、QZ、YH、ZH to calculate (only calculate the first four points, the driver's line of sight is on the line when calculating the HZ) , and then select the minimum as the value of S. This way the safety of driving on the curve can be assured.

3 Data and assumption

The basic curve is the familiar form of a curve, and this study is based on the constructed basic curve for the different designed speeds. Considering the driver's

safety, the radius and the curve length of the basic curve are set at half the value of a common minimum radius and curve length, according to the Design Specifications for Highway Alignment in China (JTG D20-2006), with transition curve:circular curve:transition curve=1:1:1.

- A higher grade highway (two-way roads and upwards) is chosen as an objective, and the modeling is based on the assumptions as follows:
- The terrain is flat (only a horizontal curve considered, and no vertical curvature);
- The weather is good; the driver isn't affected by the weather;
- Chevrons are set in the median and the one shoulder on two-way roads and upwards, and are fixed on the barrier in ramps;
- A glare screen has been installed in the median, and the distance of headlights is according to the high beam;
- The car is driving along the central line of the lane;
- The first chevron is fixed on the point of curvature (PC), ZH.

4 Models

4.1 Driver's minimum angle of view

The driver's minimum angle of view can be considered with four factors as following: driver's line of sight, cone of sight, cone of acute vision and angle of headlights. Referring to some data, we found that the minimum line of sight is 40 degrees when speed is 105km/h; the common cone of sight is 20 degrees; the readability of the edge of cone is bad, so the cone of acute vision is about 10 degrees (head is fixed); as for the angle of the headlights, the S.A.E. stipulates 24 degrees, and the E.C.E. is 10 degrees, which has not been clearly set at present, so the angle of headlights is considered as 18 degrees. Then the driver's minimum angle of view is 10 degrees, which is the minimum of the four, and around the central line is 5 degrees respectively.

4.2 Driver's minimum distance of view

The driver's minimum viewing distance can also be considered with four factors as follows: static eyesight, dynamic eyesight, the eyesight of night, and the distance of the eye focus. The static eyesight takes 0.7 as the criterion, uses the landolt ring, and chooses the type Ⅱ chevrons of which the stimulated width is 90mm, to calculate the sight distance as 205m; the dynamic eyesight is about 1/3 the time of the static eyesight; the eyesight of night and the distance of the eye focus are restricted by the headlights. Table 1 shows the distance of headlights according to the Photometric Characteristics of Headlights for Motor Vehicles (GB4599-94), which choose 127m as the distance of the headlights considering safety.

Table 1. Distance of the headlights

Lamp type	Luminous intensity(cd)	Lighting distance(m)
Using car, two lamps	24000	127
New car, two lamps	30000	137
Using car, four lamps	40000	152
New car, four lamps	48000	162

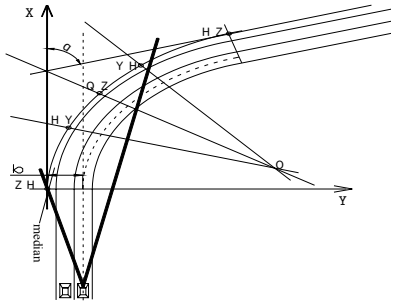


Figure 1. Building the corodition axis

As stated above, 127m is chosen as the driver's minimum distance.

4.3 Models of the basic curve

In view of limited time and energy, the study object in this paper is a two-way road which has a median, chevrons fixed in the median, and one shoulder, the same as other roads and ramps.

The highway may have six or eight lanes, so the angle between the left line of sight and the direction of driving may be larger than 5 degrees when the driver's left sight reaches the ZH; when the angle between them is 5 degrees, the distance between the driver and ZH may be more than 127m. By calculating, the two-way road with four lanes accords with it, so only calculate the four-lane road. As for the six and the eight lane road, if there is enough space, a button and other delineation devices can be placed together.

4.3.1 Choose the basal lane

In the previous research, chevrons were fixed only on one side of a low-grade road with two lanes, so one lane is chosen as the basal computing lane. But it is unreasonable to choose the midline of one side as the basal lane in atwo-way highway with four lanes and upwards. Considering the vehicles on the inside lane (near the median) may block the driver's view on the outside lane, and the basal lane must be the disadvantaged lane, so the outside lane should be chosen as the basal lane (Figure 1).

4.3.2 Parameter models of basic curve

The tangent off-set method is selected due to the complicated computer process, and the curve's tangent of chevrons is X axis; the X axis's perpendicular through the

ZH is Y axis (Figure 1), and the coordinate axis of the turn right curve is the same to the left.

According to the research (cited in PENG Fei 2006), the parameter model of a basic curve can be applied to this paper as follows:
The model of the first transition curve:

$$\begin{cases} x_1 = l - \dfrac{l^5}{40c^2} \\ y_1 = \dfrac{l^3}{6c} - \dfrac{l^7}{336c^3} \end{cases} \tag{1}$$

where $c = Rl_s$.

The model of the first transition curve with the inside offset is b:

$$\begin{cases} m = x_1 - \eta \\ n = y_1 + \xi \end{cases} \tag{2}$$

Where $\eta = b\sin\dfrac{l^2}{2c}$, $\xi = b\cos\dfrac{l^2}{2c}$.

The model of a circular curve:

$$\begin{cases} x_2 = R\sin\dfrac{l}{R}\cos\beta_0 - R(1-\cos\dfrac{l}{R})\sin\beta_0 + x_{hy} \\ y_2 = R\sin\dfrac{l}{R}\sin\beta_0 - R(1-\cos\dfrac{l}{R})\cos\beta_0 + y_{hy} \end{cases} \tag{3}$$

where $\beta_0 = \dfrac{l_s}{2R}$, x_{hy} and y_{hy} can be gained by $l = l_s$ in formula (1).

The model of a circular curve with the inside offset is b:

$$\begin{cases} m = x_2 - b\sin(\dfrac{l}{R} + \beta_0) \\ n = y_2 + b\cos(\dfrac{l}{R} + \beta_0) \end{cases} \tag{4}$$

By converting the coordinates of the first transition curve, the model of the second transition curve can be gained as follows:

$$\begin{cases} x_3 = x_{hz} - y_0 \sin\alpha - x_0 \cos\alpha \\ y_3 = y_{hz} + y_0 \cos\alpha - x_0 \sin\alpha \end{cases} \tag{5}$$

Where, $x_{hz} = T + T\cos\alpha$, $y_{hz} = T\sin\alpha$,

$$T = (R+p)\tan\frac{\alpha}{2} + q, \quad q = \frac{l_s}{2} - \frac{l_s^3}{240R^2}, \quad p = \frac{l_s^2}{24R} - \frac{l_s^4}{2384R^3}.$$

The model of the second transition curve with the inside offset is b:

$$\begin{cases} m = x_{hz} - (y_0 + \xi)\sin\alpha - (x_0 - \eta)\cos\alpha \\ n = y_{hz} + (y_0 + \xi)\cos\alpha - (x_0 - \eta)\sin\alpha \end{cases} \tag{6}$$

The coordinate value of the five character points can be achieved by the models listed above.

4.4. Models of the chevron spacing

According to the analysis above, the setting principle in this paper is that calculating the four character points (ZH, HY, QZ and YH) in turn as the jumping-off point, then the minimum is the S.

As for the ZH, which is the point of curvature and the important part from beeline to curve, it is a little different from the HY, QZ and YH. The safe decelerating distance is taken as a restricted condition, considering the importance of the curve radius perception before driver enters into the curve. That is, when the driver's left line of sight sees the ZH, the distance between the driver and ZH must be larger than the safe decelerating distance. Otherwise, an accident could happen due to not having enough time to decelerate even if enough chevrons (4 pieces) are in sight.

The author makes certain the optimum deceleration as 2.5m/s^2, considering the driving comfort (Bart & Cornelie & Ruben, 2006). Take 20km/h as the maximal speed variation in usual driving, the safe decelerating distance before the driver enters into

the curve is 67.9m when the designed speed is 120km/h. The safe decelerating distance decreases along with the decrease of the designed speed. So all the basic curves should be in accord with the restricted condition of the safe decelerating distance after calculating.

The following simply illuminates the models and steps by QZ, the other points having the same calculating principle.

4.4.1 Calculate the driver's coordinate P (m, n)

When the driver's left line of sight L_1 reaches QZ (Figure 2), the driver may be on the first transition curve or on the circular curve of HY—QZ, so the coordinate of P needs the attempt of simultaneous equations.

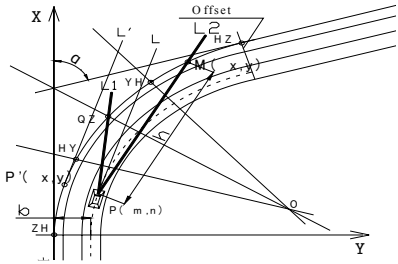


Figure 2. Figure about calculating

In Figure 2, L is the line of the driver's sight; it is the tangent of driving trochoid, the intersection point being P; the L' is the parallel of L and the tangent of curve with chevrons, the P' being the tangent point. If the chord deflect angle is δ, the transition curve angle is β, then:

$$\delta = \arctan\frac{y}{x} \approx \frac{\beta}{3} \tag{7}$$

The angle of line L and X axis is the same as the angle of line L' and X axis; it's β.

The relationship between point $P(m,n)$ and point $P'(x,y)$ can be transferred by b, which can be seen in (1) and (2), (3) and (4), (5) and (6).

According to the equation (7):

$$\tan\frac{\beta}{3}=\frac{y}{x} \tag{8}$$

The equation of the driver's left line of sight L1 :

$$y_{QZ}-n=\tan(\beta-5°)(x_{QZ}-m) \tag{9}$$

Attempt to join (1), (2), (8) and (9), where the x, y in equation (8) is x_1, y_1. If they have no right solution, then the driver is not on the first transition curve; then join (3), (4), (8) and (9), where the x, y in equation (8) is x_2, y_2, which can get the right solution.

4.4.2 Calculate the arc length L in the driver's view

The driver's right line of sight L_2 can be gained after getting the driver's coordinate P(m, n):

$$y-n=\tan(\beta+5°)(x-m) \tag{10}$$

$M(x,y)$ is the point of intersection of line L_2 and curve with chevrons, which may be on the circular curve of QZ—HY, or in the second transition curve. Therefore attempt to join (3), (4) and (10). If no right solution exits, then join the (5), (6) and (10).

As for the above, the result must meet $h=\sqrt{(x-m)^2+(y-m)^2}\le 127m$, where h is the distance between point P and point M, or else the equation (10) should be changed into equation (11):

$$y-n=\tan\beta_0(x-m) \tag{11}$$

Where $\beta_0<\beta+5°$, and:

$$(x-m)^2+(y-n)^2=127^2 \tag{12}$$

Attempt to join (3), (4), (11) and (12); if there is not a right solution, then the

point of intersection is on the second transition curve. Then the l can be achieved by joining (5), (6), (11) and (11), and the arc length LQZ which between QZ and M can be gained.

4.4.3 Spacing S

According to the steps above, LZH, LHY, LQZ and LYH could be obtained in turn. If the four chevrons are fixed in it, then:

$$S_{ZH} = L_{ZH}/3; \quad S_{HY} = L_{HY}/3; \quad S_{QZ} = L_{QZ}/3; \quad S_{YH} = L_{YH}/3 \tag{13}$$

At last, choose the minimum space as the S:

$$S = \min\left\{S_{ZH}, S_{HY}, S_{QZ}, S_{YH}\right\} \tag{14}$$

5 Results

Select six types of roads in which the design speed is 30km/h、40km/h、60km/h、80km/h、100km/h、120km/h in the criterion，and construct six basic curves according to the common minimum values prescribed in the criterion. By applying the setting principle and model in this paper, and programming through matlab, the results can be obtained as in Table 2.

Table 2. Theoretical Minimum Chevron Spacing

Design speed (km/h)		120	100	80	60	40	30
Minimum radius (m)		1000	700	400	200	100	65
Spacing (turn left) S (m)	See ZH	17.2	17.2	17.2	18.7	16.5	17.4
	See HY	22.7	24.0	28.3	14.2	8.4	6.1
	See QZ	23.0	24.6	28.1	12.3	6.3	4.3
	See YH	23.0	24.6	26.6	12.9	6.5	4.3
	Min	17.2	17.2	17.2	12.3	6.3	4.3
Spacing (turn right) S (m)	See ZH	20.7	20.7	20.7	22.2	20.0	21.4
	See HY	25.1	26.2	28.3	12.9	6.9	4.7
	See QZ	25.2	26.5	23.9	11.9	5.9	3.8
	See YH	25.2	26.5	26.4	12.5	6.0	3.8
	Min	20.7	20.7	20.7	11.9	5.9	3.8

According to the data above, the characteristic figures can be drawn as in Figure 3 and Figure 4.

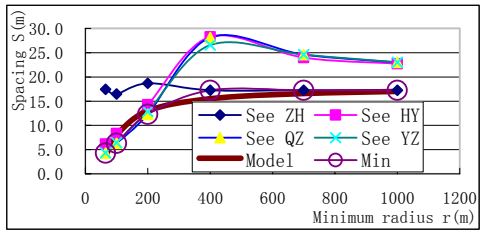


Figure 3. Spacing on the two-lane road (turn left)

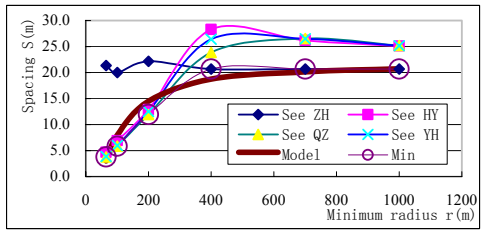


Figure 4. Spacing on the two-lane road (turn right)

From the above two figures, we can see that when the radius is smaller than 500m, the setting space of chevrons increases obviously as the radius increases. But when the radius is larger than 500m, the setting space of chevrons only increases a little. That is to say, although the space S increases as the radius increases, the increasing speed drops off gradually, especially when the radius is larger than 500m, as well as the concave of the curve is downward.

Make regression between the radius and the least spacing through SPSS, and we can get the left-turn and right-turn curve alignment chevron spacing model of the two-lane road respectively, shown in equations (17) and (18). Broad lines in Figure 3 and Figure 4 represent the regression model.

Regression model of spacing:

$$S = -\frac{A}{r} + B \tag{16}$$

Regression model of left-turn spacing:

$$S_L = -\frac{1000}{r} + 18 \qquad R^2 = 0.9477 \tag{17}$$

Regression model of right-turn spacing:

$$S_R = -\frac{1300}{r} + 22 \qquad R^2 = 0.9066 \tag{18}$$

From Figure 3 and Figure 4, it can be found that when the radius is more than 500m, the setting space of chevrons varies only a little, that is to say, the setting space approaches a certain value gradually as the radius increases. For the hyperbolic function model selected in this paper, we can get the asymptotic line equation:

$$S' = B \tag{19}$$

Then get the asymptotic line of the model, which is the largest setting space:

$$S'_{left} = 18m; \qquad S'_{right} = 22m \tag{20}$$

Through the analysis above, the regression model in this paper matches the smallest space value of the calculation, namely the model is reasonable.

6 Conclusions and future research

In this paper, the author clearly points out that the first chevron should be set at the jumping-off point of the curve (ZH). According to the former principle of chevrons, the first chevron placing position changes following the change of radius, which results in the first chevron being in the transition curve or the beeline, and that drivers can't make a correct judgment about the point of curvature. The authors set the first chevron at the ZH, effectively avoiding the driver's false judgment about the point of curvature, and enhancing the security of the road users; at the same time, it makes a supplement about the driver's visual district. There are only transverse view limits like the visual angle of the eye and the irradiation angle of the vehicle headlight in traditional methods, but during the actual driving, especially at night, the driver's visual distance is restricted by many aspects. Therefore when chevrons are set by the former methods, and when the radius increases to a certain value, the far side chevrons can't appear in the view of the driver's eye. On the basis of this, the paper adds the longitudinal distance limits like the sight distance of eyes and the irradiation distance of headlights, and further perfects the principle of the chevron placement by taking the driver's security as the fundamental jumping-off place.

In the former model, only the visible situation when driver hasn't entered into the curve was considered. The space can only guarantee the security before the driver enters the curve. When the drivers come into the curve, it can't be guaranteed that the driver will see enough chevrons. Therefore the conventional method doesn't consider the drivers' security when in the curve. This paper puts forward a method that

calculates the chevron spacing according to the characteristic points of a curve, making up the deficiency of the original method, which guarantees the security both before and after the drivers enter the curve.

From Figures 3 and 4, it can be seen that when the radius is less than 500m, the setting space of chevrons increases obviously as the radius increases. But when the radius is more than 500m, the setting space of chevrons increases only a little, which effectively illuminates the limits of the customary calculation method and principle, namely the setting space of chevrons not increasing all along as the radius increases. When the radius is more than 500m, because of the larger radius, the arc that drivers can see is limited by the sight distance of their eyes and the irradiation distance of the vehicle headlights. Because the sight distance is a fixed value, when the radius further increases, the arc that drivers can see only changes a little. At the same time, models in this paper have proved this point and calculated the largest chevron spacing.

In this paper the chevron spacing is calculated in the situation as follows: during the night, the irradiation distance of the vehicle headlight is considered to be the worst, and the various parameters chosen are the smallest value. Because it is the most secure situation, all the parameters have been chosen at the smallest value. In the ultimate results, the minimum that was calculated under the situation of four characteristic points was taken, which makes the result smaller. But on the other hand, the space can satisfy not only the demands of the worst environment and vehicle, but also can satisfy the demands of driving in the daytime. As calculated under the worst condition, with four chevrons that can be seen, drivers may see more than four chevrons in the curve, which can offset the influence of the sheltering of the chevrons caused by the vehicle from some degree. Therefore it is more dependable for four chevrons in view. The calculated results in this paper haven't been changed into integers, and in practical construction it can be adjusted according to the actual road conditions

It is necessary to illuminate: when choosing the radius, this paper has not chosen all the values, but the common smallest radius according to the criteria. When calculating, all the parameters have been chosen at the common smallest value, so that the results can be applied to all the curves of different radii.

Furthermore, the ultimate setting space can't guarantee the drivers having four chevrons in view all the time when they are driving. One of the reasons is that the setting principle in this paper takes every point in the curve as the jumping-off point in turn, but in fact it is impossible. So the author simplified it, taking the limited characteristic points. Although the ultimate space was chosen at the smallest value, it is possible that drivers may not see four chevrons at any time. The second reason is that when drivers are driving in the curve, vehicles in the fast lane may shelter the sight of the vehicles on the outside lane, and the results in this paper haven't checked concerning the sheltering.

This paper can be further improved in several aspects:

- All the curves in this paper are of a basic type. It is more difficult for the setting space in combined curves, which can be studied more concretely and thoroughly.
- As the vehicle volume on the road increases, the possibility that vehicles in the fast lane may shelter the sight of the vehicles on the outside lane increases gradually, especially on a multi-lane road. The method for calculating the vehicle shelter possibility and the dependability of the four in view still needs further research.
- The road is supposed to be flat in this paper; that is to say only the horizontal curve factors have been taken into consideration. For the road that contains a vertical curve, whether the gradient causes the shelter or not, there is a question of how to set the chevron spaces to make them more reasonable.
- During the model establishment, the influence on drivers caused by bad weather has not been considered. A question is how to take the sight and eyesight reduction status of driver under bad weather into account.
- The objective road in this paper is a two-lane road. A multi-lane road needs to combine curve chevrons, buttons and delineators. How to combine these needs further research.

Acknowledgements

This work is supported by the Headquarters of China Railway Tunnel Group Combo. We are also grateful to the Fourth Survey and Design Institute of China Railways as well as Shaoqing QIN

Notation

The following symbols are used in this paper:

R = radius of the curve with chevrons;

l = the distance between any point in the transition curve to the transition curve jumping-off point;

l_s = the length of the transition curve;

T = total tangent length;

α = route angle;

p = the offset value of original curve;

q = the distance between the former circular curve jumping-off point or end-point to the transition curve jumping-off point;

b = inside offset.

References

Bart van Arem, Member, IEEE, Cornelie J.G. van Driel, and Ruben Visser (2006). "The Impact of Cooperative Adaptive Cruise Control on Traffic-Flow Characteristics." IEEE Transactions on Intelligent Transportation Systems, 7(4), 429-436.

Carlson P.J., E.R.Rose, S.T.Chrysler and A.L.Bischoff (2004). Simplifying delineator and Chevron Applications for Horizontal Curves. Texas Transportation Institute, College Station.

Elisabeth R. Rose and Paul J (2005). "Carlson. Spacing Chevrons on Horizontal Curves." Transportation Research Record. 191B, Transportation Research Board of the National Academies, Washington, D.C., 84-91.

Jennings B.E. and M.J. Demetsky (1983). Evaluation of Curve Delineation Signs on Rural Highways. Report VHTRC 84-R16. Virginia Highway and Transportation Research Council in cooperation with the U.S. Department of Transportation, Charlottesville, VA.

Mao Hui (2001). "Chevron alignment sign placing." Journal of Highway and Transportation Research and Development. Gonglu jiaotong Keji, 18 (5), 90-92.

Niessner C.W (1983). Post Mounted Delineators. Report FHWA-TS-83-208. FHWA, U.S. Department of Transportation, Washington, D.C.

Peng Fei (2006). Discussion of the sideline's equation on subgrade in curve based on the same coordinate axis. Highway, NO.10, 97-101.

Zador P., H.S.Stein, P.Wright and J.Hall (1987). "Effects of Chevrons, Post-Mounted Delineators, and Raised Pavement Markers on Driver Behavior at Roadway Curves." Transportation Research Record. 1114, TRB, National Research Council, Washington, D.C., 1-10.

Zwahlen H.T (1993). Optimal Application and Placement of Roadside Reflective Devices for Curves on Two-Lane Rural Highways. Report FHWA/OH-94/011. FHWA, U.S. Department of Transportation.

Zwahlen H.T. and J.Y. Park (1994). "Curve Radius Perception Accuracy as a Function of Number of Delineation Devices (Chevrons)." Transportation Research Record. 1495, TRB, National Research Council, Washington, D.C., 99-106.

Asphalt Rubber and Its Potential Use in China

Terrie Bressette[1], Haiping Zhou[2], Anne Stonex[3], and R. Gary Hicks[4]

[1]California Department of Transportation, Materials Engineering and Testing Services, 5900 Folsom Boulevard, Sacramento, CA 95819; Phone (916) 227-7303; Fax (916) 227-7316; e-mail: Terrie_Bressette@dot.ca.gov
[2]MACTEC Engineering and Consulting, Inc., 4704 Roseville Road, Suite 108, North Highlands, CA 95660; Phone (916) 332-5552; Fax (916) 332-5559; e-mail: HZhou@mactec.com
[3]MACTEC Engineering and Consulting, Inc., 3630 E. Wier Avenue, Phoenix, AZ 85040; Phone: (602) 437-0250; Fax: (602) 437-3675; e-mail: AStonex@mactec.com
[4]California Pavement Preservation Center, 203 Langdon Hall, CSU, Chico, Chico, CA 95929-0930; Phone (530) 898-5981; email: RGHicks@csuchico.edu

Abstract: In 2004, the number of scrap tires in China reached 120 millions and this number is increasing by 12% every year. There is a considerable demand for using scrap tires in civil engineering applications. This paper presents a description of asphalt rubber technology and its potential use in China as an alternative use for scrap tires. Asphalt rubber has been used in the United States of America and several other countries for many years and the asphalt rubber products have been shown to provide significantly improved engineering properties over conventional paving grade asphalt. Consequently, pavements constructed with asphalt rubber mixes have many benefits such as improved durability and improved resistance to surface initiated fatigue and reflection cracking, improved aging and oxidation resistance due to higher binder contents, thicker binder films, and anti-oxidants in the tire rubber.

The use of asphalt rubber is not only an innovative way to recycle used tires that otherwise would be wasted, but also one of the most effective methods in extending pavement life through the improved engineering properties of the asphalt rubber products. China has thousands of miles of semi-rigid asphalt concrete and portland cement concrete pavements in need of maintenance and rehabilitation. This paper presents the latest advances in the asphalt rubber technology and the state-of-the-practice and discusses various asphalt rubber products used by the California Department of Transportation and their potential use in China.

Key words: asphalt rubber, rubberized asphalt concrete, crumb rubber modifier, asphalt rubber chip seal, stress absorbing membrane interlayer

1 Introduction

1.1 Background

In the recent Asphalt Rubber 2006 Conference held in Palm Springs, California, USA from October 24 to 27, 2006, several delegates from the People's Republic of China expressed great interest in the use of asphalt rubber technology in China. This is understandable as China's economy has been continuously growing at an explosive rate for the last 25 plus years and China has also made considerable investments in its roadway infrastructure. China now has become the biggest rubber consumer in the world (Huang, 2006). In 2004, 190 million tires were produced in China. Also in 2004, the number of scrap tires reached 120 million and this number is increasing by 12% every year. There is a considerable demand for using scrap tires (Huang, 2006).

Asphalt rubber has been used in the United States of America and several other countries for decades and the asphalt rubber products have been proven to provide significantly improved engineering properties over conventional paving grade asphalt. In the United States, the California Department of Transportation (Caltrans) has used asphalt rubber since the late 70s (Caltrans, 2006a; Van Kirk, 1989; Caltrans, 2005a) and is continuing to advance this technology through various evaluation approaches. With the recent passage of the California Assembly Bill 338, the use of asphalt rubber will be increased from the current 20% of asphalt concrete used to 35% by year 2013. It is believed that the asphalt rubber technology used in California and experience gained by Caltrans may well be suited and applicable to conditions in China.

When using recycled tires in pavements, there are generally two processes that are employed: the wet and dry processes. With the wet process, it can further be classified into two distinct forms. In the first form, crumb rubber modifier (CRM) is blended and interacted with asphalt cement at high temperatures under high agitation prior to mixing with aggregate in a hot mix plant. In the second form, CRM is blended with hot asphalt cement at the refinery or at an asphalt storage and distribution terminal and transported to the hot mix plant or job site for use. This form of asphalt rubber does not require agitation to keep the CRM particles evenly dispersed in the modified binder. With the dry process, the CRM is substituted as a portion of the aggregate and added to the aggregate in a hot mix plant before adding the asphalt cement (Caltrans, 2005b).

1.2 Purpose

This paper presents a description of the asphalt rubber technology with the focus on the first form of the wet process, the state-of-the-practice that has been successfully used by Caltrans, and its potential use in China. The paper provides an overview of asphalt rubber technology, benefits, limitations, cost, and environmental considerations. The paper describes in details the two of the most commonly used asphalt rubber products in California: rubberized asphalt concrete and asphalt rubber spray applications such as rubberized chip seals and their material types, design, and construction.

2 Overview of Asphalt Rubber

2.1 What is asphalt rubber?

Asphalt rubber is a blend of asphalt cement, reclaimed tire rubber, and certain additives in which the rubber component is at least 15 percent by total binder mass and is reacted in the hot asphalt cement sufficiently to cause swelling of the rubber particles (ASTM, 2001). It is produced at elevated temperatures (≥177°C, 350°F) and under high agitation to promote the physical interaction of the asphalt binder and rubber constituents, and to keep the rubber particles suspended in the blend. Various petroleum distillates or extender oils may be added to reduce viscosity, facilitate spray applications, and promote workability.

2.2 How is asphalt rubber used?

Asphalt rubber is used as a binder in various types of asphalt pavement construction including hot mixes (HMA) and spray applications. It is also used in crack sealants. For hot mixes, asphalt rubber is most commonly used in gap-graded and open-graded mixes, particularly for surface courses and for thin overlays that are 30 to 60 mm (1.2 to 2.4 inches) thick. The most common spray application is a chip seal, also called a stress absorbing membrane (SAM). Chip seals are primarily used for maintenance and pavement preservation.

Asphalt rubber chip seals may also be overlaid with hot mix, making them interlayers, typically called SAMI-R. SAMIs are used primarily for pavement rehabilitation.

2.3 Where should asphalt rubber products be used?

Asphalt rubber products can be used wherever conventional asphalt concrete or bituminous surface treatments would be used. Asphalt rubber hot mixes are typically most effective as thin overlays of distressed flexible or rigid pavements having an adequate structural capacity since they provide better resistance to reflective cracking and fatigue than conventional dense-graded HMA. This makes the product more attractive for use in areas where reflective cracking is the predominant distress mode and where there are vertical geometric constraints such as curb-and–gutter alignment or limited underpass clearance.

Asphalt rubber products should be used where and when weather conditions are favorable for placement and compaction because of temperature requirements. However, this does not prevent their use at high elevations. Asphalt rubber products have been used with success in most of the geographical and climate zones in California, and in Arizona from low desert through the mountain/alpine climate zones. The products should also be very suitable for use in many areas of China.

2.4 Where should asphalt rubber products not be used?

Asphalt rubber paving materials should not be placed in the following conditions:

- During cold or rainy weather with ambient or surface temperatures <13°C (55°F).
- Over pavements with severe cracks more than 12.5 mm (0.5 inch) wide where traffic and pavement deflection data are not available.
- Areas where considerable handwork is required.
- Where haul distances are too long to maintain mixture temperature required for placement and compaction.

2.5 Benefits of asphalt rubber

Asphalt rubber provides significantly improved engineering properties over conventional paving grade asphalt. The benefits of asphalt rubber include the following:

- From asphalt rubber binders:
  - Increased viscosity that allows greater film thickness in paving mixes without excessive drain down or bleeding.
  - Increased elasticity and resilience at high temperatures.
- Pavements constructed with asphalt rubber mixes:
  - Improved durability.
  - Improved resistance to surface initiated fatigue and reflection cracking due to higher binder contents and elasticity.
  - Reduced temperature susceptibility.
  - Improved aging and oxidation resistance due to higher binder contents, thicker binder films, and anti-oxidants in the tire rubber.
  - Improved resistance to rutting (permanent deformation) due to higher binder viscosity.
  - Lower pavement maintenance costs due to improved pavement durability and performance.
- Other benefits:
  - Reduced construction times due to thinner lifts.
  - Better chip retention in chip seals due to thick films of asphalt rubber binder.
  - Savings in energy and natural resources by using waste products.
  - Improved safety due to better long-term color contrast for pavement markings.

2.6 Limitations of asphalt rubber

Limitations on use of asphalt rubber include the following:

- Higher unit costs due to mobilization of asphalt rubber production equipment.
- Asphalt rubber is not best suited for use in dense-graded HMA.
- Construction may be more challenging, as temperature requirements during placement are more critical.
- Asphalt rubber materials are often difficult to hand work because of stiffer binder and coarser mixture gradations.
- If work is delayed more than 48 hours after blending the asphalt rubber, some binders may not be usable.
- For chip seals in remote locations, hot and/or pre-coated aggregate may not be available because there may not be a hot-mix plant within reasonable haul distance of the job site.

2.7 Cost considerations

The unit costs of asphalt rubber products are typically higher than those of conventional or polymer modified products. The unit costs escalate considerably for jobs with less than 2,250 tonnes (2,500 tons) of asphalt rubber concrete because mobilization and set up of the asphalt rubber binder production equipment cost as much for small jobs as for big ones. Therefore, smaller asphalt rubber projects may not be cost effective with respect to initial cost. The asphalt rubber binder content is also typically higher than for conventional mixes, which increases binder cost. However, Caltrans has found that gap-graded asphalt rubber can be placed in approximately half the thickness of conventional dense-graded mixes, when used to mitigate reflective cracking; thus countering the increase in cost. In addition, improved durability, particularly resistance to reflective cracking, and related reduced maintenance needs, should substantially reduce the overall life cycle costs and help offset the difference in initial cost. Generally, asphalt rubber is cost effective when used as thin gap- or open-graded surface courses or overlays of 30 to 60 mm (1.2 to 2.4 inches) compacted thickness, chip seals and interlayer applications.

2.8 Environmental considerations

There are a number of social benefits of using scrap tire rubber as briefly described below:

- Not Adding to Tire Stockpiles. The primary benefit is recycling waste tires into engineering use instead of adding to tire stockpiles.
- Value-Added Use of Waste Tires. The addition of rubber enhances the physical properties of the resulting paving materials over the life of the pavement, and thus provides a long-term benefit to tax payers and the motoring public.
- Noise Abatement. Significant reductions in traffic noise, ranging from 4 to 8 dB, have been reported for open-graded RAC.
- Emissions (fumes and smoke). When production temperatures are held to less than 325ºF, study results generally indicate little if any difference from emissions of conventional HMA materials, and no identifiable increase in risk to health or safety of hot plant or paving personnel or the public.
-

3 Rubberized Asphalt Concrete

3.1 Types of rubberized asphalt concrete

There are several types of rubberized asphalt concrete (RAC) mixes being used by Caltrans including RAC-G, RAC-O, and RAC-O-HB, as shown in Table 1. These mixes are

used for different purposes and should be properly selected and used for most cost-effective benefit and performance. In California, gap- and open-graded RAC mixes are most often used as overlays for maintenance and/or rehabilitation of existing asphalt concrete and portland cement concrete pavements. RAC is also used as surface (wearing) courses for new pavement construction (Caltrans, 2006b), most often in areas where traffic noise is a consideration. Caltrans encourages the use of RAC. When all factors (including cost) are the same, RAC should be used as a preferred material for the surface layer instead of conventional HMA (Caltrans, 2006c). These practices are very suitable to conditions in China where pavement maintenance and rehabilitation of asphalt concrete and portland cement concrete pavements are becoming major issues on state highways and demand for quieter pavement is increasing in major cities (Huang, 2006).

Table 1. Asphalt rubber hot mixes

Mix Type	Purpose	Appropriate Use	Thickness/Mix Design
RAC-G (gap-graded)	Provides a durable, flexible pavement surface with • Increased resistance to reflective cracking, rutting and oxidation, • Good surface friction characteristics due to the texture provided by the gapped aggregate grading, and • Often reduced traffic noise. RAC-G acts as a structural layer in the pavement.	Can be used for overlays for pavement rehabilitation – particularly to retard reflective cracking or in new construction for a wide range of traffic volumes and loadings when placed as a surface course. However, RAC-G mixtures are not recommended for parking areas.	Thickness governed by: Structural adequacy upgrade; reflective crack retardation; and ride quality improvement. Mix design: California Test Method 367. Hveem stability (minimum 23); VMA: 18 to 23%: Air voids 3-5%.
RAC-O (open-graded)	• Provides a durable, highly flexible pavement surface with enhanced drainage and frictional characteristics • Reduces splash and spray to improve visibility during wet weather • Reduces hydroplaning in wet conditions to reduce potential for skidding • Provides increased resistance to reflective cracking and oxidation • Provides a smooth ride	Highly effective as an overlay of PCC and AC pavements in locations where potential for reflection of joints and/or cracks is severe. It is also used as a maintenance blanket to restore surface frictional characteristics and to help preserve the underlying pavement. Caltrans does not use RAC-O in snow country due to concerns about possible damage from tire chains and snowplows.	Thicknesses are generally twice the nominal aggregate size with 3/8-and ½-inch mixes most frequently used. California Test 368, with asphalt rubber binder content set at 1.2 times the optimum neat asphalt binder content with a check test for drain off.
RAC-O-HB (open-graded with high binder content)	Restores surface friction and provides improved resistance to fatigue and reflective cracking, as well as to stripping and oxidative aging. These factors increase the durability of open-graded pavements.	Best used where resistance to reflective cracking is most critical due to higher asphalt-rubber binder contents of 8 to 10 percent by mass of dry aggregate.	Thicknesses are generally twice the nominal aggregate size with 3/8-and ½-inch mixes most frequently used. RAC-O-HB is designed according to the RAC-O procedure, but the multiplier for asphalt rubber binder content is increased to 1.65.

3.2 Asphalt rubber binder design

Asphalt rubber binders must be properly designed and produced to comply with specifications and provide a quality product suitable for the expected climate and traffic

conditions. The interaction between asphalt cement and CRM materials is material-specific and depends on a number of factors including:

- Asphalt cement source and grade
- Rubber type, source, amount, and gradation
- Interaction time and temperature

The project environment and condition determines the base asphalt cement grade to use for asphalt rubber binders. In California, performance grade (PG) 58-22 asphalt cement is specified for use in high mountain and high desert climate areas where resistance to cold temperature cracking is critical to long term performance. A stiffer grade of PG 64-16 is specified for use in the rest of California (coastal, valley, low or southern mountains, low desert). China has a wide range of environments; the selection of base asphalt cement grade must consider the specific project environment and location.

For a specific project, an appropriate asphalt rubber binder design must be developed using the designated source and grade of asphalt, asphalt modifier (extender oil, if used), and CRM materials (scrap tire and high natural). The asphalt rubber binder used by Caltrans calls for 20 ± 2 percent CRM by total binder mass. The CRM must include 25 ± 2 percent of high natural rubber CRM and 75 ± 2 percent scrap tire CRM, by mass of total CRM. The scrap tire CRM consists primarily of No. 10 to No. 30 sized particles (2 mm to 600 μm sieve sizes). The high natural rubber CRM is somewhat finer, mostly No. 16 to No. 50 (1.18 mm to 300 μm sieve sizes). Caltrans uses extender oils and high natural CRM to enhance the asphalt rubber interaction.

The physical properties specified by Caltrans for asphalt rubber binders include rotational viscosity, resilience, ring-and-ball softening point, and cone penetration (Caltrans, 2006a). A design profile is used to verify compliance of field asphalt rubber binder properties. Table 2 presents an example of a state-of-the-practice asphalt rubber binder design profile. The design profile is based on laboratory testing of the specific sources and grades of asphalt, asphalt modifier, and CRM identified for use in asphalt rubber binder production for specific project(s). The design profile must identify the specific component materials (source or supplier and grade) and proportions thereof used in the design. Due to differences between laboratory and field production, the design profile serves only as a guide to indicate expected trends in asphalt rubber viscosity during asphalt rubber binder production.

Table 2. Asphalt rubber binder design profile

Test Performed	Minutes of Reaction					45 minutes Specification Limits**
	45	90	240	360	1,440	
Viscosity*, Haake at 190°C, Pa·s, (10^{-3}), or cP	2400	2800	2800	2800	2100	1500 – 4000
Resilience at 25°C, % Rebound (ASTM D5329)	27	--	33	--	23	18 Minimum
Ring & Ball Softening Point, °C (ASTM D36)	59.0	59.5	59.5	60.0	58.5	52 – 74
Cone Pen. at 25°C, 150g, 5 sec., 1/10 mm (ASTM D217)	39	--	46	--	50	25 – 70

** The viscosity test is conducted using a hand-held Haake viscometer or equivalent.*

*** Caltrans specifications as of June 2006.*

3.3 RAC construction

The key to producing quality asphalt rubber materials and construct asphalt rubber pavements is temperature control in all aspects of the work, including asphalt rubber binder manufacture, hot mix production, delivery, placement, and compaction.

3.3.1 Asphalt rubber binder manufacture

Field production of high viscosity asphalt-rubber binder generally includes the following processes (Caltrans, 2006a; RACTC, 1999): the component materials are metered into high shear blending units to incorporate the correct proportions of extender oil and CRM into the paving grade asphalt; the blending units thoroughly mix the CRM into the combined hot asphalt and extender oil, and the blend is pumped into a heated tank where the asphalt-rubber interaction proceeds.

The quality of the resulting asphalt-rubber binder depends on proportioning, temperature, agitation, and time. Temperature is critical for process control. Tanks that store the asphalt cement and the asphalt rubber between initial blending and use are heated to temperatures from 190°C to 226°C (375°F to 435°F) and insulated. Asphalt-rubber production equipment and storage tanks generally include retort heaters or heat exchangers to heat the asphalt cement and/or asphalt-rubber binder.

Augers are used to agitate the asphalt rubber inside the tanks to keep the CRM particles well dispersed. The asphalt rubber binder must be interacted with agitation for a minimum of 45 minutes at temperatures from 190 to 218°C (375 to 425°F) to achieve the desired interaction between asphalt and rubber. To maintain the reaction temperature within the specified range, the asphalt cement temperature must be at 204 to 224°C (400 to 435°F) before the design proportions of scrap tire and high natural CRM are added. This is because the CRM is added at ambient temperature (not heated). The viscosity of asphalt rubber must meet the requirements of specifications before it can be used. This can be monitored with a hand held rotational viscometer at various time intervals for quality control and assurance.

3.3.2 Hot mix production

Either batch or continuous asphalt concrete plants can be used in RAC mix production, but typically at higher operating temperature than conventional HMA. The asphalt rubber production equipment is usually set up as close to the mixing unit as feasible to minimize the length of the heated and/or jacketed binder feed lines. Special heavy-duty pumps are needed to transfer the asphalt rubber binder. For drum plants, the asphalt rubber producer is required to use a flow meter that interlocks the asphalt rubber binder feed with the plant aggregate feeds.

Both the plant and field inspectors should visually inspect the RAC in the haul truck bed for signs of any problems with the mix and check mix temperature. When in doubt, the inspector should obtain samples immediately and have them tested for gradation and asphalt rubber binder content. Some of the potential "trouble" signs to watch for in the mix are as follows:

- Segregation: This may be particle size segregation or temperature segregation.
- Blue smoke: Mix is too hot.
- White smoke (steam): Too much moisture in aggregate.
- Stiff appearance: Mix may be too cool.
- Dull, flat appearance: May indicate low asphalt rubber binder content and/or excessive fines (minus No. 200 (0.075 mm) sieve size). Localized areas of dullness

may indicate insufficient mixing of the asphalt rubber binder and aggregates, or mix segregation.
- Slumped and shiny: May indicate high asphalt rubber binder content.

3.3.3 Delivery, placement, and compaction

Conventional equipment such as haul trucks, paver, rollers, and sand spreaders are used to deliver, place, and compact RAC materials. Trucks hauling RAC mix are tarped to retain heat during transport. Pavers should be equipped with vibratory screed and screed heaters, and automatic screed controls with skid. Rollers for RAC must be steel-wheeled (drum), and must be equipped with pads and a watering system to prevent excessive pick-up. The minimum recommended roller weight is 7.3 tonnes (8 tons). Rubber tired, pneumatic rollers are not appropriate for compacting RAC mixes because of excessive pick up of the mixture by the tires. As with all hot mix asphalt, compaction is essential to the performance of asphalt rubber mixtures.

For open-graded mixtures, vibratory compaction is not used and there are no compaction requirements. These mixes are typically placed as surface courses in thin lifts about 25 to 30 mm (1 to 1.2 inches) thick. Compaction is achieved with a few passes by rollers operating in the static mode.

4 Asphalt Rubber Spray Application

4.1 Types of spray applications

There are primarily two types of asphalt rubber spray applications being used in California: chip seal and SAMI-R, as presented in Table 3. The spray applications are used for different purposes, either as surface treatments or interlayers prior to an asphalt concrete overlay, and should be properly selected and used for the most cost-effective benefit and performance.

Table 3. Asphalt rubber spray applications

Type	Purpose	Appropriate Use	Spread Rates
Chip Seals	Provides a flexible, waterproof, skid resistant and durable surface that resists oxidation and is highly resistant to reflective cracking.	• Rehabilitation of structurally sound pavements which are cracked or raveled. • Restoration of surface frictional characteristics. • Routine preventative maintenance to extend the life of AC pavements.	• Asphalt rubber binder at 0.6-0.7 gal/yd^2 (2.5-3 l/m^2). • Hot pre-coated chips (nominal 1/2 or 3/8-inch size) at 30-44 lbs/yd^2 (15-22 kg/m^2).
SAMI-R	SAMI-R is a non-structural membrane interlayer used to retard and minimize reflective cracking in overlays placed on it, and to minimize further infiltration of surface water through the pavement structure.	• Used under thin maintenance overlays and is a pavement rehabilitation tool. SAMIs can provide structural benefit in rehabilitation projects when reflection cracking is the governing distress mode for overlay design. • A SAMI is not recommended as part of new construction.	Same as for chip seal.

A chip seal is the application of a bituminous binder immediately followed by the application of an aggregate. The aggregate is then rolled to embed it into the binder. A SAMI-R is simply an asphalt rubber chip seal that is overlaid with conventional HMA or RAC. SAMI-R acts as

an interlayer to interrupt crack propagation and has been shown to be highly effective in delaying reflective cracking in overlays of existing distressed asphalt and jointed portland cement concrete pavements.

4.2 Binder design

It is important that the asphalt rubber binder properties, particularly the primary field control of viscosity, remain in compliance with specifications. Guidelines provided in the RAC section are also applicable to spray application. The extender oil dosage for chip seals ranges from of 2.5 to 6 percent by mass of the asphalt cement. To minimize potential for flushing and bleeding for hot climate and heavy truck traffic locations, the extender oil content may further be reduced.

4.3 Spray seal construction

4.3.1 Chip seal

Chip seals are surface treatments that are extremely sensitive to the effects of construction operations and site conditions, including temperature (air, cover aggregates, and underlying pavement). Clean or pre-coated chips are also critical and are required to be hot (143ºC to 183ºC, 260ºF to 325ºF). Embedment and adhesion of the chips must be accomplished by rolling while the asphalt rubber membrane is still hot. It is not advisable to place chip seals when ambient or pavement temperature is less than 15ºC (60ºF). Hot pre-coated chips should be applied immediately behind the asphalt rubber binder spray. Rolling of a chip seal is done to orient and embed the rock (get the flat sides down). Pneumatic rollers are normally used for rolling chip seals because the kneading action of the rubber tires promotes embedment, and the tires ride on the cover aggregate rather than binder. Sweeping is done at the completion of chip sealing to remove surplus aggregate from the surface of the new chip seal to minimize loose rocks.

4.3.2 SAMI-R

The construction process for SAMI-R is similar to that of chip seals, except that an additional layer of hot mix asphalt will be placed on top of the SAMI-R. Chip retention is not an issue unless the interlayer is opened to traffic prior to overlay. Otherwise, the aggregate chips are sandwiched in. They are keyed into the overlay during compaction and prevent formation of a slippage plane along the relatively thick asphalt rubber membrane. Tack coat, fog seal or sand should not be applied over an interlayer because it could interfere with bonding of the overlay.

5 Conclusions and Recommendations

5.1 Conclusions

Asphalt rubber materials have been used for decades and have been improved through the years by advancements in materials sciences and pavement technology. Asphalt rubber products have many benefits ranging from the use of waste scrap tires to the improvement of pavement performance. However, the decision to use asphalt rubber products should consider where and when to use them and how to realize the identified benefits. The asphalt rubber materials must be carefully and properly designed, produced, and constructed to achieve the desired pavement performance.

Based on the historical experience and current practices in the USA, asphalt rubber hot mixes are most appropriate for use as a thin surface layer on a structurally sound pavement because of their superior performance in controlling reflective cracking, improving pavement durability, and reducing tire-pavement noise. Asphalt rubber chip seals are best used as surface treatments for improving surface frictional characteristics and extending existing

pavement life and the SAMI-R is best used to minimize reflective cracking in overlays and infiltration of surface water into the pavement structure.

5.2 Recommendations

As China continues its economic growth and the expansion of its infrastructural construction, pavement maintenance and preservation will become more and more important and urgent to protect the investment in China's roadway systems. China has thousands of miles of semi-rigid asphalt concrete and portland cement concrete pavements in need of maintenance and rehabilitation and plans to invest billions of dollars to expand its roadway network to rural villages and to construct quieter pavements in major cities (Huang, 2006). Asphalt rubber products, as one of the preferred choices by Caltrans to deal with pavement maintenance and rehabilitation needs described in this paper, can have a great potential in China.

6 Disclaimer

The information presented in this paper does not necessarily represent the official views of Caltrans nor is Caltrans responsible for its contents. It only reflects the views and experience of the authors, who are responsible for the facts and accuracy of the information presented herein.

References

American Society of Testing and Materials, (2001). "D 6114 Standard Specification for Asphalt-Rubber Binder," in *Vol. 4.03, Road and Paving Materials; Vehicle-Pavement Systems*, Annual Book of ASTM Standards 2006, ASTM, West Conshohocken, PA, 2001.

Caltrans, (2005a). "Synthesis of Caltrans Rubberized Asphalt Concrete Projects," Office of Flexible Pavement Materials, Sacramento, CA, October, 2005.

Caltrans, (2005b). "Use of Scrap Tire Rubber-State of the Technology and Best Practices," Office of Flexible Pavement Materials, Sacramento, CA, February, 2005.

Caltrans, (2006a). "Asphalt Rubber Usage Guide," Office of Flexible Pavement Materials, Caltrans, Sacramento, CA, July 2006.

Caltrans, (2006b). "Use of Rubberized Asphalt Concrete for New Construction. Memorandum," Caltrans, Sacramento, CA, April 24, 2006.

Caltrans, (2006c). "Highway Design Manual Chapter 600," Caltrans, Sacramento, CA, June 2006.

Huang, Wen-Yuan, Chiang Yung-Chieh, and Xu Li-Ting, (2006). "Study and Application of Asphalt Rubber and Rubberized Pavement in China," Proceedings, Asphalt Rubber 2006, Palm Spring, California, USA, pp 327-340.

RACTC, (1999). "Informational Brochure," Rubberized Asphalt Concrete Technology Center, Sacramento County, 1999.

Van Kirk, J., (1989). "Caltrans Experience with Asphalt-Rubber Concrete-An Overview and Future Direction," Proceedings, National Seminar on Asphalt Rubber, Kansas City, Missouri, 1989.

LABORATORY EVALUATION OF PERFORMANCES OF ASPHALT RUBBER AND GAP GRADED MIXTURES

Rongji CAO[1,2] and Qifeng BAI,[2]

[1] Transportation College, Southeast University, 210096, Nanjing, PH(025)-86576810; FAX(025)86576666; email: crj@ti.js.cn
[2] Jiangsu Transportation Research Institute, 211112, 2200 Chengxin Road, Nanjing, PH(025)-86576819; FAX(025)86576666; email: bqf@ti.js.cn

Abstract: Till now, only dry process is used to manufacture asphalt rubber (AR) in highway in China. This paper describes the production of asphalt rubber with wet process and the application of asphalt rubber for concrete pavement overlays. This paper presents: (1) selection of crumb rubber and base asphalt; (2)producing parameter tests that affect the properties of asphalt rubber; (3) evaluation of performances of asphalt rubber Stress Absorbing Membrane Interlay (AR-SAMI); (4)evaluation of different design methods of gap graded asphalt rubber mixture... After a series of tests, some conclusions were drawn from this research. Asphalt rubber using truck tire crumb rubber has better performances than that using car tire crumb. Using #20 mesh size crumb rubber can give better asphalt rubber than #40 mesh and #60 mesh crumb rubber can. As for base asphalt, asphalt rubber using penetration 70 base asphalt has better high temperature performance than that using penetration 90 base asphalt, but its low temperature performance and fatigue resistance were not as good as asphalt rubber using penetration 90 asphalt. AR-SAMI had good fatigue and shear resistance. Gap graded asphalt rubber mix designed using Superpave Gyratory Compactor(SGC) method had better rutting and low temperature cracking resistance than that using Marshall method.
Key words: *crumb rubber, SAMI, gap graded, wet process, performance*

1 Introductions

Asphalt rubber mixture has good rutting resistance, low temperature cracking resistance, fatigue resistance and anti-aging performance. It has been widely used in Arizona, Texas and Florida in the US. In 1980's, asphalt rubber using dry process has been studied in China, but for many reasons, this technique was not widely used. In the last two years, asphalt rubber using wet process was introduced to China. In this paper, the properties of asphalt rubber, the performance of asphalt rubber Stress Absorbing Membrane Interlayer (AR-SAMI) and gap graded asphalt rubber mixture are presented.

2 Asphalt Rubber Binder Characterization

2.1 Materials

2 types of base asphalt: penetration 70 and Penetration 90.

2 types of crumb tire types: truck tires and car tires

4 types of crumb rubber particle sizes: #20, #40 and #60 mesh.

2.2 Production of Asphalt Rubber

The base asphalt and crumb rubber were mixed at temperature of 175□ for 60 minutes in lab, with a high speed shear blender.

2.3 Experiment Results and Analysis

2.3.1 Crumb Rubber Type and Particle Size

Two kinds (truck tire and car tire) and 4 particle sizes (#20, #40 and #60 mesh) of crumb rubbers were used to evaluate the influence on properties of asphalt rubber. In these tests, asphalt rubber was produced with 80% of penetration 70 asphalt and 20% of crumb rubber content by mass.

The characterization of the asphalt rubber was performed using conventional binder tests and Superpave binder tests. The test results are presented in Table 1 and Table 2.

Generally, all the asphalt rubbers have good performances. Compared with base asphalt, their penetration decreases and softening point increases. Asphalt rubbers can meet the requirements of PG82-28, which indicates that asphalt rubbers have good high temperature and low temperature performances. The values of $G^*\cdot\sin\delta$ at 31□ are about 300kPa, which are less than 5000kPa, the limit of PG specification. This indicates asphalt rubber has good anti-aging and fatigue resistance performances.

From softening point, $G^*/\sin\delta$, it was shown that asphalt rubber using truck tire has better high temperature performance than that using car tire. At the same time, asphalt rubber using truck tire has lower $G^*\cdot\sin\delta$ value, which shows that it has better fatigue resistance. On the basis of S value and elastic recovery, asphalt rubber using truck tire has better low temperature properties.

In conclusion, truck tire can provide better performances for asphalt rubber than car tire, because truck tire has more rubber hydrocarbon, which can form better network structure and result in better performances.

From test results, #20 mesh crumb tire has better high temperature performance than #40 mesh and #60 mesh crumb tire, while their low temperature cracking resistance and fatigue resistance are similar.

In conclusion, truck tire is better than car tire in asphalt rubber. Asphalt rubber with #20 mesh truck tire has the best properties. In the following tests, truck tire of #20 mesh was used to produce asphalt rubber.

2.3.2 Crumb Rubber Content

Rubber content is one of the main factors that influence the properties of asphalt rubber. Table 3 shows the test results of different crumb rubber contents. In these tests, asphalt rubber was produced with #20 mesh truck tire. With the increasing of rubber content, asphalt rubber's high temperature and low temperature performances, anti-aging properties were improved significantly. As rubber content increases from 12% to 24%, the penetration of asphalt rubber increases 28.8□, elastic recovery value increases 11%, G*/sinδ value increases 6~8 times, and creep stiffness at -18□ decreases 75%. This indicates that increasing crumb rubber content is very effective in improving the properties of asphalt rubber. However, increasing crumb rubber content will decrease workability. It is recommended that the viscosity of asphalt rubber at 177□ should be 1.5 to 4.0 Pa•s [2].

Viscosities of asphalt rubber with different rubber contents were tested and shown in Table 4. The viscosity of asphalt rubber 16% or 18% crumb rubber are between 1.5 to 3.0 Pa•s. Therefore the appropriate rubber content is between 16% and 18%.

Table 1. Conventional binder test results of asphalt rubber

Tests	Test Method	Truck tire rubber			Car tire rubber		
		#20 mesh	#40 mesh	#60 mesh	#20 mesh	#40 mesh	#60 mesh
Penetration, 25℃, 0.1mm	ASTM D5	28.6	30.7	32.2	44.0	47.3	62.9
Softening point,℃	ASTM D36	72.6	75.7	72.7	67.6	66.2	63.1
Recovery, 25℃,%	ASTM D6084	89	89	88	82	83	72

Table 2. Superpave binder tests results of asphalt rubber

Test	Truck tire rubber			Car tire rubber		
	#20 mesh	#40 mesh	#60 mesh	#20 mesh	#40 mesh	#60 mesh
Original						
Dynamic Shear: G*/sinδ at 76℃,kPa	10.564	8.283	7.331	5.441	3.771	3.328
G*/sinδ at 82℃,kPa	7.601	5.556	4.845	3.642	2.521	2.126
RTFOT aging						
Dynamic Shear: G*/sinδ at 76℃,kPa	11.661	9.995	8.298	5.700	4.336	4.347
G*/sinδ at 82℃,kPa	8.246	6.786	5.430	3.676	2.842	2.620
PAV aging						
Dynamic Shear: G*sinδ at 31℃,kPa	291	258	305	337	307	297
Bending Beam:						
S-value at -18℃,MPa	92	97	73	132	139	141
m value at -18℃	0.327	0.334	0.310	0.316	0.348	0.346
S-value at -24℃,MPa	228	214	174	303	298	344
m value at -24℃	0.227	0.269	0.265	0.269	0.280	0.269
Performance Grade	86-28	86-28	86-28	86-28	86-28	86-28

Table 3. Test results of asphalt rubber with different rubber content

Tests	Rubber content, %				Spec.
	12	16	20	24	
Conventional binder tests					
Penetration, 25℃ , 0.1mm	41.9	38.0	28.6	27.8	≥25
Softening point，℃	60.7	65.6	72.6	89.5	≥57
Elastic recovery, 25℃，%	81	86	89	92	/
Superpave binder tests					
Original					
Dynamic Shear: G*/sinδ at 76℃,kPa	2.658	4.949	10.564	18.235	≥1.0
G*/sinδ at 82℃,kPa	1.505	3.226	7.601	13.715	
RTFOT aging					
Dynamic Shear: G*/sinδ at 76℃,kPa	4.583	7.508	11.661	24.689	≥ 2.2
G*/sinδ at 82℃,kPa	2.862	5.341	8.246	18.126	
PAV aging					
Dynamic Shear: G*sinδ at 31℃,kPa	427	339	291	337	≤5000
Bending Beam:					
S-value at -18℃,MPa	216	144	92	50	≤300
m value at -18℃	0.300	0.312	0.327	0.332	≥ 0.1
S-value at -24℃,MPa	369	329	228	175	≤ 300
m value at -24℃	0.249	0.252	0.227	0.268	≥ 0.1

Table 4. Viscosity of asphalt rubber at 177℃

Rubber content, %	12	16	18	20	Test Method
Viscosity, Pa·s	1.30	1.90	2.00	8.10	ASTM D4402

2.3.3 Types of Base Asphalt

The influence of types of base asphalt was showed in Table 5. The asphalt rubber using penetration 90 base asphalt exhibits better low temperature cracking resistance than asphalt rubber using penetration 70 base asphalt. But high temperature performance is on the contrary. Since different types of base asphalt will lead to different properties of asphalt rubber, it is suggested that base asphalt type should be chosen properly according to the climate and pavement structure conditions. In the following tests base asphalt of penetration 70 was used.

Table 5. Test results of asphalt rubber with different types of base asphalt

Tests	Test Method	Base Asphalt		Spec.
		Pen. 70	Pen. 90	
Conventional binder tests				
Penetration, 25℃，0.1mm	ASTM D 5	28.6	44.7	≥25
Softening point，℃	ASTM D 36	72.6	69.9	≥57
Elastic recovery, 25℃，%	ASTM D 6084	89	91	/
Superpave binder tests				
Original				
Dynamic Shear: G*/sinδ at 76℃,kPa	AASHTO TP5	10.564	7.093	≥1.0
G*/sinδ at 82℃,kPa		7.601	5.262	
RTFOT aging	AASHTO T240			
Dynamic Shear: G*/sinδ at 76℃,kPa	AASHTO TP5	11.661	7.397	≥ 2.2
G*/sinδ at 82℃,kPa		8.246	5.035	
PAV aging	AASHTO PP1			
Dynamic Shear: G*sinδ at 31℃,kPa	AASHTO TP5	291	192	≤5000
Bending Beam:	AASHTO TP1			
S-value at -18℃,MPa		92	89	≤300
m value at -18℃		0.327	0.380	≥ 0.1
S-value at -24℃,MPa		228	196	≤ 300
m value at -24℃		0.227	0.315	≥ 0.1

3 Asphalt Rubber Stress Absorbing Membrane Interlayer Tests

Asphalt Rubber Stress Absorbing Membrane Interlayer (AR-SAMI) is proved to

be useful to prevent reflective cracking [1]. In this paper flexural beam fatigue tests and shear strength tests at 45 degree were conducted. Asphalt rubber was produced with base asphalt of penetration 70 and #20 mesh crumb rubber, 17.5% by mass.

3.1 Fatigue Tests of AR-SAMI

Flexural beam fatigue tests with compound specimens were done to evaluate the cracking resistance of AR-SAMI. As shown in Figure1, compound beam was made up of three parts: the top part is 30 mm thick HMA, the bottom part is 10 mm thick HMA and the middle part is 10 mm interlayer. The bottom part was cut, 10 mm in width, to simulate pavement crack. The HMA is conventional dense graded mixture. The bottom part was cut from compacted mixture and put in the bottom of a mould. Then, sprayed interlayer was put on it and the mould was filled with hot mixture and compacted with rolling compactor. Finally, the specimen was cut to the right size.

Constant strain, 250με, tests were conducted at load frequency of 10 Hz, and at temperature of 15□. The test results are presented in Table 6.

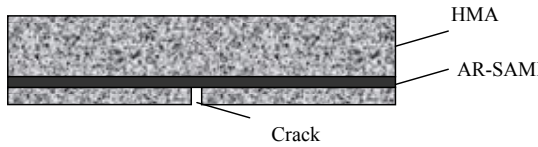


Figure1. Compound beam for fatigue test

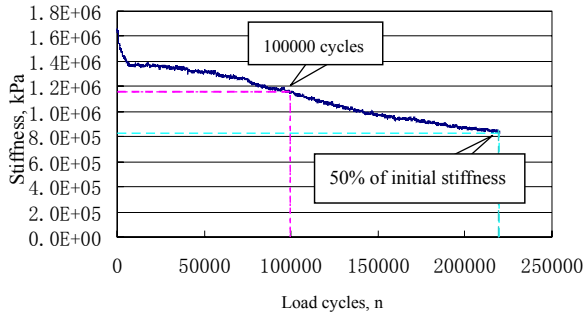


Figure 2. Load cycles and stiffness of compound beam in fatigue test

Semi-rigid pavements are common in China, and seal coat with polymer modified asphalt emulsion is usually used on the top of cement treated base. Tests with AR-SAMI and modified asphalt emulsion were conducted to compare. Table 6 presents the results. After 100000 load cycles, the stiffness of compound beam with modified asphalt emulsion seal decreased to 55% of the initial stiffness, but the stiffness of compound beam using AR-SAMI is still 69.4% of initial stiffness. Therefore, AR-SAMI is more effective in preventing reflective cracking than polymer modified asphalt emulsion seal coat. AR- SAMI could be used as interlayer between semi-rigid base and asphalt pavement to reduce reflective cracks.

Table 6. Fatigue test results

Interlayer type	Polymer modified asphalt emulsion seal coat	AR-SAMI
% of initial stiffness	55.0	69.4

3.2 Shear Strength Test

Shear strength test, with an angle of 450, was designed to evaluate the shear strength of AR-SAMI, as shown in Figure 3. Two kinds of specimens were used: laboratory compacted and field cores. The following was the procedure of making specimens in laboratory. First cut cement treated base to 50mm in height. Then put sprayed interlayer on it. Put them in a gyratory compactor mould which was filled with hot mixture and compacted in Superpave Gyratory Compactor.

The specimens were sheared with the angle of 450, at temperature of 50□, with the loading speed of 50 mm/min.

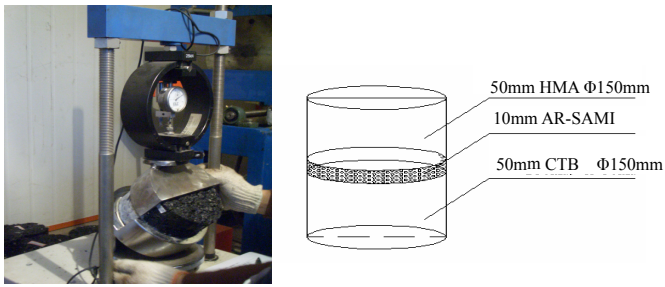


Figure 3. Shear strength test

The test results are shown in Table 7. Compared with samples made in laboratory, specimens taken core from pavement exhibited a little higher shear strength. Figure 4

shows the specimen after shear strength test. Obviously, shear failure occurred in asphalt mixture, not in AR-SAMI. This indicated that shear failure was not due to the deficiency of AR-SAMI shear strength. The shear strength of AR-SMAI was actually higher than 0.35MPa.

Table 7. Shear strength test results of AR-SAMI, 30℃

	Compacted in lab	Taken core from pavement
Strength, MPa	0.33	0.35

Figure 4. Samples after shear strength test

4 Asphalt Rubber Mixtures Design and Performance Tests

Marshall Method and Superpave Gyratory Compactor (SGC) method are used in asphalt rubber mix design. Marshall Method is used in Arizona, while SGC method test is used in Texas [3].

This research used the same aggregate and binder to design gap graded mixtures with each method and studied their performance.

4.1 Gap Graded Asphalt Rubber Mixture

Two kinds of gap graded asphalt rubber mix were designed, using Marshall Method and SGC method, as shown in Table 8 to Table 10. In Marshall Method, the specimens were compacted with 75 blows on each side. In the SGC method, the N_{ini}, N_{des}, N_{max} are 8, 75, and 115 respectively.

Table 8. Gradation of gap graded rubber asphalt mix

Mixture	Percentage Passing, %									
	19	13.2	9.5	4.75	2.36	1.18	0.6	0.3	0.15	0.075
Mix1	100	90.9	60.6	31.8	21.0	15.8	10.4	5.5	2.8	1.2
Mix2	100	93.7	56.2	25.4	21.0	17.5	13.3	10.8	8.7	6.8

Table 9. Marshall Test results of mix1

Mixture	AC %	G_{mb} g/cm^3	G_{mm} g/cm^3	VMA %	VV %	VFA %	MS kN	FL 0.1mm
Mix1	7.6	2.421	2.548	20.7	5.0	75.9	7.9	41.1
Arizona Spec.	/	/	/	≥19	4.5～6.5	/	/	/

Table 10. SGC test results of mix2

Mixture	AC %	Volumetric index at N_{des},			% G_{mm} at N_i,	% G_{mm} at N_m,
		% G_{mm}	VMA, %	VFA, %		
Mix2	7.7	96.1*	19.9	80.6	87.0	96.8
Texas Spec.	/	97	≥19	/	/	/

4.2 Comparison of Two Kinds of Asphalt Rubber Mixtures

Tests were conducted to compare the performances of moisture susceptibility, rutting resistance and crack resistance of the two kinds of gap graded asphalt rubber mixtures.

4.2.1 Stone-to-stone skeleton Analysis

Voids in the Coarse Aggregate in dry rodded condition (VCADRC) and Voids in Coarse Aggregate in asphalt mixture (VCA_{mix}) of the two types of mixtures were measured according to AASHTO T19.

Table11 shows that mix2 has better aggregate skeleton because its VCA_{mix} is smaller than its VCADRC. However in mix1, the VCA_{mix} is bigger than VCADRC, which indicates that too much fine aggregate in the mixture influences the stone-to-stone contact of coarse aggregate.

Table 11. Void in coarse aggregate (VCA)

Mixture	VCA_{DRC}	VCA_{mix}
Mix1	40.98	41.65
Mix2	40.97	35.76

4.2.2 Moisture Susceptibility

In the research, Marshall Stability Test under saturated and Indirect Tensile Test under frost-thaw conditions were adopted to evaluate the moisture stability of the mixture. The results are presented in Table 12 and Table 13. Two mixtures all meet the specification and their performances are similar.

Table 12. Results of Marshall stability test under saturated

Mixture	Control		Saturated (48h)		TSR, %
	VV, %	Stability, kN	VV, %	Stability, kN	
Mix1	4.9	7.78	5.1	7.04	90.5
Mix2	3.8	7.24	3.6	6.57	90.7
Spec.	/	/	/	/	>85

Table 13. Results of indirect tensile test under frost-thaw

Mixture	Control		Frost-thaw		TSR, %
	VV, %	Strength,	VV, %	Strength,	
Mix1	6.2	0.85	6.4	0.69	81.1
Mix2	4.9	1.05	4.8	0.86	81.9
Spec.	/	/	/	/	>80

4.2.3 Wheel Tracking Test

Wheel tracking tests (T0719)[4] at 60□ were done to evaluate the rutting resistance. The test results of dynamic stability and cycles per 1mm of rutting are shown in Table 14. The dynamic stability of mix2 is about 30% more than that of mix1, so mix2 has better rutting resistance performance.

Table 14. Wheel tracking test results

Mixture	Dynamic Stability (cycles/mm)			
	1	2	3	Aver.
Mix1	3938	3150	3000	3363
Mix2	4200	4200	5250	4550
Spec.	/	/	/	>3000

4.2.4 Low Temperature Beam Bending Test

Beam bending test (T0719) [4] at -10□ was carried out to characterize the low temperature performance of mixture. The results are presented in Table 15. The results shows that the bending strength and the failure strain of mix2 are higher than that of mix1. Therefore mix2 is better in low temperature performance, because of its stone-to-stone structure and higher asphalt content.

Table 15. Beam bending test results (-10℃)

Mixture	Max. load, kN	Deflection, mm	Strength ,MPa	Strain, με
Mix1	0.96	0.63	7.861	3283.0
Mix2	0.97	0.65	7.924	3400.6
Spec.	/	/	/	>2000

After comparison of two mixtures, it could be concluded that mix2 has better high and low temperature performances than mix1. The moisture susceptibility of each mixture is similar.

5 Conclusions

This paper presents the results of a laboratory study which evaluated the effects of rubber type, size, rubber content and base asphalt type on the properties of asphalt rubber binder. The performances of AR-SAMI and AR-GAP were evaluated. The main conclusions achieved so far are as following:

Based on laboratory tests, asphalt rubber has superior performances. The asphalt rubber binder tested in this research could meet PG 82-22 requirement.

The test results showed that truck tire rubber is better than the performances of asphalt rubber binder car tire. The #20 mesh size rubber is better than #40 and #60 mesh size rubbers.

The pavement performance and construction workability considered, 16~18% rubber content was preferred in the research.

As for base asphalt, asphalt rubber using penetration 70 asphalt has better high temperature performance than that using penetration 90 asphalt, but its low temperature performance and fatigue resistance are not as good as asphalt rubber using penetration 90 asphalt.

The mixture designed with SGC is better than the one designed with Marshall Method on high and low temperature performances. Their performances of moisture susceptibility were similar.

References:

State of California Department of Transportation(2005). “Use of Scrap Tire Rubber State of The Technology and Best Practices.” Sacramento, California.

Arizona Department of Transportation(2000). “Standard Specifications for Road and Bridge Construction.” Phoenix, AZ.

Texas Department of Transportation(2004). “Standard Specifications for Construction and Maintenance of Highways, Streets, and Bridges.”

Minister of communication of China, Test Specification of Asphalt and Asphalt Mixtures, (JTJ052-2000)

New Technical Method for Production Homogenous and Storage Stable Polymer Modified Asphalt

Al-Hadidy AI[1] and Tan Yi-qiu[2]

[1] PhD Student, School of Transportation Science and Engineering, (HIT), China-Harbin; PH(+86) 139-460-33276;email : abd_et76@yahoo.com
[2] Prof. of School of Transportation Science and Engineering, (HIT), China- Harbin; PH(+86) 139-046-37262;email : yiqiutan@163.com

Abstract: The main objective of this research study is to find a new technical method for producing a homogeneous and storage-stable polymer modified asphalt paving materials instead of the dynamic vulcanization process. A simple machine work under heat and pressure was designed and manufactured for this purpose.

In this paper, polypropylene-modified asphalt is reported as a justification for this method. Seven different blends including control mix were subjected to binder testing such as rheological tests, as well as to some other tests related to the homogeneity of the system.

Optimum asphalt content was obtained by Marshall method and used in all the modified mixes. The engineering properties, including Marshall stability and indirect tensile strength were determined. Research results indicate that modified binders showed a higher softening point, keeping the values of ductility at minimum range of specification of (100^{+} cm), and caused a reduction in the percentage of weight loss due to heat and air (i.e. increase durability of original asphalt). The results indicated that the inclusion of produce polypropylene by this method with asphaltic concrete mixtures gave a quite satisfactory result in terms of stability and other Marshall properties, and increase in indirect tensile strength.

Keywords: *Polypropylene, Polymers, Modified asphalt, Asphalt binders, Asphalt concrete, Marshall test, Indirect tensile strength.*

1 Introduction

It is well known that rheological and engineering properties of the asphalts and asphalt concrete mixtures can be improved by the addition of small quantities of polymers (3-8% wt), such as polyelefins (polyethylene, ethylene/propylene dienecopolymer, chlorinated polyethylene), styrene-butadiene rubber (SBR), styrene-butadiene block (SBS) and ethylene/ vinylacetate (EVA) coplymers(Boutevin,1989 and Lu X,2001). Polymer modified asphalts are used in high performance specially pavements to suppress the noise level and to drain out water readily among other desirable effects. During its service life, the pavement deteriorate; as a consequence of multiple factors that operate simultaneously, such as load and frequency of traffic and environmental factors(M.S. et al,2004).

However, the major obstacle to widespread usage of polymer modified asphalts (PMAs) in paving practice has been their tendency towards gross phase separation under quiescent conditions at elevated temperature because of the poor compatibility between polymer and asphalt(S.M.A. et al,1991).

(Hailong Jin et al,2002) reported that polystyrene has been found to be very difficult to disperse into asphalt even using a higher shear because of its incompatibility with asphalt.

Consequently, the coarsely dispersed phase separates quickly once agitation stops, they also added sulfur to improve the storage stability of polystyrene-modified asphalt for practical use.

This procedure is similar to the vulcanization of rubber/plastic blends (i.e., dynamic vulcanization as reported by (G.A. Wen et al,2001).

It is well known that the inclusion of sulfur in asphalt forming hazardous fumes of H_2S, SO_2 and CO_2, when they are exposed to temperatures higher than 140 °C, and most of asphaltic concrete mixtures are prepared in a temperature of more than 150 °C, this reason will be limited the usage of a dynamic vulcanization process for storage stable polymer modified asphalt.

(Al-Hadidy,2006) used the machine mentioned by(Lobna,1992) for the thermal degradation of polypropylene and used the resultant polymer in the modification of asphalt paving materials. Also he mentioned that the inclusion of pyrolisis polypropylene by this procedure gave a quite satisfactory result in terms of stability, increase in tensile strength and keeping the ductility values of virgin asphalt at minimum range of specification at (100^{+}cm).

(Fred,1971) defined that the thermal degradation process means the breaking down of structure; in one usage relating to polymers. It is taken to mean any process leading to deterioration of properties (i.e. reduction of molecular weight).

However, in this paper, a simple machine was manufactured in our laboratory as shown in Figure 1 for thermal degradation in order to simulate the resultant polymer from this machine with the one produced from the machine mentioned by (Lobna,1992).

In the present work, the polypropylene was used after thermal degradation as a resultant polymer from this machine for asphalt paving materials modification to investigate and explain.

To investigate and explain the existence of this thermal degradation machine, the rheological behavior of pyrolisis polypropylene-modified asphalt paving materials was determined and compare with the results of the rheological properties obtained by the method mentioned by (Lobna,1992). Also to verify its compatibility with asphalt a photomicrographs was obtained using optical microscopy.

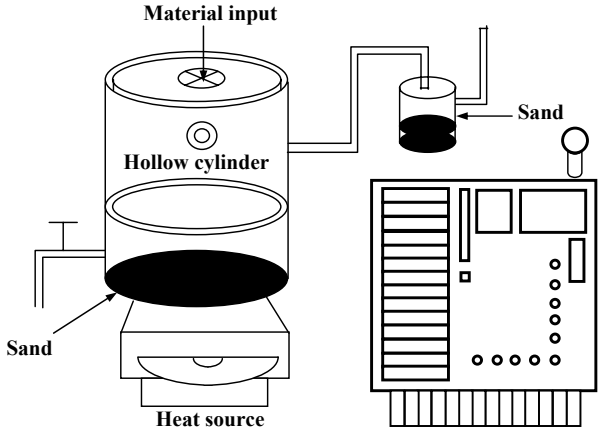


Figure 1. Thermal degradation machine.

2 Preparation of Pyrolisis Polymer (Polypropylene)

A given weight of polymer was placed inside the machine and heated to 400 °C with slow speed stirring for 2 hrs, and at the end of this time the resultant monomer was collected in a flat pan, then it grind to powder by mechanical grinding after cooling to room temperature.

3 Materials and Methods

3.1 Polypropylene, Asphalt Cement and Aggregate

The polypropylene (PP) material was obtained from the petrochemical company in Basrah (200 km south of Baghdad). The physico-chemical properties of PP before and after thermal degradation processes are shown in Table 1.

50/60 penetration grade asphalt cement taken from Baiji refinery (200 km north Baghdad) was used in this study, and its physico-chemical properties are shown in Table 2.

Al-Kazer aggregate was utilized in the preparation of stone matrix asphalt (SMA) specimens. Figure 2 illustrates the gradation limits of the used aggregates according to the South Carolina Department of Transportation SC DOT (Bradley et al,2004). The rock dust ($CaCO_3$) was used as filler and its physico-chemical properties are listed in Table 3.

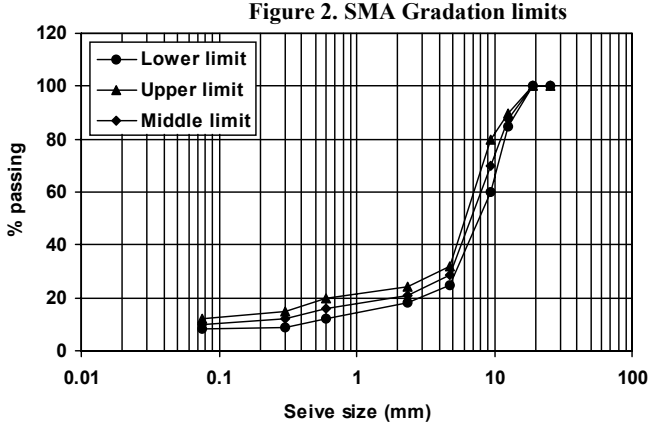
Table 1. Physico-chemical properties of PP before and after thermal degradation process.

Property	Result	Unit
	before pyrolisis process	
Chemical unit	$-CH_2-CH-CH_3-$ (Fred,1971)	
Thermal degradation (pyrolisis) temp.	308	°C
Color	white	-
Density	0.82	gm/cm^3
	After pyrolysis process	
Density	0.64	gm/cm^3
Melting point	153-157	°C
Decomposed	138-141	°C
Color	Brown	-

Table 2. Physico-chemical properties of asphalt cement.

Property	ASTM Designation No.	Test conditions & units	Result	SCRB limits
Penetration	D-5	25 °C,100g,5 s,dmm	51	50-60
Softening point	D-36	R&B, °C	52.5	50-60
Ductility	D-113	25 °C, 5 cm/min, cm	150^+	>100
Sp. gr	D-70	25 °C/ 25 °C	1.051	-
Flash point	D-92	COC, °C	259	-
Loss on heat	D-1754	5hrs, l63 °C, %	0.253	-
Asphaltenes	D-2006	%	22.4	-
Saturates	D-2006	%	27.6	-
Naphthenes aromatics	D-2006	%	35	-
Polar aromatics	D-2006	%	11	-

Figure 2. SMA Gradation limits



Property	Test result
Silicon dioxide (SiO_2)	2.7
Aluminum oxide (Al_2O_3)	0.35
Ferric oxide (Fe_2O_3)	0.93
Calcium oxide (CaO)	50.67
Magnesium oxide (MgO)	0.64
Sodium oxide (Na_2O)	0.02
Potassium oxide (K_2O)	N.D.
Sulfur trioxide (SO_3)	0.75

Table 3. Physical and chemical properties of filler.

3.2 Polypropylene-Asphalt Formulation

A different percentages of pyrolisis (PP) 1, 3, 5, 7, 9 and 11% wt. was mixed with asphalt at temperature of 160 °c for 5 min. by slow speed stirring.

3.3 Optimization of the mixtures

Marshall mix design (ASTM D-1559,2000) procedure is normally used to optimize the SMA mixtures in Iraq. In SMA mix design, usually the Marshall method of mix design is used to verify satisfactory voids in SMA mixtures. Laboratory specimens were prepared using fifty blows of the Marshall hammer per side. Seventy-five compaction blows were not used since they would not result in a significant increase in density over that provided by 50 blows. SMA mixtures have been more easily compacted on the roadway to the desired density than the effort required for conventional HMA mixtures[7, 8]. The optimum asphalt content for SMA mixtures is usually selected to produce 3-5% air voids and a drain down not to exceed 0.3%.

Marshall stability and flow values are generally measured for information but not used for acceptance(Fred,1971). In this research, compaction of all the SMA samples was performed using fifty blows of the Marshall hammer per side. The optimum asphalt content for the control SMA mixture was found to be 5.82% at 4.2% air voids.

3.4 Preparation of Marshall, indirect tensile strength and compressive strength specimens of modified SMA mixtures

An optimum asphalt content of 5.82% as found from Marshall control mix design (by wt. of total mix) was used in preparing all other PP modified mixes to maintain consistency through the study.

The following steps are performed for the formulation of compacted specimens:

1. The combined aggregate and filler were heated to 160 °C in a controlled electrical oven.
2. The modified binder was heated up to 150 °C in an electrical controlled oven.
3. The combination was mixed by mechanical mixing at a temperature of 150 ± 5 °C for 1.5 min.
4. The specimens formulated were then compacted at 135 °C using electrical Marshall apparatus specified by (ASTM D1559,2000).

One hundred and eighty samples for all percentages of LDPE were fabricated for each percent of LDPE. Five specimens were used for Marshall stability at 60 °C for 35 minutes. 80 specimens were prepared to determine the tensile strength values. These 80 specimens were divided into two groups (40 specimens each). The first group was tested at 25 °C for 2hrs and the second set tested after immersion at 60 °C for 24 hrs followed by 2 hrs immersion at 25 °C at 2 in/min deformation rates then the percent loss in tensile strength, tensile strength ratio, tensile strength and tensile stiffness modulus were determined using the equations described by (Lottman et al,1970 and Lottman,1978).

Also another 80 specimens were prepared to determine compressive strength values and divided into two groups similar to those for an indirect tensile strength test but tested with a deformation rate of (3.2 mm/min for Marshall specimen). Then the percent loss of compressive strength and compressive strengthof each were calculated.

4 Laboratory Testing

A series of tests were carried out on modified binders according to (ASTM,2000)methods to characterize the mixes designed for different percentages of PP as additive. The tests that were conducted include the following:

- Rheological tests, such as penetration, ductility, softening point . . . etc.
- Temperature susceptibility.
- Short-term aging using the thin film oven test (TFOT), which simulates the aging effect of the asphalt mixtures production and construction.
- Compatibility test:

The tests were conducted on asphaltic concrete mixtures modified with PP include:

- Marshall test (ASTM D1559,2000)
- Indirect tensile strength test (ASTM D4124,2000)
- Compressive strength test (ASTM D1074,2000)

5 Results and Discussion

5.1 Rheological tests

The rheological properties of PP/A binders were evaluated and the results are presented in Table (4).

The results indicate that PP is effective in improving the rheological properties of asphalt cement.

Examining Table (4a), it can be seen that PP is keeping the ductility values of 100^{+} up to 7% and raising the softening point of virgin asphalt by 33% at 7% wt. PP.

Table 4a. Rheological properties of PP/A binders.

% PP	Penetration (25 °C, 100g, 5s, dmm)	Ductility (25 °C, 5 cm/min)	Softening point °C	P.I.	% Loss of heat
0	51	150^{+}	52.5	-0.547	0.253
1	44	145	54.8	-0.365	0.232
3	38	136	58	-0.017	0.201
5	30	120.5	63.5	+0.519	0.175
7	26.8	103	70	+1.369	0.141
9	22.3	96.5	77	+2.024	0.120
11	16.5	79.0	83	+2.229	0.103

Table 4b. Rheological properties of PP/A binders, after Lobna[8].

% PP	Penetration (25 °C, 100g, 5s, dmm)	Ductility (25 °C, 5 cm/min)	Softening point °C	P.I.	% Loss of heat
0	42	150^{+}	54	-0.644	0.25
1	39	133	55.5	-0.478	0.21
2	35.5	117.5	57	-0.365	0.18
3	31.5	115	59	-0.118	0.15
4	29	101.5	65	+0.71	0.137
6	23.7	82	66	+0.475	0.121
8	21.5	74	75	+1.67	0.105

5.2 Temperature susceptibility

The penetration index relationship was used to investigate the influence of the addition of PP on temperature susceptibility of asphalt cement using the following equation mentioned by(Yang,1993):

$$\text{Penetration index (P.I.)} = \frac{20 - 500A}{1 + 50A}$$

$$A = \frac{\log(\text{pen.@ T}) - \log 800}{T - T_{R2B}}$$

where

T = Testing temperature

T_{R2B}= Ring and Ball softening point.

Figure 3 illustrates the relationship between the penetration index and PP contents. It can be seen that all modified binders are less susceptible to temperature changes than virgin asphalt cement.

5.3 Durability test

Figure (4) shows that the percent loss of PP binders decreases as the PP content increases. This is because PP was occupied a space of total mix and causes reduction in asphalt volume, which leads to a decrease in loss by dehydrogenation and oxidation of asphalt in the mix (i.e. durability increased slightly with the addition of PP content in the mix).

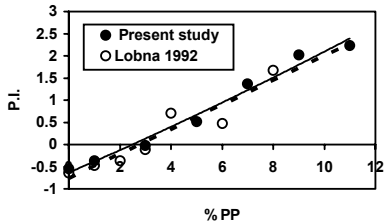


Figure 3. Relationship between PP content and P.I.

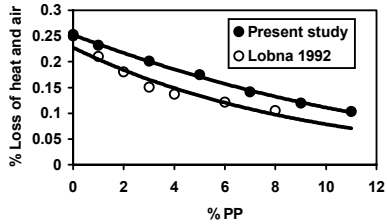


Figure 4. Relationship between PP content and %loss of heat and air

5.4 Compatibility test

Light microscopy was used to evaluate the compatibility between PP and asphalt. The microstructures of asphalt and modified asphalt with PP were observed on a light microscope and shown in Figure 5. Because of the identification photographs of four percent of PP which are 1, 3, 5 and 7% are therefore, one of them is listed and discussed here: The photomicrographs demonstrated that PP is well dissolved in the asphalt matrix at a temperature of 160 °C for all contents.

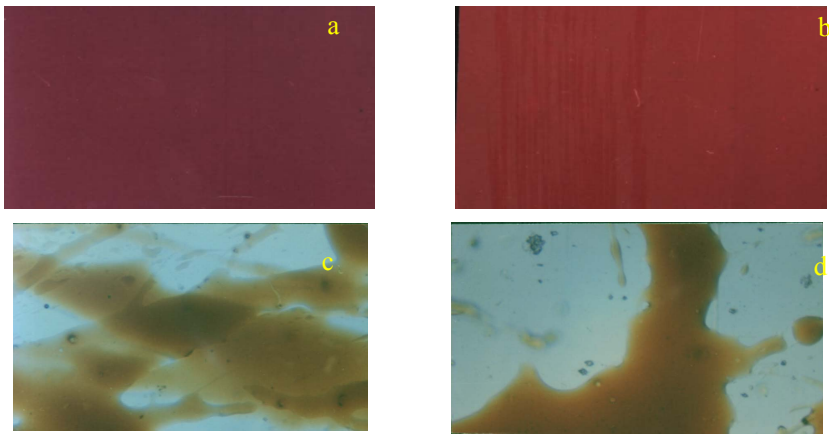


Figure 5. The photomicrograghs of PP modified asphalt(a=0%,b=7%,c=9%&d=11%)

5.5 Marshall, indirect tensile strength, and compressive strength tests

5.5.1 Marshall test

The relationship between Marshall properties and PP content were plotted at optimum binder content (i.e. 5.82% wt of total mix) as shown in Figure 6.

Figures 6a and b represents the effect of PP content on Marshall stability and flow. The figure indicates that as PP content increases the stability increases and the flow decreases up to 5% PP. This is attributed to the specific gravity of PP, which is less than that of pure asphalt, which serves to penetrate between particles and enhanced interlock of aggregates which increase stability and decrease flow. Beyond 5% PP content the stability decreases and the flow increases.

Figure 6c shows that air voids decreased with the addition of PP content due to the filling property attributed by PP addition.

Examining Table 5 indicates that 5% PP content satisfies the specified limits of 3-5% air voids and a drain down less than the minimum allowable of 0.3% mentioned by SC DOT specification(Bradley et al,2004)for AC 13 mm.

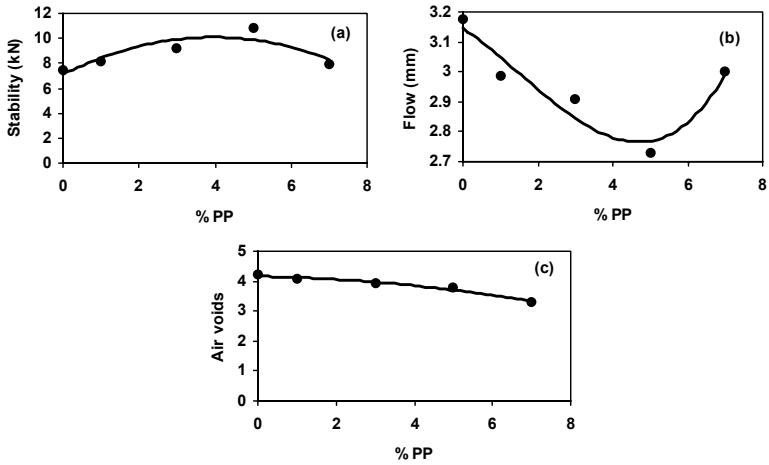


Figure 6. Relationship between PP content and (a) stability, (b) flow and (c) air voids.

Table 5. Marshall result of PP/A mixtures.

% PP	Stability kN	Flow mm	Rigidity ratio kN/mm	% air voids	% drain down
0	7.412	3.175	2.335	4.2	0.303
1	8.213	2.984	2.715	4.08	0.295
3	9.192	2.91	3.179	3.92	0.280
5	10.876	2.73	3.956	3.78	0.243
7	7.943	3.0	2.64	3.29	0.225

5.5.2 Indirect tensile strength test

Tensile strength for PP/A mixtures are given in Figure 7. The results indicate that tensile strength, tensile stiffness modulus and tensile strength ratios were increased, while the percent loss in tensile strength was decreased for both testing temperatures (i.e. increase the adhesion between aggregate and asphalt, which leads to decrease the stripping of SMA).

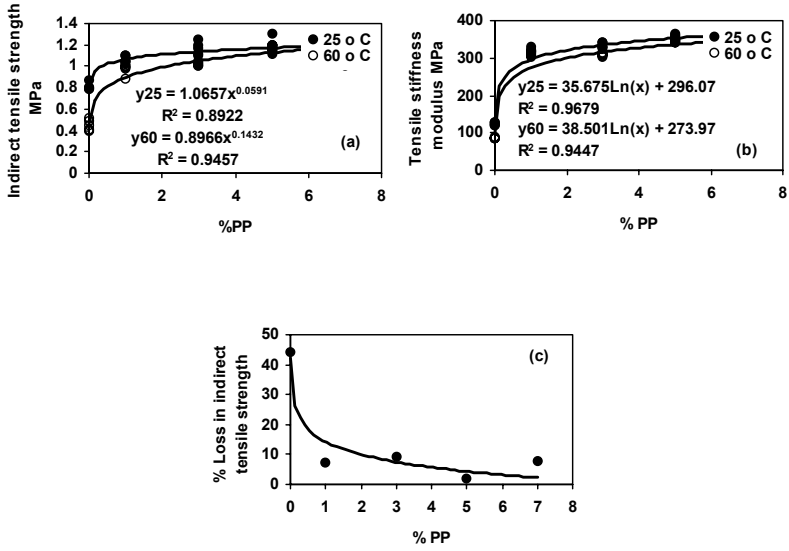


Figure 7. Relationship between PP content (a. tensile strength, b. tensile stiffness modulus and c. % loss in indirect tensile strength).

5.5.3 Compressive strength test

From the cited results in Figure 8, it can be seen that the compressive strength of modified SMA mixes is increasing with the increase of PP content.It was found that the index of retained compressive strength was increased 19% at 5% PP content.

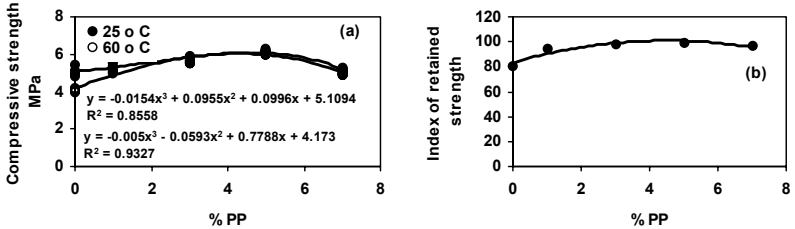


Figure 8. Relationship between PP content and (a) compressive strength (b) index of retained strength.

6 Conclusions

Based on the testing and analysis presented, the results of the study warrant the following conclusions:

A) Comparing the engineering properties for the pyrolisis polypropylene modified asphalt cement obtained by the machine manufacturing by authors with those obtained by modified binders with pyrolisis polypropylene from using (Lobna,1992) procedure, see Figures 3 & 4 and Tables (4a & b) .

B) A review of the standard binder testing and grading results indicated the following:

(1) Penetration at 25 °C generally will decrease as polypropylene content increases, which indicates an improved shear resistance in medium to high temperatures.

(2) The addition of polypropylene kept the ductility values at minimum range of ASTM and SCRB specifications (100^{+}cm) up to 5% PP content.

(3) The softening point tends to increase with the addition of polypropylene, which indicates improvement in resistance to deformation.

(4) The percent of loss of air and heat decreases with the addition of polypropylene (i.e. the resistance of asphalt to the action of temperature and temperature changes and the action of heat increases).

(5) The photomicrograph indicated that polypropylene and asphalt are compatible to extent; and.

(6) Penetration index values indicated that polypropylene was reduced the temperature susceptibility of asphalt.

C) A review of the Marshall and indirect tensile strength mixes design result indicated the followings:

(1) The pyrolisis polypropylene was effective in preventing excessive drain down of the SMA mixtures (i.e. bleeding phenomenon).

(2) The mixtures containing the polypropylene greatly exceeded the indirect tensile strength requirements set forth by the SC DOT.

(3) The tensile strength ratios for the mixes containing the polypropylene were greater than 85%. This indicates that this type of additive does not cause the mixture to weaken when exposed to moisture.

(4) The stiffness determined from the mechanical properties of Marshall and indirect tensile strength tests indicated that SMA modified with polypropylene may resist the pavement deformation forces, rutting and shoving, therefore it should be used on busy intersections or truck stops and parking lots where standing loads cause extended periods of such deformation; and

(5) Results indicated that achieving a flexible, more economic pavement with high performance and durability can be obtained with 5% of pyrolisis polypropylene.

(6) From statistical analysis, a significant correlation was found between PP content and each of indirect tensile strength, tensile stiffness modulus, percent loss in indirect tensile strength, compressive strength and index of retained strength.

References

Al-Hadidy A.I. (2006). "Evaluation of pyrolisis polypropylene modified asphalt paving materials", Al-Rafidain Engineering Journal, Vol. (14), No. 2: 36-50.

ASTM Standard Specification(2000). Section 4, Vol.04-03.

Boutevin B. Pietrasanta Y. Robin J.(1989). "Bitumen-polymer blends for coating applied to roads, and public constructions", Prog. Org. Coating; 17(3): 221-49.

Bradley J. Putman, Amirkhanian,(2004). "Utilization of waste fibers in stone matrix asphalt mixtures", Elsevier Journal Resources Conservation and Recycling 42.

Fred W. Billmeyer, Jr.(1971). "Textbook of polymer science", 2nd ed.

G.A. Wen, Y. Zhang, Y.X. Zhang. (2001) ."Journal of applied polymer science. 82(4),: 989-996.

Hailong Jin, Guangtao Gao, Yong Zhang, Yinxi Zhang, Kang Sun, Yongshong Fan. (2002). "Improved properties of polystyrene-modified asphalt through dynamic vulcanization". Polymer Testing 21: 633-640.

Ibrahim M.Asi,(2006). "Laboratory comparison study for the use of stone matrix asphalt in hot weather climates", Elsevier Journal, Construction and Building Materials 20.

Lobna A.S.(1992). "Studies of the changing of rheological properties of asphalt", M.Sc. thesis, College of Science, Chemistry Dept., University of Mosul.

Lottman R. and Dennis L.G.(1970). "Pressure-induced stripping in asphaltic concrete", Transportation Research Board. 340, pp13-25.

Lottman R,(1978). "Pressure moisture-induced damage to asphaltic concrete. NCHRB, Report No. 192, TRB.

Lu X, Isacsson U.(2001) "Modification of road bitumens with thermoplastic polymers". Polymer Testing; 20(1): 77-86.

State cooperation of road and bridges (SCRB)(1999). "Hot mix asphalt concrete pavements" Iraqi standard specifications, Ministry of Housing and Construction, Department of Design and Study, section R9.

M.S. Cortizo, D.O. Larsen, H. Bianchetto, J.L. Alessandrini.(2004). "Effect of the thermal degradation of SBS copolymers during the aging of modified asphalts". Polymer Degradation and Stability 86: 275-282.

S.A.M. Hesp, R.T. Woodhams(1991)." Colloid polymer and science", 269: 825.

Yang H.H.,(1993). "Pavement analysis and design", University of Kentucky. Jones Wiley and Sons, Inc. pp. 336, 337, 409 and 410.

PERPETUAL PAVEMENTS: APPLYING THE APPROACH TO VERY HEAVY LOADS

Xishun Zhang Ph.D.[1], Gerald A. Huber[2], David H. Timm Ph.D.[3] and Yang Yongshun

[1]Heritage Research Group, 7901 West Morris Street, Indianapolis, IN 46231; PH (317) 390-3147; FAX (317) 486-2985; email: xishun.zhang@heritage-enviro.com
[2] Heritage Research Group, 7901 West Morris Street, Indianapolis, IN 46231; PH (317) 390-3141; FAX (317) 486-2985; email: gerald.huber@heritage-enviro.com
[3] National Center for Asphalt Technology, 238 Harbert Engineering Center, Auburn University, AL 36849; PH (334) 844-6282; FAX (334) 844-6290; email: timmdav@eng.auburn.edu
[4] Shandong DOT Highway Administrative Bureau, 29 Shun Geng Road, Jinan, Shandong 250002, P.R.China; PH (86) 531 82973642

Abstract: The concept of perpetual pavements has been well documented in Europe and North America for several years. This paper presents an experiment to design and construct perpetual pavement sections to carry very heavy loads with a design single axle of 19 tonnes. The main objective is to construct pavement sections designed using different methods, to measure their response and predict performance.

The experiment has three objectives:

1. Evaluate structural design methods.
2. Evaluate asphalt mix properties to carry heavy loads.
3. Characterize traffic loads being applied to the pavement.

Five pavement sections were built on a new motorway in Shandong Province, China. One section was designed using a tensile strain at the bottom of the pavement of 70 microstrain under the design load. Another was designed using the same principle except 125 microstrain was used as the limiting strain. A third section also used 125 microstrain but was constructed with mixtures that are more strain tolerant. Two composite sections were built with stabilized soil and aggregate layers under hot mix asphalt.

All sections contain embedded instrumentation to evaluate pavement response under known loads. A weigh-in-motion scale will categorize the number and types of trucks as well as the axle loads to which the pavement is being subjected.

This paper presents the design and construction of the experiment and the measured response of the pavements to known loads. A methodology is also presented to estimate the expected life of the pavement sections.

1 Introduction

1.1 Motorways in China

The concept of perpetual pavements has been accepted by the asphalt paving community during the last ten years, which stated simply means that if the pavement layers are not subjected to strains above a critical value the pavement will not fail

structurally. Only the surface will require renewal every 15 to 20 years but the pavement structure will remain intact for 50 years or more.

The Peoples Republic of China is currently building a motorway system. Until 2005 China's Ministry of Communications has constructed 40,000 km of motorway in China and by 2015 the network is expected to contain more than 80,000 km. As of 2005 Shandong Province had built 3100 km of motorway and by 2015 the total network in the province will be nearly 5000 km (Yang, 2006).

Heavier than expected truckloads has caused early structural failure of some pavements. The Highway Bureau in Shandong Province has reported that 40 percent of the truckloads are overweight with single axle loads as high as 300 kN (67,000 lbs.) (Yang, 2006) As a result, the Provincial Highway Bureau is interested to evaluate the concept of perpetual pavements as applied to the heavy loads used in China.

1.2 Concept of Perpetual Pavements

Historically pavement design in North America has been based on the principle of increased thickness for increasing repetitions. The core of the AASHTO structural design method is rooted in the AASHO test road of 1959-61 in which approximately one million repetitions of a 40 kN (9000 lb) axle were applied to pavements of various thicknesses. The relationship of fatigue damage to thickness for different numbers of axle passes allowed design charts to be developed. For each level of subgrade support, these charts require thicker pavements to be used for increasing repetitions of a 40 kN axle (Equivalent Standard Axle Load, ESAL.)

Although the basis for the AASHTO design charts was limited to one million repetitions, highways today are commonly designed for more than 30 million repetitions over a 20-year design life. Some pavements are designed for more than 100 million repetitions. Subsequent editions of the AASHTO design procedure have extrapolated the thickness versus traffic charts assuming that the relationship from the AASHO test road continued to be valid as repetitions increased beyond one million.

In the 1990s the basis of the AASHTO design method came under scrutiny. Similar to other materials, hot mix asphalt was known to experience fatigue cracking related to the amount of applied strain. Metals, such as steel or aluminum experience fatigue in the same way, but it is known that these materials require a minimum strain level to be applied before fatigue damage occurs. If applied strains were kept below this threshold, the metal could sustain an unlimited number of repetitions.

For hot mix asphalt the strain threshold is believed to be approximately 70 to 90 micros-strain. Currently research is underway in the USA under the National Cooperative Highway Research Program exploring the concept of threshold values for hot mix. Until research indicates what the threshold is and the conditions that influence the threshold, industry has accepted 70 microstrain as the threshold value.

2 Experiment Objectives

This experiment has two main objectives: (1) To evaluate structural design methods, and (2) To characterize the spectrum of traffic loads.

This research project evaluates four different structural designs. Two of the methods use Full Depth Asphalt (FDAC) pavement design and two of the methods use

composite design. The two composite designs are based on hot mix asphalt (HMA) over pozzolanic bound layers (lime-flyash and lime-flyash-cement).

Usual mixes for motorways in Shandong Province include an SMA surface and dense graded mixtures designed using the Marshall method of mix design. In this research the mixtures are designed using the Superpave method of mix design. SMA and dense-graded mixtures that are similar to Chinese specifications are designed to meet both the Marshall and Superpave specifications. Two new mixtures are also evaluated. Drainage asphalt (a mix with 20 percent air voids) is used as a drainage layer in the pavement. At the bottom of the hot mix pavement structure a fatigue-resistant layer was used.

This experiment will obtain the number and weights of axles on the motorway as well as the types of trucks. The results will define the spectra of axle weights, that is, for each type of axle (steer, single, tandem or triple) there will be a number of axles within each weight category.

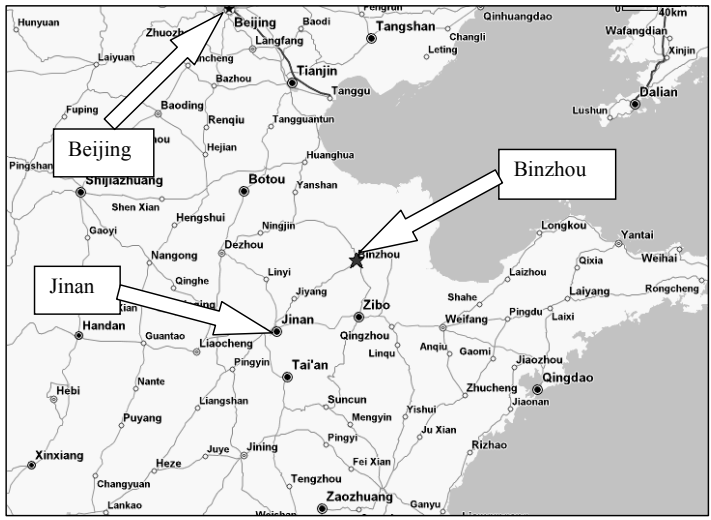
3 Binzhou Experiment Description

3.1 Location Description

This experiment was built near the city of Binzhou in Shandong Province as shown in Figure 1. Shandong Province is located in the east central part of China. Binzhou, located on the banks of the Yellow River, 80 km from the Yellow Sea, is about 300 km southeast of Beijing and 150 northeast of Jinan, the capital of Shandong Province. The topography of the area is very flat with little vertical relief and no discernable river valley.

The soil is composed of river-deposited silt with a low plasticity index (approximately eight) and a CBR of eight. The water table in the area is high, about two meters below the ground surface. The pavement is built on a roadway embankment that is a minimum of three meters above the surrounding land.

Figure 1. Location of Binzhou Perpetual Pavement Experiment



3.2 Location of the Experimental Sections

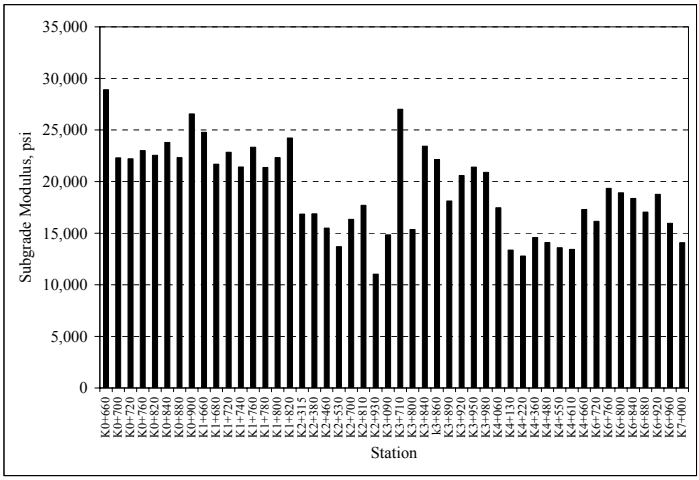
The motorway is a four-lane, divided highway with controlled access running north to south just to the west of Binzhou. This experiment contains five sections, four experimental and a control section of the originally designed pavement. All the sections are built on the northbound lanes with both lanes built using the experimental sections. The length of the sections was selected based on the distance between bridges that are convenient locations to separate the sections. Average length of the sections is about one kilometer, with the shortest being 0.7 km and the longest 1.7 km.

3.3 Material Properties

Embankment density specifications call for 90% of Proctor density in the lower layers (more than 130 cm from the top of the embankment.) The next 50 cm require 93% density and the top 80 cm require 95% density. The top 20 cm of subgrade is stabilized with 8% limekiln dust.

Falling weight deflectometer data was used to evaluate the subgrade for pavement design. Testing was conducted directly on the prepared embankment at loads ranging from 8.8 to 35.6 kN. The standard AASHTO (AAHSTO 1993) approach using the outermost deflection to estimate soil stiffness (Mr) was employed, and the results are summarized in Figure 2. The data followed an approximately normal distribution with an average of 132 MPa. Since the data were from one testing date and no information regarding seasonal changes was available, a number of potential designs were developed using various values of soil stiffness. Specifically, designs were developed for soil stiffness of 48 MPa and 103 Mpa (7,000, and 15,000 psi).

Figure 2. Subgrade Modulus Measured with Falling Weight Deflectometer



For the hot mix asphalt layers, little advance information was available regarding the stiffness. Therefore, assumptions were necessary. Through team discussions and review of relevant data sets such as those from the National Center for Asphalt Technology (NCAT) Test Track, the design stiffness for HMA was set at 4,826 MPa. Similar materials tested at the NCAT Test Track (World Climate 2006) had an average stiffness, from backcalculation, of approximately 5,860 MPa at 20C. Given that the average annual temperature at the project location is 14.8C (Timm 2006a), the design value was considered conservative yet reasonable.

3.4 Design Traffic Loads

The highway at Binzhou was a new route. Therefore, other similar highways were used to estimate traffic data. A traffic load spectrum from another route was used for design of the Binzhou sections. The design load was selected to be the ninetieth percentile load. The weight for each axle group is shown in Table 1.

Table 1. Design Loads for Each Axle Group

Axle Group	90^{th} Percentile	Axle Group	90^{th} Percentile
Single	150 kN	Triple	600 kN
Double	340 kN		

4 Experiment Design

4.1 Design Methodology

The concept of perpetual pavement design is based on limiting strains in the pavement layers. Traditionally, limiting the horizontal strain at the bottom of the HMA layer controls fatigue cracking and limiting the vertical strain at the top of the subgrade controls structural rutting (Newcomb 2001). Previous laboratory and field research have conservatively estimated fatigue and rutting thresholds of 70 □□ and 200 □□, respectively (Monismith 1972, 1999). Other studies have suggested that the fatigue threshold may be significantly higher; horizontal tensile microstrain exceeding 150 □□ Nishizawa 1997). Based upon these previous efforts, it was decided to develop designs at two theoretical fatigue strain thresholds, 70 □□ and 125 □□, and use the results of the experiment to help guide future perpetual pavement designs. The vertical strain criterion was set at 200 □□; consistent with a previous field study (Monismith 1972).

4.2 Mechanistic Empirical Design

The data described above were used in the perpetual pavement design software, PerRoad 2.4. This mechanistic-empirical (M-E) pavement design program uses layered elastic analysis and Monte Carlo simulation to develop stochastic pavement designs. The computational algorithms within PerRoad are beyond the scope of this paper, but can be found in other literature (Timm 2005, 2004a, 2006b).

Using PerRoad 2.4, multiple designs were developed to keep the critical pavement responses below the respective thresholds 90% of the time. In other words, a small percentage of "overloads" were allowed in the design. It should be noted that a laboratory fatigue study conducted by Thompson and Carpenter

(Thompson 2004) indicated that a small proportion of overloads exceeding the strain threshold did not alter the existence of the strain threshold.

Two sets of material properties and failure criteria were used in developing the designs. The first set was a conservative design using a soil stiffness of 48 MPa with the fatigue threshold at 70 □□. This design resulted in a 500 mm (20 in.) thick pavement layer. The second set of conditions, using a soil stiffness of 103 MPa and a fatigue threshold of 125 □□ yielded a less conservative design thickness of 380 mm (15 in.) The rutting criteria of less than 200 □□was met in each case, and did not control the design of these sections.

4.3 HMA Mixture Selection

The asphalt mixtures were selected to achieve rut resistance in the upper portions of the pavement and fatigue resistance in the lower portions. SMA and dense-graded Superpave mixtures were used in the upper part of the pavement.

Each experimental section featured a bottom HMA layer as a fatigue-resistant dense-graded material with the asphalt content 0.6% above optimum. Based upon previous studies (8), it is believed that the added asphalt will improve overall fatigue resistance of the pavement where the tensile strains are the highest. One of the sections designed with a threshold strain of 125 □□ used a polymer modified-binder to see if the polymer modified asphalt binder was better able to withstand higher strain than unmodified asphalt binder.

The FDAC designs and one of the composite designs included an open-graded drainage layer to help remove moisture and mitigate stripping problems. Moisture damage (stripping of asphalt binder from the aggregate) has been observed in several failure investigations of pavements in Shandong Province that had been in service for 5 to 8 years. Reducing moisture infiltration or providing an outlet for internal pavement drainage has shown benefit in reducing moisture related distress. Therefore, the drainage layer was included in the perpetual pavement sections. The open-graded drainage layers are daylighted to the outside shoulder rather than the typical U.S. method of utilizing pipe underdrains. The high embankment height and steep sideslope make daylighting a practical option.

4.4 Final Pavement Designs

The final FDAC pavement designs are shown in Figure 3. The more conservative design is labeled “70 □□” and the less conservative one is labeled “125 □□”. The section containing polymer-modified asphalt binder in the fatigue layer is labeled “125 □□ (Modified)”.

The final composite designs are shown in Figure 4. The section labeled “Section 5” is the original design for the motorway and will act as the control section for the experiment. Section 4 is an alteration of the standard pavement design in which the lime-kiln dust-flyash stabilized layer has been reduce in thickness and replaced with HMA. Section 4 has 150 mm of dense-graded HMA. In Section 5 this thickness is increased to 180 mm. In addition, 150 mm of open-graded drainage layer is placed between the dense-graded mixtures and the lime-kiln dust-flyash stabilized granular layer.

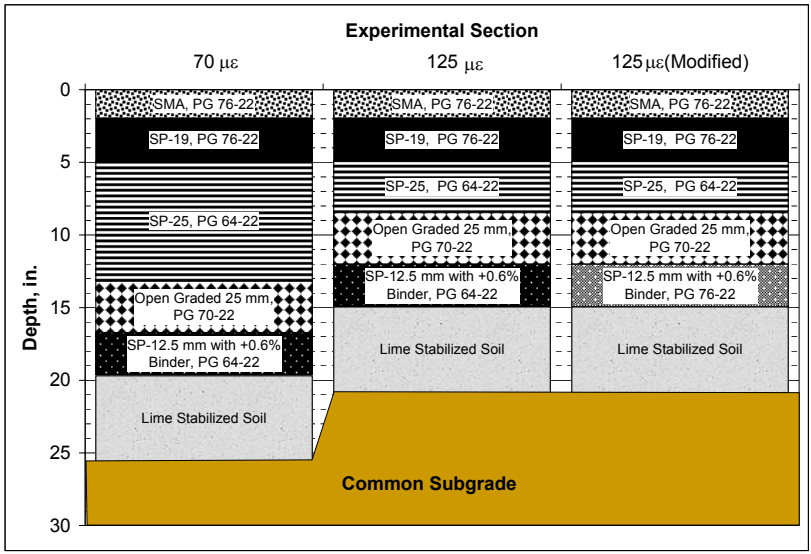


Figure 3. Layer Type and Thickness in Full Depth Asphalt Experiment Sections

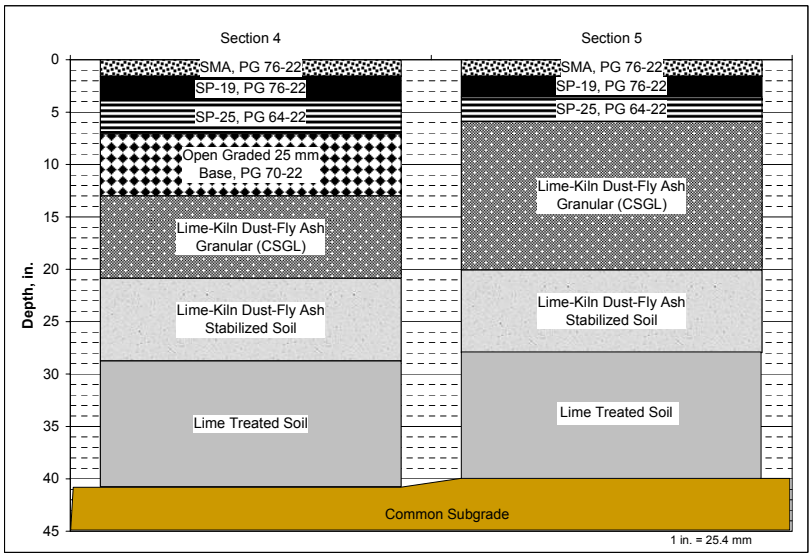


Figure 4. Layer Type and Thickness in Composite Experiment Sections

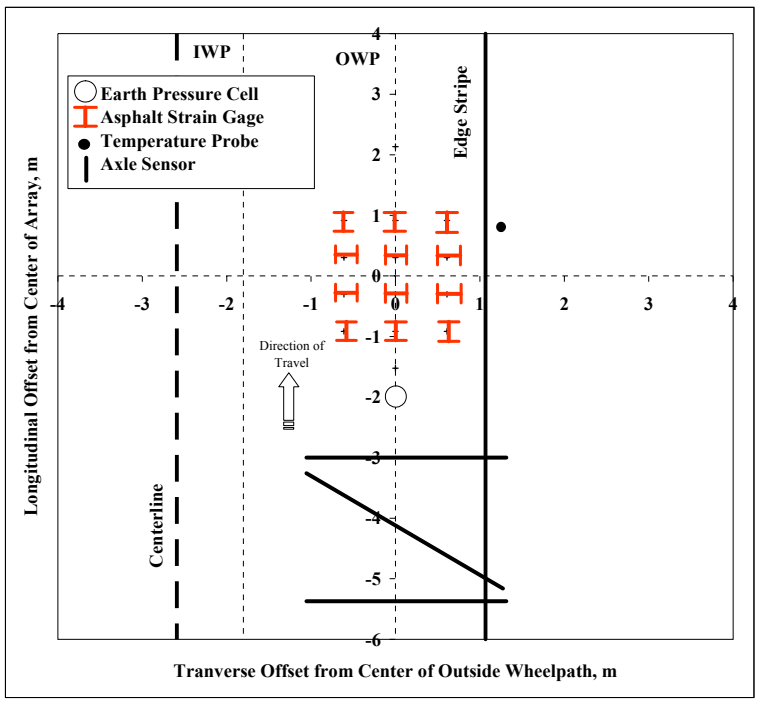
5 Instrumentation and Construction

Instrumentation was embedded in the roadway to enable pavement response measurements under the expressway traffic. The instrumentation layout, shown schematically in Figure 5, was patterned after the scheme used at the NCAT Test Track (Timm 2004b). Axle sensors are used to pinpoint the location of loads as they move across the gauge array in the outside wheelpath (OWP). The earth pressure cell measures vertical stresses on top of the lime-stabilized soil (experimental sections) and CSGL (control sections) and will support efforts to characterize limiting stress/strain criteria.

The asphalt strain gauges measure horizontal strain, in the lateral and transverse directions, at the bottom of the HMA and are spaced to capture the natural wander of vehicles. These gauges are important to establishing critical strain thresholds for fatigue and evaluating the design methodology. A temperature probe was installed vertically in the pavement to capture temperature at the top, middle and bottom of the HMA, which helps develop relationships between environmental parameters, material properties and pavement response. Not shown in Figure 7 is a WIM station that was installed at the end of the project to document live traffic volume, weights, and axle configurations.

Construction of the test sections occurred in June 2005. The HMA mixtures were produced at a new batch plant located near the test sections. Rolling patterns were established on trial mixes and two pavers were used in tandem to pave widths up to 40 ft (12.2 m). During construction, HMA samples were obtained from the delivery trucks and shipped to the U.S. for further testing.

Figure 5. Embedded Pavement Instrumentation



6 Data Collection

Due to the nature and length of this experiment, the data collection efforts are a massive undertaking. Collection and analysis of WIM, FWD and dynamic response data through the life of the test sections (or sensors) is a major component of the study. Laboratory material testing, continuous pavement performance monitoring, and forensic investigations as the sections begin to show distress are also vital. Further, all of the mentioned collection efforts are particularly challenging because the experimental sections are segments of a public access expressway.

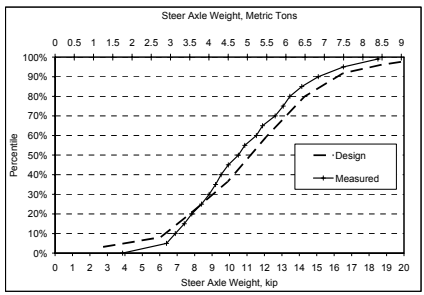
6.1 Traffic Loads

Axle loads and vehicle classification are done by the Weigh-in-Motion. Measure load spectra for each of four axle types is shown in Figure 6 along with the design load spectra. For steer axles and single axles the measured weights typically are less than the design weights. For the single axle the measured 90th percentile is less than the design but note that the heaviest loads are considerably higher than the highest design load (49 metric tons vs. 29 metric tons.)

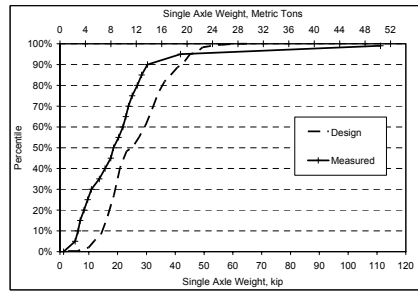
For the tandem and tridem axle groups, the 90th percentile loads are higher as measured than as designed. For the tandem axles the measured load is 43 metric tons compared to the design of 32 metric tons. Likewise, the measured tridem axle load is 92 metric tons compared to the design of 60 metric tons.

In general, the measured weight data shows that there are more of the lighter load trucks than had been assumed for design and that the heavily loaded axle groups have higher weights than expected.

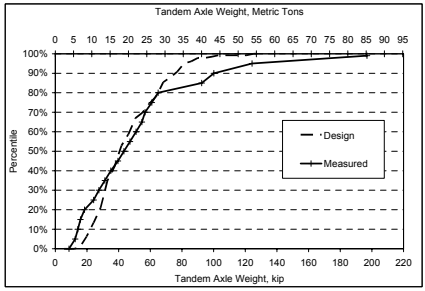
Figure 6. Axle Load Spectra



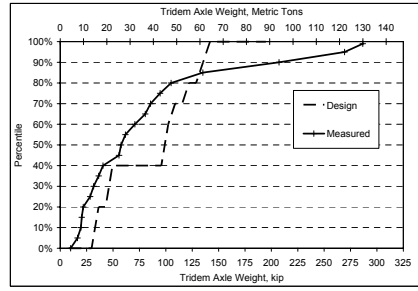
a) Steer Axle



b) Single Axle



c) Tandem Axle

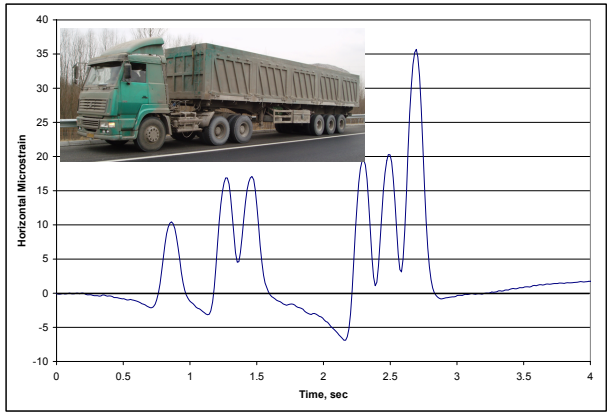


d) Tridem Axle

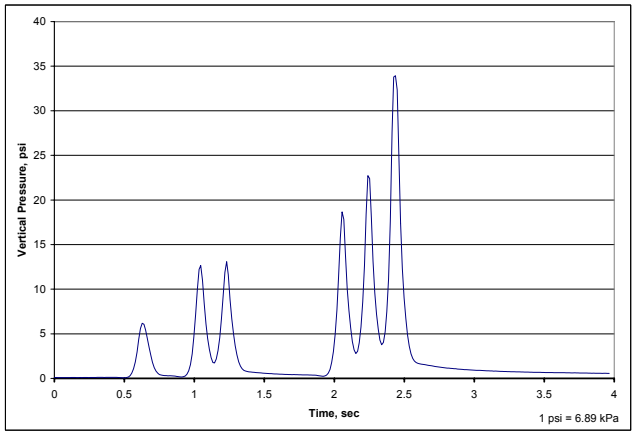
6.2 Measured Strain and Vertical Pressure

Dynamic pavement instrumentation is a powerful tool used to measure the in situ pavement response, which allows for the direct evaluation of design and comparison between different pavement cross-sections and materials. The sensors and data acquisition used for the project are capable of recording live pavement response in excess of 2,000 samples/sec per sensor. At these high collection rates, the responses show not only the peak response, but also the nature of the response as a truck travels over the pavement. Figure 7 shows an example of the data collected from the asphalt strain gauges and earth pressure cells, in addition to the test vehicle. The steer, tandem and tridem axles are clearly discernible with the tridem axle generating the highest strain and pressure responses.

Figure 7. Pavement Dynamic Response to Vehicle Loading



a) Horizontal Strain



b) Vertical Pressure

7 Data Evaluation Methodology

The main focus of this experiment is the long-term performance of the experimental sections. The pavement test sections in this experiment are located on a public highway and therefore the loading is not controlled. Loads and their effect on the pavement structure are evaluated in two ways. First, on a routine basis, controlled loads will be passed over the instrumented pavements. Strain and pressure response is measured under the load. Second, weights of the live traffic will be monitored using the weigh-in-motion scale.

7.1 Use of Controlled Load Testing

The response to controlled loading will be used for several purposes:

1. Compare response among the test sections
2. Validate structural design methodology
3. Predict structural performance of the test sections
4. Enhance theoretical models
5. Characterize seasonal and temperature effects on pavement response
6. Evaluate load – response interaction

7.1.1 Compare Response among the Test Sections

Horizontal strain and vertical pressure are influenced by the visco-elastic response of the pavement layers. Therefore, the pavement response will depend on pavement temperature and vehicle speed. Controlled load testing will be used to develop response characteristics of each experiment section.

Table 2. Load Levels Used for Controlled Load Testing.

Load Level	Single Axle	Tridem Axle
1	9000 kg	42,000 kg
2	12,000 kg	48,000 kg
3	15,000 kg	54,000 kg
4	18,000 kg	60,000 kg

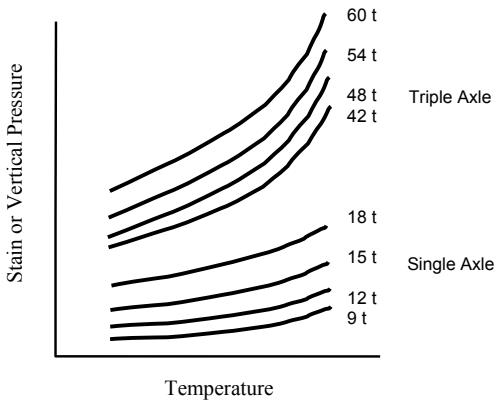


Figure 8. Experimental Section Response to Load and Temperature (Conceptual)

The US-China team decided to develop response charts for single and tridem axle groups. Controlled vehicles will be driven at 70 kph over the gauges. Axle loads are listed in Table 2.

Based on the results of the controlled load testing it will be possible to develop figures such as that shown in Figure 8 for each of the experimental sections. By comparing these curves it will be possible to compare the dynamic response among the experimental sections.

7.1.2 Prediction of Structural Performance for the Experimental Sections

The first step to predicting structural performance of the experimental sections is to estimate the strain spectra experienced by the experimental sections. The weigh-in-motion system will measure all of the applied axle loads and the pavement temperature when the loads passed over the strain gages. Using the strain response charts it is possible to estimate the strain that occurred and construct a strain spectra such as that shown in Figure 9.

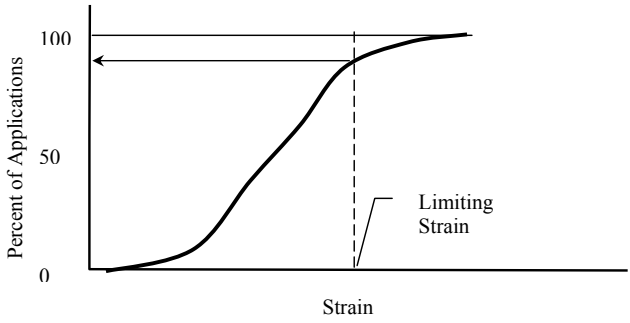


Figure 9. Experimental Section Strain Spectra (Conceptual)

Using the strain spectra curves for each experimental section the percentage of strains above a limiting strain can be determined. As results from on-going research into limiting strain becomes available, it will be possible to evaluate the sections with the most current information.

By comparing the strain spectra for each of the experimental sections is will be possible to estimate the relative fatigue performance of the sections that contain the same mix in the bottom layer of the pavement structure. Four of the five experimental sections have the same high asphalt content, fatigue layer at the bottom and therefore, the relative performance of these four sections can be estimated.

The fifth experimental section contains polymer-modified asphalt binder. To obtain an estimate of fatigue performance for this section, it is necessary to perform

some laboratory fatigue testing. Beam fatigue testing is being done on the two different fatigue layer mixtures: one containing unmodified asphalt binder and the other containing polymer-modified asphalt.

8 Experiment Status

The Shandong Perpetual Pavement Project was constructed in the summer of 2005 and was opened to traffic in December 2005. Data collection efforts with the controlled loading have occurred and are being analyzed. Monitoring of live-traffic conditions is in progress. Laboratories in Shandong Province and in the U.S.A. are performing tests to determine fatigue and stiffness properties. This testing will be completed during the winter of 2006-7.

9 Conclusions

This paper describes the design and construction of experimental sections on a new motorway near Binzhou in Shandong Province, Peoples Republic of China. The main purpose is to evaluate structural design methods and perpetual pavement concepts for heavy truckloads. Four different structural designs and two different fatigue resistant mixtures are being tested. Traffic loading is very heavy as compared to North American loads, approximately double.

This paper describes the instrumentation used to monitor the pavement response to load and presents pavement response measurements. Actual truck weights were found to be less on average than designed for but the heaviest of the truck loads were found to be heavier than the 90th percentile used as the design value.

Also, this paper describes the methodology that will be used to compare the experimental sections and determine if the sections will behave as perpetual pavements.

10 Acknowledgements

This project is the result of efforts by partners from the People's Republic of China and the United States of America. The Shandong Highway Bureau is the major supporting partner in China. The Shandong Highway Research Institute and the Shandong Highway Engineering Group are supporting partners in China. In the U.S.A., the Indiana Department of Transportation initiated this experiment as part of a joint INDOT-Shandong Highway Bureau cooperation agreement. The Federal Highway Administration, the National Center for Asphalt Technology and the Heritage Research Group are supporting partners in the U.S.A. This international cooperation will lead to a better understanding of HMA perpetual pavement performance under heavy truck loading.

References

American Association of State Highway and Transportation Officials (AASHTO), 1993. AASHTO Guide for Design of Pavement Structures. Washington, D.C., 1993.

Monismith, C. L., & McLean, D. B. 1972, Structural Design Considerations. Proceedings of the Association of Asphalt Paving Technologists, Vol. 41, pp. 258-304.

Monismith, C.L. & Long, F. 1999, Overlay Design for Cracked and Seated Portland Cement Concrete (PCC) Pavement – Interstate Route 710. Technical Memorandum TM UCB PRC 99-3, Pavement Research Center, Institute for Transportation Studies, University of California, Berkeley.

Nishizawa, Shimeno, T., S., & Sekiguchi, M, 1997, Fatigue Analysis of Asphalt Pavements with Thick Asphalt Mixture Layer. Proceedings of the 8th International Conference on Asphalt Pavements, Vol. 2. University of Washington, Seattle, WA, August, pp. 969-976.

Newcomb, D.E., Buncher, M. & Huddleston, I.J. 2001. Concepts of Perpetual Pavements. Transportation Research Circular No. 503, Transportation Research Board, pp. 4 – 11.

Thompson, M. R. & Carpenter, S. H. 2004, Design Principles for Long Lasting HMA Pavements. Proceedings International Symposium on Design and Construction of Long Lasting Asphalt Pavements, International Society for Asphalt Pavements, Auburn, AL, pp. 365-384.

Timm, D.H. & Young, J. 2004a, The Effects of Load Spectra and Variability on Perpetual Pavement Design. International Symposium on Design & Construction of Long Lasting Asphalt Pavements, International Society for Asphalt Pavements, Auburn, AL, pp. 131-152.

Timm, D.H., Priest, A.L. & McEwen, T.V., 2004b, Design and Instrumentation of the Structural Pavement Experiment at the NCAT Test Track. Report No. 04-01, National Center for Asphalt Technology, Auburn University.

Timm, D.H., Turochy, R.E. & Bower, J.M.. 2005, Effect of Load Spectra on Mechanistic-Empirical Flexible Pavement Design. Transportation Research Record, Transportation Research Board, in press.

Timm, D.H. & Priest, A.L., 2006a. Material Properties of the 2003 NCAT Test Track Structural Study. Report No. 06-01, National Center for Asphalt Technology, Auburn University.

Timm, D.H. & Newcomb, D.E. 2006b, Perpetual Pavement Design for Flexible Pavements in the U.S. International Journal of Pavement Engineering, Vol. 7, No. 2, pp. 111-119

World Climate. 2006, Climate Data for 36°N 116°E – Jinan. http://www.worldclimate.com/cgi-bin/grid.pl?gr=N36E116, accessed May 12, 2006.

Yang, Y.; Gao, X, Lin, W; Timm, D.H., A.L. Priest, Andrewski, D.A.; & Huber, G.A., 2006, Perpetual Pavement Design in China, Proceedings International Conference on Perpetual Pavements, Columbus, Ohio.

EXPERIMENT ON EFFICIENCY OF NANO TITANIUM DIOXIDE PHOTOCATALYTIC MATERIALS DEGRADING FACTORS OF AUTOMOTIVE EMISSION

Qiang GUAN[1], Zhanyu WANG[2], Haiying BAI[3]

[1]College of Transportation, North-east Forestry University, 26 Hexing Road, Harbin, 150040, P R China, PH (086)451-82191004; Email: jtxygq@nefu.edu.cn
[2]College of Transportation, North-east Forestry University, 26 Hexing Road, Harbin, 150040, P R China, PH (086)451-82191004; Email: zywang77@yahoo.com.cn
[3]Heilongjiang Provincial Communications Department, 72 Heping Road, Harbin, 150040, P R China, Email: baihaiying@mail.hljjt.gov.cn

Abstract：A test chamber with degrading materials on automotive emissions is developed to test the degradation efficiency of automotive emission factors with four different contents: 2%, 3%, 4%, 5% of nano titanium dioxide photocatalytic materials made by dispersed mechanism methods, which are mixed round with high shear, ultrasonic wave and ultrasonic wave dispersant. The sample is the emission when the vehicle is on idle. The average degradation efficiency of different dispersed nano titanium dioxide materials on NO_x, HC, and CO are calculated according to the equation of degradation efficiency. The experimental results show that the photocatalytic materials with four contents of nano titanium dioxide, made by different dispersed methods, have certain influence on NO_x, HC, CO and the trends of the degradation are the same. Nano titanium dioxide made by ultrasonic dispersant has the best degradation efficiency on NO_x of automotive emission. To the nano titanium dioxide made by high shear, ultrasonic wave and ultrasonic dispersant, the optimum content of nano titanium dioxide of degrading NO_x, HC, CO on automotive emission is 5%, 3%, 2%; 5%,3%,4%; 5%, 2%, 3% respectively.

Keywords: *transportation environment, automotive emission, degradation efficiency, nano titanium dioxide*

1 Introduction

Having the characters of nonpoisonous, steady light chemical property, effective photocatalysis, strong oxidized ability and high photocatalytic degradation efficiency, nano titanium dioxide has been applied in the field of catalysis frequently in recent years. It is an uncomplicated technology to degrade the environmental pollutants using Nano titanium dioxide photocatalysis with the low cost (ZHANG Wan-zhong, 2006). Nano titanium dioxide photocatalytic material is a late-model material to apply in the field of traffic environmental protection. Foreign scholars regard Nano titanium dioxide as the main element of the coating or the gushing pharmaceutical while degrading harmful factors in the air (LIU Shou-Xin, 2005). Certain research in degradation of NO_x with photocatalytic material have been done both at home and

abroad. IBUSUKIT (IBUSUKIT, 1994) have reported the research in this respect at first, and TAKAMIK (TAKAMIK, 1999) have made Nano titanium dioxide into photocatalytic coating to purify NO_x and measured the effect of it in the urban atmosphere, but the test that was carried on in the atmosphere received great impact from external factors. In Japan, the researchers have already developed environmental protection materials, which can coat tunnels, bill-boards, and railings on both sides of the road to degrade automotive emissions. Additionally, in the past, the researchers mostly focused on the research of a single gas, such as benzene, formaldehyde, etc. But there are hardly any reports on the study of how the degradation efficiency of different contents of Nano titanium dioxide materials, which are obtained from high shear, ultrasonic wave and ultrasonic dispersant methods, affects automotive emissions. In the light of the current situation of Nano titanium dioxide photocatalytic materials applied in the market, the experiment system is developed to test Nano titanium dioxide of 2%, 3%, 4%, 5% content in real time. The differences among the degradation efficiency of different content Nano titanium dioxide, made by different dispersed methods, are analyzed with the test data. The cause of automotive emissions usually are NO_x, CO, HC, CO_2, O_2, etc. Based on the contamination of factors, NO_x, HC and CO are harmful to humans; therefore, the degradation efficiency of NO_x, HC and CO of Nano titanium dioxide is studied in this paper.

2 Experiment System and Experiment Materials

2.1 Experiment System

According to the ASTM D 5116-1997 (Standard Guide for Small-Scale Environmental Chamber Determinations of Organic Emissions from Indoor Materials/Products) and the principle of photocatalysis, the experimental system is composed of a degradation test chamber, simulated light source and NO_x, HC, CO, temperature, pressure sensor and ADAM4107+ module, RS-232/RS-485 interface converter, computer, etc. Using the Configuration software Kingview 6.5, the functions, such as high speed inquiry and memory of the historical data of one millisecond and the dynamic test on the sample, can be achieved. The external diameter of the degradation test chamber is 408mm. The height is 1020mm. Two terminal surfaces are oval: one of 2:1, welded with HP295 Hot-rolled plate. The thickness of the degradation test chamber shell is 4mm. The testing system is equipped with a HID-H7 Xenon lamp simulating sunlight. Therefore, the photocatalytic reaction could take place under the circumstances of the xenon lamp shining (HU An-zheng, 2001), and the factors of the automotive emission are degraded. The variable value of concentration of each harmful factor in automotive emission is measured by a computer assisted test system, which can be used to analyze the degradation efficiency of Nano titanium dioxide (HAN Xiang-chun, 2005). Equipped with a circulation fan, it could promote automotive emissions flowing in the degradation test chamber. The system had real-time character, the function of showing and controlling; it presents the parameters during the process of measuring, drawing out the corresponding curve, and

showing the results of measuring. The flow chart of the test system is shown in Figure 1.

2.2 Experiment Materials

The CYC-1 Nano titanium dioxide is produced by Shenzhen Cheng-yin Technology Co., Ltd. It is white ultra-fine powder in small particle size (average crystal size is no more than 30nm), and large specific area. The special composite structure of the powder can help separation and transportation of the photo carrier. Therefore, it has a high level of photocatalytic activity and a long lifespan. The Nano titanium dioxide is dispersed in the Matrix material with high shear, ultrasonic mix and ultrasonic dispersant mix dispersed methods (BAI Hai-ying, 2005). The photocatalytic materials, which contain 2%, 3%, 4% and 5% Nano titanium dioxide respectively, are coated uniformly on the surface of the test glass carrier. The size of the glass carrier is 150mm*180mm*3mm. Then, the photocatalytic pellicle samples are made. It is the matrix material that has acid-base resistance, swagging resistance, aging resistance, strong adhesion, mildew preventive and is contamination free. The matrix material can be brushed and sprayed as coating which feels smooth. The Ingredients of the Matrix Material are as shown in table 1.

Table 1. Ingredients of Matrix Material

Material	Content, %	Function	Material	Content, %	Function
Styrene-Acrylic Emulsion	55 to 65	Base -material	Calcium Carbonate	15	Filling-material
Primal 1020PR	2.0 to 3.0	Leveling	Anti-mildew agent	0.1	Anti-mildew, Disinfect
Texanol	0.7 to 1.5	Filming	Ammonia	0.3 to 0.4	Regulate PH
Tamol 731	0.5 to 1.5	Disperse	Water		Disperse
General Defoamer	0.5	Anti-foam	Thickener	0.3 to 1.0	Increase viscosity
SPA-202	0.2 to 0.4	Anti-foam			

3 Degradation Experiment on Factors of Automotive Emission of Nano Titanium Dioxide

3.1 Preparation of the emission sample

The test vehicle is a PONTIAC TRANS SPORT, whose engine displacement is 3.8 L. The source of emission is the automotive exhaust gas that comes out when the Rotational Speed of the engine crankshaft is 900r/min and the temperature of the engine cooling liquid is preheated to 90□. The water and the vapor in automotive emission will affect test precision. Then a dehumidifier is concatenated with an automotive emission gatherer, which can get rid of water and vapor in automotive emissions during the process of emission leading into the degradation test chamber. The drier of the dehumidifier must be replaced before each test. The test temperature should be at 25□, and the emission charging interval is 10 seconds.

3.2 Experiment program

We take the content of 2%, 3%, 4%, 5%, made by dispersed mechanism methods mixing round with high shear, ultrasonic wave and ultrasonic wave dispersant, as test samples. According to the different dispersed mechanism methods and the different contents, test samples are divided into three groups. In the condition of the same dispersed mechanisms and the same content, there are five test samples to be tested. The test conditions are the same and the test time is 360 minutes.

The test sample is put into the degradation test chamber. The side brushed with Nano titanium dioxide is exposed to the Xenon lamp so that test samples can be exposed directly to the Xenon lamp. The automotive emission is sealed into the degradation test chamber without leakage. The flow fan and power switch are opened, which have the function of mixing automotive emissions uniformly and preheating the sensors of the computer assisted test system respectively. Meanwhile, the variation of every item of data is observed. The Xenon lamp is opened until all the data output is in a stabile state. Then, the real test begins. The corresponding variable values of NO_x, HC, CO, temperature, and pressure can be recorded by the computer assisted test system in real time. It is 20 minutes to gather the time interval in the data.

4 Experiment Results and Analysis

For the experiment on degradation efficiency of Nano titanium dioxide, take the average degradation efficiency of NO_x, HC, CO as evaluating indicators in the whole test time. According to the experiment data, the degradation efficiency of NO_x, HC, CO of different contents Nano titanium dioxide, made by three different dispersed methods above, is calculated. Then, the average degradation efficiency is considered as evaluating indicators of Nano titanium dioxide degradation on NO_x, HC, CO of automotive emission, in the condition of the same dispersed method and the same density of Nano titanium dioxide. Figures 2, 3 and 4 show variable trends of the contents of 2% to 5% Nano titanium dioxide, made by the dispersed mechanism method mixing round with high shear. Figures 5, 6 and 7 show variable trends of the contents of 2% to 5% Nano titanium dioxide, made by the dispersed mechanism method mixing round with an ultrasonic wave. Figures 8, 9 and 10 show variable trends of the contents of 2% to 5% Nano titanium dioxide, made by the dispersed mechanism method mixing round with an ultrasonic wave dispersant.

The degradation efficiency of Nano titanium on the factors of automotive emission is calculated by equation 1, when the test time is 360 minutes. The degradation efficiency equation of any kind of factor of automotive emission is

$$\eta_i = \frac{c_i - c_{i0}}{c_{i0}} \tag{1}$$

Where η_i is degradation efficiency of the No. i factor of automotive emission, c_{i0} is the initial content of the No. i factor of automotive emission, and c_i is the final content of the No. i factor of automotive emission.

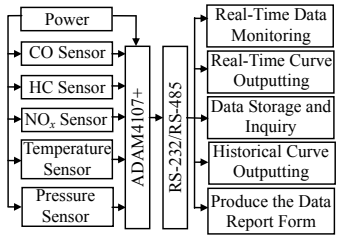


Figure 1. Flow chart of the test system

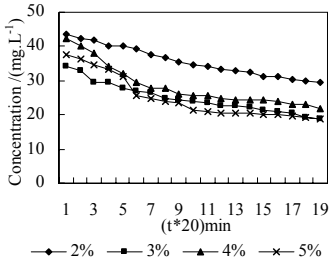


Figure 2. The decreasing curve on NO_x of Nano titanium dioxide prepared in high shear

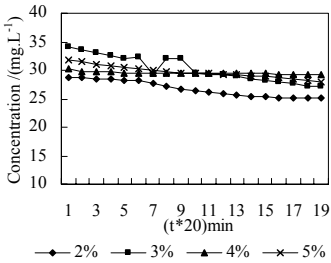


Figure 3. The decreasing curve on HC of Nano titanium dioxide prepared in high shear

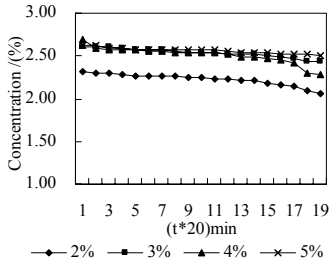


Figure 4. The decreasing curve on CO of Nano titanium dioxide prepared in high shear

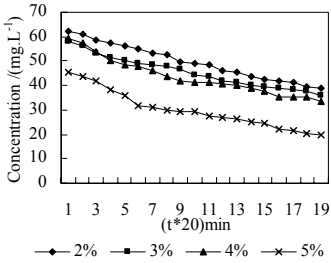


Figure 5. The decreasing curve on NO_x of Nano titanium dioxide prepared in ultrasonic

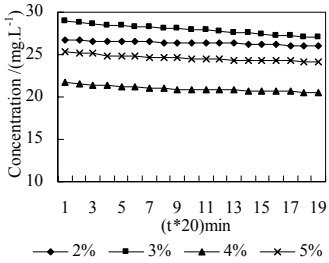


Figure 6. The decreasing curve on HC of Nano titanium dioxide prepared in ultrasonic

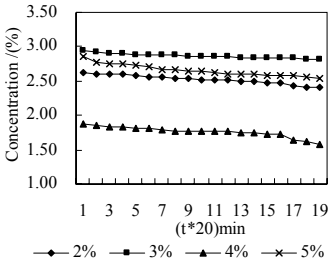


Figure 7. The decreasing curve on CO of Nano titanium dioxide prepared in ultrasonic

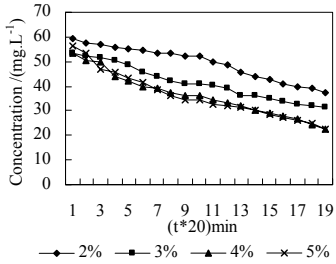


Figure 8. The decreasing curve on NO_x of Nano titanium dioxide prepared in ultrasonic dispersant

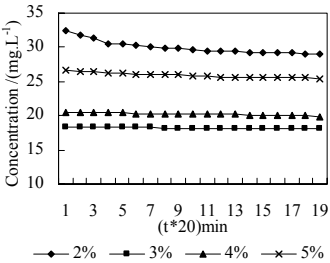


Figure 9. The decreasing curve on HC of Nano titanium dioxide prepared in ultrasonic dispersant

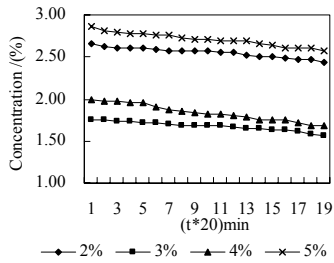


Figure 10. The decreasing curve on CO of Nano titanium dioxide prepared in ultrasonic dispersant

Table 2 shows the average degradation efficiency of the different contents and the different dispersed methods of Nano titanium, which is calculated by the result of degradation efficiency.

Table 2. The Average Degradation Efficiency of Different Content Nano titanium of Three Dispersed Methods on NO_x, HC, CO (%)

		2%TiO_2	3%TiO_2	4%TiO_2	5%TiO_2
High Shear dispersion	NO_x	29.48	36.53	42.58	50.21
	HC	11.90	17.88	3.35	11.57
	CO	24.57	6.11	19.44	4.48
Ultrasonic dispersion	NO_x	36.76	38.09	43.41	55.97
	HC	3.91	6.75	5.36	4.50
	CO	8.39	4.81	15.03	11.10
Ultrasonic Dispersant Mix	NO_x	37.37	41.71	57.16	60.16
	HC	10.57	3.95	4.05	4.20
	CO	8.67	10.27	7.89	9.95

It can be seen from the graphs and the average degradation rates listed in the table that the contents of NO_x, HC, CO vary in the effect of Nano titanium; the degradation efficiency of different contents and dispersed methods can be reflected by the degradation curves. For some factors of automotive emission, the degradation curves follow the same trend since different contents of Nano titanium and different dispersed methods were adopted. Nano titanium has good degradation efficiency on NO_x. The best degradation efficiency of Nano titanium is 5% content, which is made by dispersed mechanism methods mixing round with an ultrasonic wave dispersant. The maximum average degradation efficiency on NO_x is 60.16% during the time of 360 min, and the best method is the dispersed mechanism method mixing round with an ultrasonic wave dispersant in the condition of the same content of Nano titanium, which has the best degradation efficiency. The best degradation efficiency content of Nano titanium on NO_x, HC, CO, made by the method of high shear, ultrasonic wave and ultrasonic wave dispersant is 5%, 3%, 2%; 5%, 3%, 4%; 5%, 2%, 3% respectively.

5 Conclusions

The experiment system with degrading materials on automotive emissions is developed to test the degradation efficiency of automotive emissions factors with four different contents of Nano titanium made by dispersed mechanism methods, which mixed round with high shear, ultrasonic wave and ultrasonic wave dispersant. The results show that the degradation trends of different contents of Nano titanium, made by the three dispersed mechanism methods above, are the same. If the optimum dispersed method is adopted, the degradation efficiency of Nano titanium on the factors of automotive emission will be improved. The degradation deficiency of different content Nano titanium made by different dispersed methods on NO_x, HC, CO is calculated by the degradation equation. It can be seen that Nano titanium has good degradation efficiency on NO_x, and the best content is 5%, which is made by dispersed mechanism methods mixing round with ultrasonic wave dispersant. It will be a good reference for the purification of the atmosphere in the future.

References

BAI Hai-ying, WANG Rong-guo(2005). "Property characterization and experimentation of TiO_2 photocatalyst and degradation material [J]", *Low Temperature Architecture Technology*, 85-86.

HAN Xiang-chun, BAI Hai-ying, GUAN Qiang(2005). "Design on testing system for degrading automobile emission with photocatalytic material TiO_2 [J]", *Journal of Northeast Forestry University*, 89-91.

HU An-zheng, TANG Chao-qun(2001). "Progress in research of the nanosized TiO_2 photocatalytic materials and their application in environment protection [J]", *Journal of Functional Materials*, 586-594.

IBUSUKIT T K(1994). "Removal of low concentration nitrogen oxides through photoassisted heterogeneous catalysis [J]", *J.Catalysis*, 93-99.

LIU Shou-Xin, LIU Hong(2005). "The foundation and application of photocatalysis and photoelectric catalysis [M]", *Beijing: Chemical Industry Press*, 206-208.

TAKAMIK, SAGAWAT, UEHARA H, et al(1999). "Photocatalytic De-NO_x-ing building materials [J]". *Catalysts Catal*, 295-296.

ZHANG Wan-zhong(2006). "The preparation and application of titania-based photcatalytic nano materials [J]", *Chemical Reagents*, 523-526.

Discussion of Coverage for Airport Asphalt Pavement Design

Hongduo ZHAO[1] and Jianming LING[2]

[1]PH.D., Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, NO.1239 Siping Road, Shanghai China, 200092; PH (8621) 65982292; FAX (8621) 65985891; email: hdzhao@mail.tongji.edu.cn
[2]Professor, PH.D., Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, NO.1239 Siping Road, Shanghai China, 200092; PH (8621) 65981431; FAX (8621) 65985891; email: jmling01@yahoo.com.cn

Abstract: Coverage is an important parameter in the airport pavement design. Coverage calculating method in CBR pavement design method of FAA and UFC are detailed. In those methods the load repetitions can be considered using pass-to-coverage ratio. The ration is calculated base on the wheel-path distribution on pavement surface, which is not corresponding to the damage caused by the stress or strain within pavement depth. The calculating position is moved from surface to the top of subgrade in LEDFAA 1.3, while the lateral distribution of strain is considered as uniform not a normal distribution. There are no specific definitions of coverage in the specifications for asphalt concrete pavement design in China. For a pavement design method which can satisfy the New Generation Large Aircraft, such as A380, the coverage should be considered base on pavement spatial responses. The general pavement design concept advanced by Monismith gives a good idea to develop coverage calculation method.

Key words: *airport, pavement design, coverage, pass-to-coverage ratio*

1 Introduction

Aircraft affects pavement via landing gear. Coverage is the basic parameter of airport pavement design. It is original defined as the maximum number of tire prints or partial tire prints applied to the pavement surface at the point where maximum accumulation occurs. Coverages resulting from operations of a particular aircraft type are a function of the number of aircraft passes, the number and spacing of wheels on the aircraft main landing gear, the width of the tire contact area and the lateral distribution of the aircraft wheel-paths relative to the pavement centerline or guideline marking. There are several methods to calculate coverage that is corresponding with pavement design method. In this paper, current methods of coverage calculation are discussed, and then give some advice.

2 Coverage in CBR design method

FAA CBR method is the popular airport pavement design method in the world. The pass-to-coverage ratios are used in developing the flexible pavement design charts in Advisory Circular 150/5320-6D. Fundamental to this procedure for converting aircraft passes to coverages is the assumption that the aircraft wheel-paths conform a normal distribution. For a single wheel, maximum accumulation occurs at the midpoint of the normal distribution curve (see Figure. 1). However, for a tire contact width of W_t, accumulations will occur at the midpoint for only those wheel passes for which the wheel centerline is within $W_t/2$ distance from midpoint. So the pass-to-coverage ratios can calculate by equation (1).

$$\frac{P}{C} = \frac{1}{C_x W_t} \tag{1}$$

$$C_x = \varphi(x)\Big|_{x=\mu} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\Big|_{x=\mu} \tag{2}$$

Where, P is aircraft passes; C is coverages; C_x is the wheel centerline passes per inch (see Figure 1).

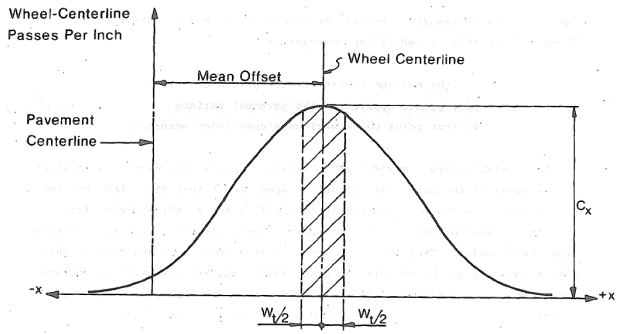


Figure 1. Theoretical normal distribution for single wheel (Hosang)

For an aircraft with many wheels, the maximum ordinate, C_{xc}, of the cumulative distribution of all the wheels must be determined. For example, the cumulative result of Boeing 727 main gear is shown in Figure 2 (Victor A. Hosang). The top curve is superposed with four wheels. Figure 2 shows that the maximum accumulation location moved from the midpoint of the normal distribution curve to midpoint of the wheel spacing. The pass-to-coverage ratios equations for 4 wheels are shown as (3) ~ (4). If there are more wheels on the aircraft main gear, it can also calculate. The pass-to-coverage ratios equations for IL-76, which has 16 wheels in the main gear, are

shown as (3) and (5).

$$\frac{P}{C} = \frac{1}{C_{xc} \mathrm{W}_t} \tag{3}$$

$$C_{xc} = f(x)\Big|_{x=\frac{s}{2}} = \frac{1}{\sigma\sqrt{2\pi}}\left[e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} + e^{-\frac{1}{2}\left(\frac{x-s}{\sigma}\right)^2}\right]\Bigg|_{x=\frac{s}{2}} \tag{4}$$

$$C_{xc} = f(x)\Big|_{x=s_1+\frac{s_2}{2}} = \frac{1}{\sigma\sqrt{2\pi}}\left[e^{-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2} + e^{-\frac{1}{2}\left(\frac{x-s_1}{\sigma}\right)^2} + e^{-\frac{1}{2}\left(\frac{x-s_1-s_2}{\sigma}\right)^2} + e^{-\frac{1}{2}\left(\frac{x-2s_1-s_2}{\sigma}\right)^2}\right]\Bigg|_{x=s_1+\frac{s_2}{2}} \tag{5}$$

Where, s is the spacing of dual wheels; S_1 is the spacing of out wheels for 4 wheels gear; S_2: is the spacing of inner wheels for 4 wheels gear.

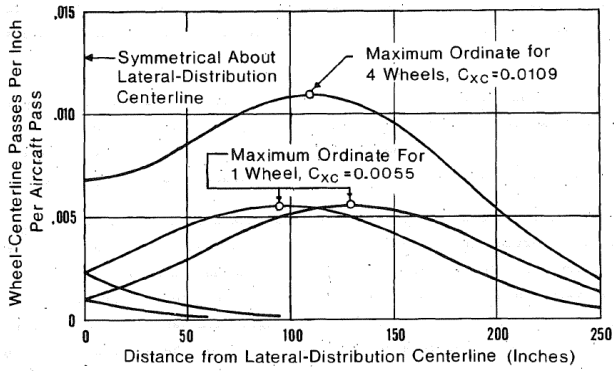


Figure 2. Theoretical normal distribution for Boeing 727 main gear (Hosang)

The pass-to-coverage ratios used in developing the flexible pavement design charts in AC 150/5320-6D are listed in Table 1.

Table 1. Pass-to-coverage ratios for flexible pavement (FAA)

Design Curve	Pass-to-coverage ratio	Design Curve	Pass-to-coverage ratio
Single Wheel	5.18	B-757	3.88
Dual Wheel	3.48	B-767	3.90
Dual Tandem	3.68	C-130	4.15
A-300 Model B2	3.51	DC10-10	3.64
A-300 Model B4	3.45	DC-30	3.38
B-747	3.70	L-101 I	3.62

The method used in CBR airport pavement design method of UFC is similar to

FAA. The calculate equation shown as (6) and (7).

$$C_x = n_0 \cdot \sum_{i=1}^{m} P_i \tag{6}$$

$$P_i = \int_{x_0-\frac{w_t}{2}}^{x_0+\frac{w_t}{2}} \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-x_i}{\sigma}\right)^2} dx \tag{7}$$

Where, P_i is the probability of i[th] wheel passes a point of pavement surface; n_0 is the pass number of a particular type aircraft; w_t is the width of wheel print; x_0 is the distance to the pavement centerline; x_i is the distance from centerline of wheel and aircraft; σ is the standard deviation of wheel lateral normal distribution.

The coverage calculation methods for CBR pavement design method of FAA and UFC are based on the statistic parameters of surface distribution of wheels. It is only suit the empirical pavement design method.

3 Coverage in LEDFAA design method

The pass-to-coverage ratios used in FAA Advisory Circular 150/5320-6D are based on the wheel-path distribution on pavement surface and wheel width. Those assumptions are not consistent with the damage control indexes of asphalt pavement, such as vertical strain on the top of subgrade, horizontal stress on the bottom of asphalt concrete layer. For those indexes, coverage should calculate in deep of pavement structure. In the layered elastic design procedure of FAA, LEDFAA 1.3, the coverages are calculated on the top of subgrade instead of on the surface in CBR methods. This time the definition of coverage has changed, but the term is still used. The area and peak number of stain transferred from the surface of gear load should consider in the coverage calculation. The coverages are the function of pavement thickness, stiffness, and gear type. Coverages are calculated on the top of subgrade is larger than on the pavement surface, it results more coverages when one aircraft passes.

In LEDFAA1.3, it assume that the wheel print is a ellipse which a:b equals 1.6, and the vertical transfer angle is with slop of 1:2. For dual wheel it is shown in Figure 3 and Figure 4 (Rodney Joel). In those figures, it also assumes that the strain in the area of equivalent tire is uniform and equals the maximum strain. For the thin flexible pavement, dual gear wheel will produce two equivalent tire prints with no overlap; otherwise for thick pavement, the equivalent tire prints reduced to one according to overlap. For main gear with multiple axles, the coverage can multiply by an axle index which is shown in Figure 5 (Rodney Joel). In the procedure of LEDFAA 1.3, coverages are calculated for each 10 inch wide strip along the pavement over a total width of 820 inches. Pass-to-coverage ratio is computed for each strip based on a normally distributed aircraft wander pattern with standard deviation of 30.5 inches

(equivalent to aircraft operation on a taxiway).

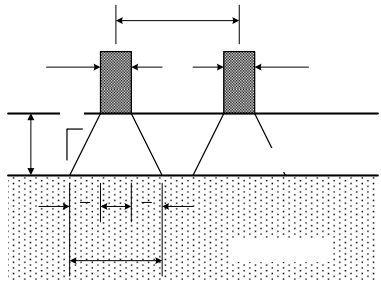


Figure 3. Equivalent tire width at the top of the subgrade with no overlap (two equivalent tires)

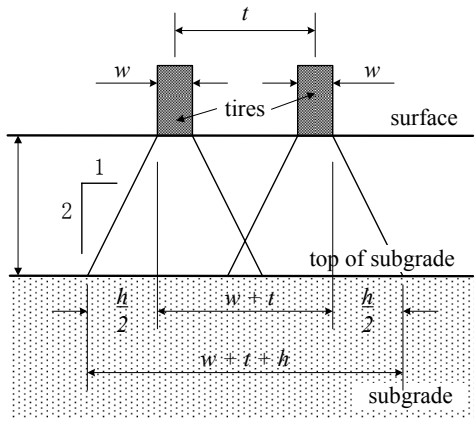


Figure 4. Equivalent tire width at the top of the subgrade with overlap (one equivalent tire)

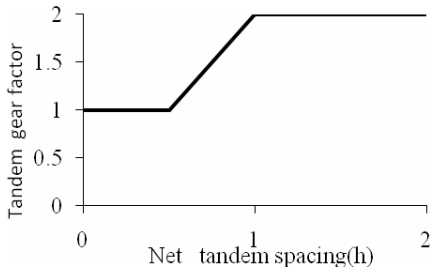


Figure 5. Tandem gear factor as a function of tandem spacing for two axles

4 Coverage in design method of China

In the specifications for asphalt concrete pavement design of civil airports, there are no details about pass-to-coverage ratios calculation. Instead of pass-to-coverage ratios, it uses a cumulate index of wheel-path lateral distribution (listed in Table 2). The coverage can obtain by aircraft annual operations multiply the cumulate index.

Table 2. Wheel-path cumulate index of lateral distribution (CCAA)

Pavement Width(m)	Δ
18	0.05
23	0.04
30	0.03
45	Dual 0.02, Dual Tandem 0.03
60	Single 0.01, Dual Tandem 0.03

In the military specification, it assumes that the wheel-path is uniform in the trip width. According to that assumption, coverages can calculate by equation (8).

$$N_c = \frac{N_j \eta N W_t}{100T} \tag{8}$$

Where, N_c is annual coverages of design aircraft; N_j is annual operations of design aircraft; η is the probability of design aircraft pass within pass width, for fighter plane is 0.90; for bomber and transporter is 0.85; N is the wheel number of one main gear; W is tire width (cm); T is pass width (m), for fighter plane is 3.8m on runway, 1.2m on taxiway; for or bomber and transporter is 5.0m on runway when the main gear spacing is larger than 7m and 4.4m on the contrary; 1.5m on taxiway.

The method of coverage calculation in cement concrete pavement design of civil airports is similar to the method in the army specification. The equation is shown as (9).

$$N_e = \frac{0.75 n_w W_t}{100T} N_s t \tag{9}$$

Where, N_e is Cumulate operations of design aircraft in design year life; N_s is annual operations of design aircraft; n_w the Wheel number of one main gear; W_t is tire width (cm); T is Pass width (m), on runway is 11.4m, on taxiway and apron is 2.3m; t is design year life.

According to function (8) and (9), the pass-to-coverage ratio calculation equation can be derived as (10):

$$\frac{P}{C} = \frac{T}{\eta n_w W_t} \tag{10}$$

5 Load repetition based on spatial responses

In the above depiction, there are two assumptions about wheel-path distribution: uniform distribution and normal distribution. Otherwise, the maximum stress or strain is used to calculate pass-to coverage ratio, and assume that these value are uniform in the effective depth. In the LEDFAA method the strain on the top of subgrade is the only strain used to consider traffic analyze. Actually, responses of pavement structure are spatial, the lateral distribution are not uniform in one depth of pavement. In order to consider the spatial responses, Monismith put forward a general pavement design concept.

In the general design concept advanced by Monismith, it assumes that wheel-paths conform normal distribution, for any point in the pavement, the responses at this point is different when aircraft passes at different lateral position. So, the actual and permit load repetition under different strain and stress level should calculate which based on pavement spatial responses. For example, if the fatigue of asphalt pavement is critical under dual wheel gear, the critical strain curve under gear load should be observed first; then for a given standard deviation of the wheel-path normal distribution, the actual load repetitions n_i in unit width can be determined; according to the strain level at that point and performance prediction function, the permit load repetition can also be determined; the total cumulate damage of pavement is the area under the n_i/N_i to aircraft position relationship curve.

For an aircraft with many axles, the total damage of one axle can be calculated uses the critical response section, then multiply the axle number to obtain the total damage of whole aircraft.

6 Conclusion

Several coverage calculation methods are discussed in this paper. For different pavement design methods, such as CBR method, LEDFAA method, general concept, the functions of coverage are varies.

(1) For the CBR method, the coverage is calculated on the surface of pavement. It is only suit the empirical pavement design method.

(2) Mechanistic-Empirical methods are developed to replace empirical method in order to satisfy the New Generation Large Aircraft. The LEDFAA is an M-E method developed by FAA. In this method, the coverage calculation position is moved from surface to the top of subgrade, while the only index considered is vertical strain and the lateral distribution of strain is assumed as uniform not a normal distribution.

(3) There are no specific definitions of coverage in the specifications for asphalt concrete pavement design in China.

(4) The coverage calculation methods should consider different pavement critical index, different pavement stiffness, and pavement spatial responses. The general pavement design concept advanced by Monismith gives a good idea to develop coverage calculation method.

References

C.L. Monismith, F.N. Finn, G. Ahlborn, N. Markevich. "A general analytically based approach to the design of asphalt concrete pavements." *Proceedings, 6th international conference on the structural design of asphalt pavements*, University of Michigan and Delft University, 1987.

China Civil Aviation Administration. *Specification for cement concrete pavement design for civil transport airports (MHJ 5004-95)*. Beijing, 1995.7.

China Civil Aviation Administration. *Specification for asphalt concrete pavement design of civil airports (MH 5010-1999)*. Beijing, 2000.5.

Gordon F. Hayhoe, Izydor Kawa, David R Brill. *New developments in FAA airport pavement thickness design software*. FAA William J. Hughes technical center airport technology research and development branch.

PLA General Logistics Department. *Specification of design of cement concrete pavement for military airports (GJB 1278-91)*. Beijing, 1992.3.

Rodney Joel, P.E. *Background information for engineering Brief #65, minimum requirements to widen existing 150-Foot-Wide runways for Airbus A380 operations.* Civil engineering/airfield pavements airport engineering division, Federal Aviation Administration, office of airport safety and standards.

U.S. Department of Transportation, Federal Aviation Administration. *Airport pavement design and evaluation, AC No: 150/5320-6D*. U.S. Department of Transportation, Federal Aviation Administration, 1995.7.

Unified Facilities Criteria (UFC). *Pavement design for airfields (UFC 3-260-02)*. U.S. Army corps of engineers (Preparing Activity), naval facilities engineering command, air force civil engineer support agency, 2001.6.

Victor A. Hosang. *Field survey and analysis of aircraft distribution on airport pavements, report No. FAA-RD-74-36*. FAA, 1975.2.

DYNAMIC MODULI FOR M-E DESIGN OF ASPHALT PAVEMENTS

Zhanping You[1], Sanjeev Adhikari[2], Shu Wei Goh[3], and Qingli Dai[4]

[1]Tomasini Assistant Professor of Transportation Engineering and Associate Director of the Transportation Materials Research Center, Department of Civil and Environmental Engineering, Michigan Technological University, Houghton, Michigan 49931, USA, zyou@mtu.edu
[2]Ph.D. Student and Research Assistant, Department of Civil and Environmental Engineering, Michigan Technological University, Houghton, Michigan 49931, USA. sadhikar@mtu.edu
[3]Ph.D. Student and Research Assistant, Department of Civil and Environmental Engineering, Michigan Technological University, Houghton, Michigan 49931, USA. sgoh@mtu.edu
[4]Research Assistant Professor, Department of Mechanical Engineering-Engineering Mechanics, Michigan Technological University , 1400 Townsend Drive ,Houghton, Michigan, 49931-1295, USA, qingdai@mtu.edu

Abstract: In the USA, many highway agencies have successfully implemented the Superpave volumetric mixture design procedure. Yet, a number of studies have shown that the Superpave volumetric mixture design method alone is insufficient to ensure reliable mixture performance over a wide range of traffic and climatic conditions. As highway agencies begin the implementation of the AASHTO Mechanistic-Empirical Design Guide (MEPDG), they will be tasked with the development of local calibration factors specific to the materials, construction practices, loading regimes and environment of their region. The local calibration factors specific to the materials include the test of the dynamic modulus of asphalt mixtures. In fact, the dynamic modulus of asphalt mixtures is an important input parameter for pavement thickness design. Therefore, the objectives of this paper are to: 1) review the current status of Superpave simple performance tests for asphalt mixes; 2) measure the dynamic modulus of Superpave asphalt mixes, and; 3) compare with the predictive equations for the dynamic modulus to establish a correlation. The dynamic modulus data will be used in the MEPDG design procedure for asphalt pavements.

Keywords: *Dynamic Modulus, Mechanistic -Empirical Design, Asphalt Pavements, Asphalt Mixes*

1. Objectives

The objectives of this paper are to: 1) review the current status of Superpave simple performance tests for asphalt mixes; 2) measure the dynamic modulus of Superpave asphalt mixes, and; 3) compare with the predictive equations for the dynamic modulus to establish a correlation. The dynamic modulus data will be used in the MEPDG design procedure for asphalt pavements.

2. Introduction

Mechanistic–Empirical (M-E) design is a combination of mechanistic and empirical approaches. The M-E design is directly involved with climate, materials, traffic, and pavement distress. The user enters trial design with repeated traffic, climate and materials input. The user also inputs the design life and acceptable performance in terms of key distress and smoothness. The MEPDG Software shows performance versus time in graphical and tabular formats. The user either retains or modifies the design. The M-E design allows a wide range of pavement structures, like new construction, reconstruction, rehabilitation and overlays. The asphalt pavement M-E design process allows traffic volume adjustment factors such as the monthly and hourly distribution, the vehicle class distribution and the axle load distribution. The climate information is processed with Enhanced Integrated Climatic Model (EICM). It uses specific and detailed inputs of asphalt pavement like modulus and thermal conductivity, heat capacity, absorption and drainage. It has multiple performance indicators like rutting, fatigue cracking, thermal cracking, and smoothness (IRI).

Many states' Department of Transportation (DOT) have successfully implemented the Superpave volumetric mixture design procedure. Yet, a number of studies have shown that the Superpave volumetric mixture design method alone is insufficient to ensure reliable mixture performance over a wide range of traffic and climatic conditions. Some research projects have been conducted in Michigan to evaluate the performance of mixtures designed using the volumetric design procedure. However, there has been a lack of a simple performance test (SPT) criteria to evaluate pavement rutting, fatigue cracking, and low temperature cracking of flexible pavements.

The development of a SPT has been the focus of considerable research efforts in the past several years. In fact, some aspects of the tests have been available for decades, such as the dynamic modulus test of asphalt mixes. The term "dynamic modulus" was around even earlier to describe concrete behavior as described by Valore and Yates (1942), Preece (1948), and Linger (1963). The dynamic modulus test was introduced into the asphalt pavement area decades ago (Papazian 1962).

A few recent research projects on the SPT are introduced here as part of the background information. Researchers (Shenoy and Romero 2002) focused on the use of dynamic modulus \E*\ data to predict asphalt pavement distresses. Researchers (Pellinen and Witczak 2002) reported the possibility of using the stiffness of asphalt mixes as the basis for the SPT. Asphalt Pavement Analyzer (APA) was used to assess rutting performance of mixtures. The resilient modulus, tensile strength, and simple shear tests were used to evaluated the moisture sensitivity and performance of lime-modified asphalt mixes (Martin and Park 2003). In Texas, researchers (Zhou and Scullion 2003) preliminarily validated the SPT for permanent deformation in a field case study, finding that both the dynamic modulus test ($E^*/\sin\delta$) and the repeated-load test (Fn) can distinguish good performing and poor performing mixtures. Researchers (Sotil et al. 2004) investigated the reduced confined dynamic modulus testing protocol for asphalt mixtures. Researchers (Tandon et al. 2004) investigated the results of integrating a SPT with an environmental conditioning system. Researchers (Galal et al. 2004) also investigated in-service accelerated pavement testing in order to model permanent deformation. Yet, even with all this research, a SPT specification

that considers specific trafficking levels for engineering application is not available at this time.

As this incomplete summary of past research indicates, a considered number of potential SPTs have been investigated to measure and assess fundamental engineering material properties that can link the advanced material characterization to the development of criteria for asphalt mixes design. A number of tests evaluated for the SPT include the dynamic modulus test, shear modulus test, triaxial repeated test, triaxial and uniaxial creep test, triaxial compressive strength test, asphalt pavement analyzer, gyratory shear stress test, indirect tensile strength and fatigue test, and the direct tensile strength test (NCHRP Project 9-19). It was recommended that the following properties are important material parameters: (1) the dynamic modulus term, $E^*/\sin\varphi$, which is determined from the triaxial dynamic modulus test; (2) the flow number, *fn*, which is determined from the triaxial repeated load test, and; (3) the dynamic modulus, E*, measured at low test temperatures. These tests were designed to control pavement rutting, fatigue, and thermal cracking. Currently, a large number of these tests are going on in Michigan and many other states.

3. Dynamic modulus

Currently, there are two general approaches in determining the mixture dynamic modulus E*: one is experimental tests and the other is prediction. In the prediction approach, there are two major methods, one is the discrete and finite element methods (Dai et al. 2006a; Dai and You 2006; Dai et al. 2006b; You and Buttlar 2004; You and Buttlar 2005; You and Buttlar 2006); the other method is using the empirical equations (Bari and Witczak 2006; Witczak et al. 2001) or micromechanical predictive equations (Christensen et al. 2003). In this paper, the experimental measurements of E* and predictive E* are discussed and compared.

The dynamic modulus test was conducted according to AASHTO TP62-03. A sinusoidal (haversine) axial compressive stress is applied to a specimen of asphalt concrete at a given temperature and loading frequency. Figure 1 shows the test set up, where the sample of an asphalt mix specimen is loaded under the compressive test. The applied stress and the resulting recoverable axial strain response of the specimen is measured and used to calculate the dynamic modulus and phase angle. The dynamic modulus is defined as the ratio of the amplitude stress (σ) and amplitude of the sinusoidal strain (ε) that results in a steady state response at same time and frequency as shown in Figure 2:

$$E^* = \frac{\sigma}{\varepsilon} = \frac{\sigma_0 e^{i\omega t}}{\varepsilon_0^{\,i(\omega t-\phi)}} = \frac{\sigma_0 \sin(\omega t)}{\varepsilon_0 \sin(\omega t-\phi)} \qquad (1)$$

Where,

σ_0 = peak (maximum) stress
ε_0 = peak (maximum) strain
ϕ = phase angle, degrees
ω = angular velocity
t = time, seconds

Scientifically, the complex modulus is a composite number including the viscus part and elastic part. Most of the researchers define the dynamic modulus as the absolute value of the complex modulus, or:

$$|E^*| = \frac{\sigma_0}{\varepsilon_0} \qquad (2)$$

The phase angle appears to be 0° for a pure elastic material and 90° for a pure viscous material. Figure 3 shows the stress and strain versus testing time for a typical test at 25 Hz frequency and 13°C. In this test, three strain readings are obtained for the three strain gauges under the cyclic load. The dynamic modulus and phase angles were computed for all the test specimens at the given temperatures (-5, 4, 13, 21.3, and 39.2°C) and loading frequencies (0.1. 0.5, 1, 5, 10, 25 Hz).

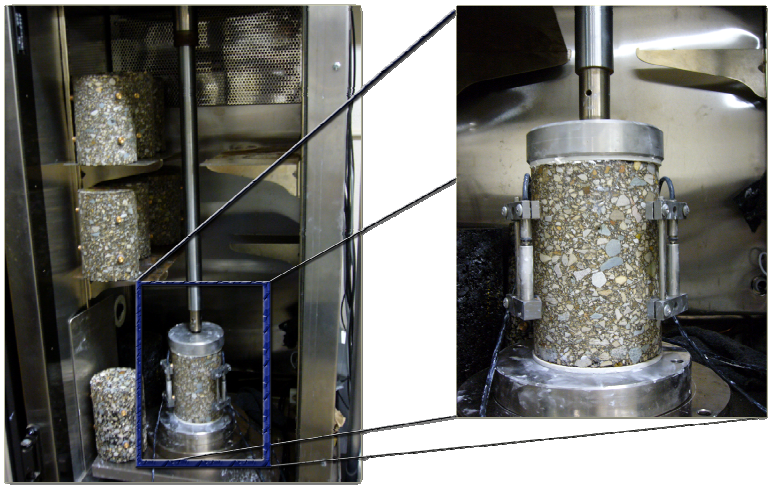


Figure 1. E* test setup

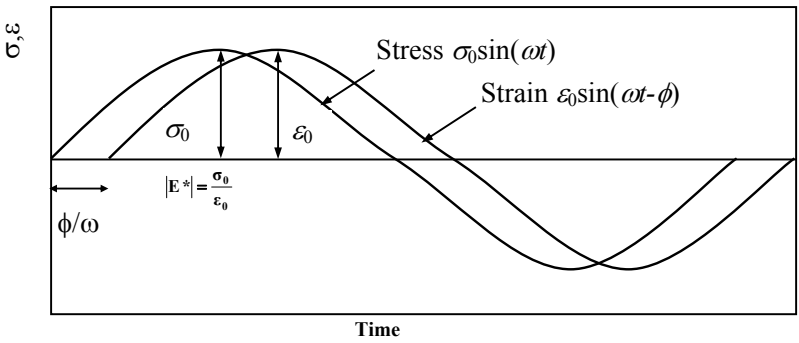


Figure 2. E* is a function of temperature, rate of loading, age, and volumetric properties

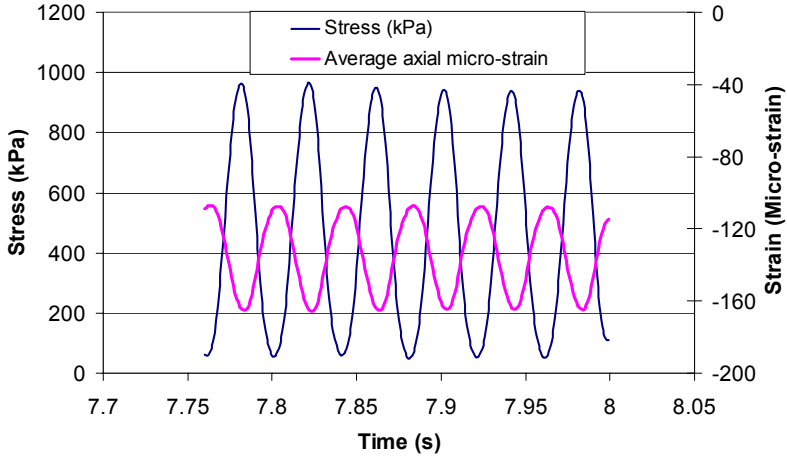


Figure 3. Stress (1 curve) and strain (average of 3 curves) in dynamic loading using 25 Hz frequency at temperature 13°C

4. Compare experimental measurements and predictive dynamic modulus

Dynamic modulus of asphalt mixes is dependent upon the properties of the individual components and volumetric composition of the mixes. Dynamic modulus of the asphalt mixes is predictable if the properties of components are known. An asphalt mixture shows visco-elastic phenomena due to viscous characteristics of the binder. There are some empirical relationships available for predicting dynamic modulus for mixes. The dynamic moduli predicted by the Witczak and Hirsch models were compared against the measured data to determine the quality of each model. The Witczak and Hirsch models were developed using all test temperatures and frequencies. The measured dynamic moduli are at temperature ranges from -5°C to 39.2°C and frequencies ranges from 25 Hz to 0.1 Hz. At each temperature and frequency, testing was conducted at three air void levels, 4%, 7%, and 10%. RTFO aged binder was used to measure the shear dynamic modulus (G_b*). In the following sections, the measured dynamic modulus versus the predicted dynamic modulus at all temperature, frequencies, and air voids were compared.

Witczak's Predictive Equation

In the current MEPDG (NCHRP I-37A), the stiffness of the asphalt mixes is determined from a sigmoidal E* master curve using one of three alternate input levels. The master curve for input Level 1 design is developed using numerical optimization to shift the laboratory mixture E* test data into a master curve (Bari and Witczak 2006; Witczak 2005a; Witczak 2005b). Before shifting the E* data, the relationship between binder viscosity and temperature is established by:

$$\eta = \frac{\left|G_b^*\right|}{10}\left(\frac{1}{\sin\delta_b}\right)^{4.8628} \quad (3)$$

$$\log\log\eta = A + VTS\log T_R \quad (4)$$

Where,

η	=	binder viscosity, cP
$\lvert G_b^* \rvert$	=	binder complex shear modulus, Pa
δ_b	=	binder phase angle, degree
A, VTS	=	regression parameter
T_R	=	temperature, ^{o}Rankine

The master curve for the Level 2 input is directly determined from the "Witczak E* Predictive Equation", using specific laboratory binder test and mixture data. The Level 3 input is developed using this predictive equation and certain (typical) properties of the binder and mix. The current version of the Witczak E* Predictive Equation is as follows:

$$\log_{10} E^* = -0.349 + 0.754\left(\left|G_b^*\right|^{-0.0052}\right)\times\left(\begin{array}{l} 6.65 - 0.032\rho_{200} + 0.0027\rho_{200}^{\ 2} - 0.011\rho_4 - 0.0001\rho_4^{\ 2} \\ + 0.006\rho_{38} - 0.00014\rho_{38}^{\ 2} - 0.08V_a - 1.06\left(\frac{V_{beff}}{V_a + V_{beff}}\right) \end{array}\right)$$

$$+\frac{2.56 + 0.03V_a + 0.71\left(\frac{V_{beff}}{V_a + V_{beff}}\right) + 0.012\rho_{38} - 0.0001\rho_{38}^{\ 2} - 0.01\rho_{34}}{1 + e^{\left(-0.7814 - 0.5785\log\left|G_b^*\right| + 0.8834\log\delta_b\right)}} \quad (5)$$

Where,

E*	=	dynamic modulus, psi
ρ_{200}	=	percent of aggregate (by weight of the total aggregate) passing through no. 200 sieve, %
ρ_4	=	percent of aggregate (by weight of the total aggregate) retained on no. 4 sieve, %
ρ_{38}	=	percent of aggregate (by weight of the total aggregate) retained on the 3/8 inch sieve, %
ρ_{34}	=	percent of aggregate (by weight of the total aggregate) retained on the 3/4 inch sieve, %
V_a	=	air voids (by volume of the mix), %
V_{beff}	=	effective binder content (by volume of the mix), %
$\lvert G_b^* \rvert$	=	dynamic shear modulus of binder, psi
δ_b	=	phase angle of binder associated with $\lvert G_b^* \rvert$, degree

The following presents a comparison between the measured dynamic modulus versus predicted dynamic modulus using Witczak's equation (Bari and Witczak 2006). Figure 4 shows the comparison between predicted dynamic moduli and measured

dynamic moduli. The relation of the predicted and measured dynamic modulus is as follow.

$$|E^*|_{measured} = 0.7959|E^*|_{Predicted} \quad (6)$$

Hence, for this type of mixes, it is possible to correct for the predicted dynamic modulus by multiplying the calibration factor by 0.7959. It is interesting to note that the modulus predictions at low temperatures (i.e. high modulus value) are closer to the measured values. However, for this type of mixture, the difference between the predicted and measurements is significant for the given temperatures and frequencies.

Hirsch's Model

The Hirsch Model was developed by Christensen et al. (2003) to estimate the dynamic modulus of asphalt concrete using the binder dynamic modulus and volumetric properties of the mixture (VMA and VFA). The equation was expressed as following equation.

$$|E^*|_{mix} = P_C\left[4200000\left(1-\frac{VMA}{100}\right)+3|G^*|_b\left(\frac{VFA*VMA}{1000}\right)\right]+(1-P_C)\left[\frac{1-\frac{VMA}{100}}{4200000}+\frac{VMA}{3VFA*|G^*|_b}\right]^{-1} \quad (7)$$

Where,

$\|E^*\|_{mix}$	=	Complex modulus for mixture, lb/in^2
$\|G^*\|_b$	=	Complex modulus for binder, lb/in^2
P_C	=	The contact factor, which was expressed as following equation

$$P_C = \frac{\left(20+\frac{VFA*3|G^*|_b}{VMA}\right)^{0.58}}{650+\left(\frac{VFA*3|G^*|_b}{VMA}\right)^{0.58}} = \text{aggregate contact volume} \quad (8)$$

VFA	=	Voids filled with asphalt, %
VMA	=	Voids in mineral aggregate, %
$\|G^*\|_b$	=	Dynamic shear modulus of binder, lb/in^2

Figure 5 shows the comparison between predicted dynamic moduli and measured dynamic moduli using Hirsch's model. The relation of the predicted and measured dynamic modulus is as follow.

$$|E^*|_{measured} = 1.05|E^*|_{Predicted}$$

Therefore, a small calibration factor 1.05 would be good to obtain the predicted dynamic modulus. For this case, all the modulus predictions are close to the measured values.

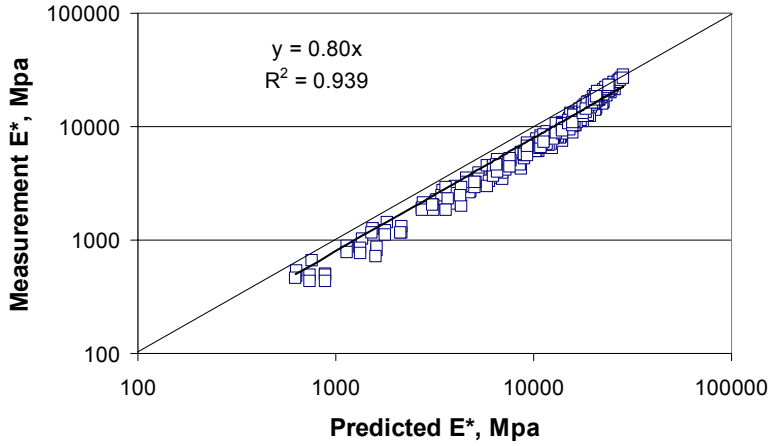


Figure 4. Measured value versus predicted value of |E*| using Witczak's model

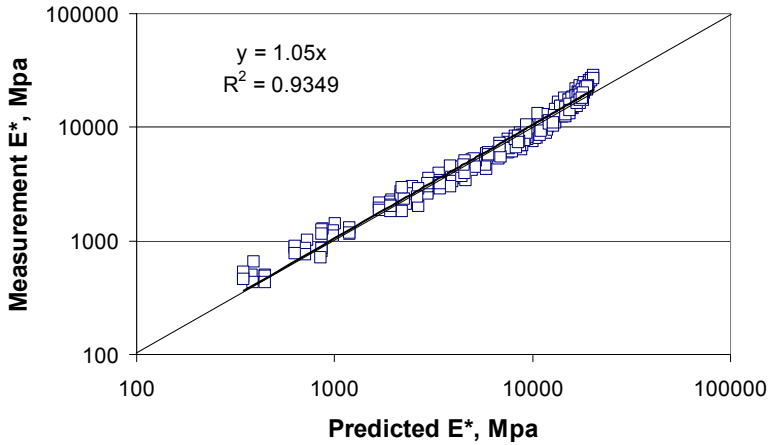


Figure 5. Measured |E*| versus predicted value using the Hirsch's model

5. Summary and conclusions

In this paper, the current Superpave simple performance tests including the dynamic modulus for asphalt mixture is briefly reviewed and some experimental tests on the dynamic modulus is discussed. In addition, the predictive dynamic modulus

from both the newly revised Witczak equation in 2006 and the Hirsh model developed in 2003 are used to compare the measurements. The reason for using the predictive equations (or models) is that the simple method may save time and cost in the experimental measurements for the MEPDG design procedure. It was found that the current Witczak equation did not accurately predict the dynamic modulus. The predictions from the Hirsh model are very close to the measurements. It should be noted that the comparison was completed on limited data obtained in the State of Michigan and should be further validated. In addition, experimental measurements are always recommended. The predictions can be used only when the measurements are not available.

6. Acknowledgments and Disclaimer

The authors acknowledge visiting scholars Dechao Li and Xiaguang Hu's assistants in experimental tests. The views expressed in this paper and the accuracy of the data and facts contained herein are the sole responsibility of the authors, and do not necessarily represent the official views of the listed sponsoring agencies. This paper does not constitute a standard, specification, or regulation. Comments contained in this paper related to specific testing equipment should not be considered an endorsement of any commercial product or service; no such endorsement is intended or implied.

References

Bari, J., and Witczak, M. W. (2006). "Development of a New Revised Version of Witzacak E* Presictive Mocxel for Hot Mix Asphalt Mixtures. " *Journal of the Association of Asphalt Paving Technologists* 75, 381-423.

Christensen, D. W., Pellien, T., and Bonaquist, R. F. (2003). "Hirsch Models for Estimating the Modulus of Asphalt Concrete." *J. Assoc. Asphalt Paving Tech.*, 72, 97-121.

Dai, Q., Sadd, M. H., and You, Z. (2006a). "A Micromechanical Finite Element Model for Viscoelastic Creep and Viscoelastic Damage Behavior of Asphalt Mixture." *International Journal for Numerical and Analytical Methods in Geomechanics*, 30, 1135–1158.

Dai, Q., and You, Z. (2006). "Prediction of Creep Stiffness of Asphalt Mixture with Micromechanical Finite Element and Discrete Element Models." *Journal of Engineering Mechanics, ASCE.*, 133(2), 163-173.

Dai, Q., You, Z., and Sadd, M. H. (2006b). "A Micromechanical Viscoelaso-Plastic Model For Asphalt Mixtures." *Geotechnical Special Publication (GSP): Asphalt Concrete: Simulation, Modeling, and Experimental Characterization. American Society of Civil Engineers (ASCE)* 12-20.

Galal, K. A., Noureldin, A. S., White, T. D., and Haddock, J. E. (2004). "Laboratory, Prototype, And In-Service Accelerated Pavement Testing To Model Permanent Deformation." *Transportation Research Record, National Research Council*, 1896.

Linger, D. A. (1963). "Effect of Stress on the Dynamic Modulus of Concrete." *Highway Research Record, Hwy Res Board*, 3, 62-73.

Martin, A. E., and Park, D. (2003). "Use of the Asphalt Pavement Analyzer and Repeated Simple Shear Test at Constant Height to Augment Superpave Volumetric Mix Design." *Journal of Traffic and Transportation Engineering*, 129(5), 522-530.

Papazian, H. S. "The Response of Linear Visoelastic Material in the Frequency Domain with Emphasisi on Asphaltic Concrete." *International Conference on Structural Design of Asphalt Pavements*, Rackham Lecture Hall, University of Michigan, Ann Arbor, 454-563.

Pellinen, T. K., and Witczak, M. W. (2002). "Use of stiffness of hot-mix asphalt as a simple performance test." *Transportation Research Board*, p. 80-90.

Preece, E. F. "Determination And Use Of The Dynamic Modulus Of Elasticity Of Concrete." *Highway Research Board Proceedings*, 233-237.

Shenoy, A., and Romero, P. (2002). "Standardized Procedure For Analysis Of Dynamic Modulus \E*\ Data To Predict Asphalt Pavement Distresses." *Transportation Research Record, National Research Council*, 1891.

Sotil, A., Kaloush, K. E., and Witczak, M. W. (2004). "Reduced Confined Dynamic Modulus Testing Protocol For Asphalt Mixtures." *Transportation Research Record, National Research Council*, 1891.

Tandon, V., Kambham, B. S., Bonaquist, R., and Solaimanian, M. (2004). "Results of integrating simple performance tests and environmental conditioning system." *Transportation Research Board*, 140-152.

Valore, R. C., and Yates, J. C. "Effects Of Calcium Chloride On Dynamic Modulus Of Elasticity Of Concrete Test Specimens." *Highway Research Board Proceedings*, 293-305.

Witczak, M. (2005a). *Appendixes to NCHRP Report 547: Simple Performance Tests and Advanced Materials Characterization Models*, Transportation Research Board- National Cooperative Highway Research Program.

Witczak, M. (2005b). *Simple Performance Tests: Summary of Recommended Methods and Database*, Transportation Research Board.

Witczak, M. W., Bari, M., and Quayum, M. M. (2001). "Sensitivity of Simple Performance Test Dynamic Modulus |E*|." Arizona State University, Department of Civil and Environmental Engineering, Tempe, AZ.

You, Z., and Buttlar, W. G. (2004). "Discrete Element Modeling to Predict the Modulus of Asphalt Concrete Mixtures." *Journal of Materials in Civil Engineering, ASCE*, 16(2), 140-146.

You, Z., and Buttlar, W. G. (2005). "Application of Discrete Element Modeling Techniques to Predict the Complex Modulus of Asphalt–Aggregate Hollow Cylinders Subjected to Internal Pressure." *Journal of the Transportation Research Board, National Research Council*, 1929, 218-226.

You, Z., and Buttlar, W. G. (2006). "Micromechanical Modeling Approach to Predict Compressive Dynamic Moduli of Asphalt Mixture Using the Distinct Element Method." *Transportation Research Record: Journal of the Transportation Research Board, National Research Council, Washington, D.C.*, 1970, 73-83.

Zhou, F., and Scullion, T. (2003). "Preliminary Field Validation Of Simple Performance Tests For Permanent Deformation: Case Study." *Transportation Research Records* 1832, 209-216.

Laboratory Study of Fatigue Limit for LSM Base Mixture

Zhixiang Zhang[1] and Qi-feng Bai[2]

1.Zhi-Xiang ZHANG Jiangsu Transportation Research Institute, Nanjing, China, 210017, zzx@ti.js.cn
2.Qi-Feng BAI Jiangsu Transportation Research Institute, Nanjing, China, 210017, bqf@ti.js.cn

Abstract: Recently the research on Perpetual Pavement has been very popular, and the basis of the structure design method for the Perpetual Pavement is the theory of the HMA "Fatigue Limit". In this article, the laboratory study for the fatigue characteristics of LSM Base Mixture chooses the Four Point Beam Bending Test method with four kinds of strain level and two kinds of asphalt content, and the two kinds of data analysis methods which include the AASHTO T321 method and the reduced energy ratio method are used. Based on the analysis of the fatigue tests, the fatigue limit for the LSM Base Mixture on the lower strain level (70～100με) is verified.

Key Words: Fatigue Limit LSM base mixture Laboratory study

1 INTRODUCTION

In the perpetual pavement design，selection of a mixture type depends on the function of each layer. Generally, the surface layer needs rutting resistance, crack resistance, and water resistance; the middle layer needs rutting resistance, stability, and water resistance; and the base layer needs fatigue resistance. Perpetual pavement means to have enough pavement thickness so pavement distress is only retained on the surface, so the perpetual pavement can be used for a long enough time and rehabilitated periodically. The designed life of perpetual pavement is defined as 40 to 50 years, and the pavement period rehabilitation time is about 10 to 15 years. The structure design method of Perpetual Pavement is totally different from the traditional asphalt pavement. The design traffic load input is axle load and not cumulative volume of traffic as is the design traffic traditional pavement [3]. The basis of the Perpetual Pavement design method is the "Fatigue Limit" theory of asphalt mixture. According to this theory, asphalt mixture has a bending critical point. The pavement will never be damaged when the strain of the mixture in the pavement is lower than this value.

Scholars and institutes in America and Japan have studied the fatigue limit of asphalt mixtures in the laboratory and have concluded that the asphalt mixture depends on the fatigue limit, and that the fatigue limit is related to the kind of asphalt mixtures, raw material properties and the test conditions. In this paper, the "Fatigue Limit" of LSM was investigated and verified through beam fatigue tests at low strain levels.

2 MIXTURES

The asphalt binder used in the study meets Chinese asphalt specifications as AH70. The aggregates used in the mixtures are limestone. The Standard Marshall mix design procedure was used to determine the gradation of aggregates and the asphalt content of theLSM mixture. Table 1 presents the job mix formula for LSM.

Table 1. Mixture Job Mix Formula for LSM

Gradation	Asphalt content (%)	Total % Passing Through Following Meshes (mm)											
		31.5	26.0	19	13.2	9.5	4.75	2.36	1.18	0.6	0.3	0.15	0.075
LSM-25	3.6	100	91.6	71.5	52.8	42.9	30.9	24.0	17.7	11.6	8.4	6.5	5.6

3 LSM MIXTURE BEAM FATIGUE TESTS

3.1Test equipment

LSM mixture's fatigue characteristics were evaluated through beam fatigue tests (AASHTO TP8-94). The test equipment used in the study was asphalt mixture test system NU-14 which was produced by COOPER of Britain. Figure1 presents the loading mode of the experiments.

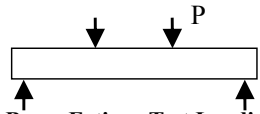
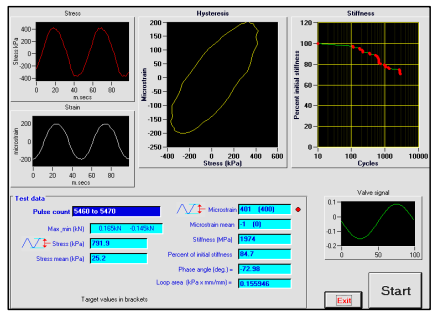


Figure1 Beam Fatigue Test Loading Mode

Loading and data gathering in fatigue tests were controlled automatically by computer. NU-14 and its controlling interface are presented in Figure 2.



（a） Fatigue Test equipment NU-14 （b）Fatigue Test Controlling Procedure Interface

Figure2 Four Point Bending Test Equipment And Controlling Interface

3.2 Test procedure

（1）First, press mixture into a board with static pressure mode, then cut the board into small beams with 380mm length, 63mm width and 50mm height.

（2）Set the temperature of NU-14 at 20.0℃ and warm up for 15 minutes, then put beam into the machine for about 2 hours.

（3）Put the test beam into the loading system, then lock the grippers and make LVDT touch the beam properly.

（4）Set the controlling strain level, loading frequency, beam's size in three dimensions, etc., then begin the test.

3.3 Test mode

The beam fatigue test has two loading controlling methods. One is the strain controlling method, and the other is the stress controlling method. The strain controlling method was chosen in this study in order to match the structure design index of perpetual pavements. In this way, we can evaluate the fatigue performance of base layer material available and integrate it with the pavement structure design.

A suitable controlling stain level is very important to evaluate the mixture fatigue performance. Monismith, CL and F.Long [3] recommended asphalt mixture's fatigue limit criterion as 60με. Carpenter.et.al[1] concluded that the fatigue limit of asphalt mixture with modified binder can reach 300με，and as for a mixture with an ordinary binder, its fatigue limit was around 70με.

Four levels of controlling strain were selected in the study, which were 70με, 100με, 200με and 400με. The loading frequency was 10Hz and the loading mode was sine waveform.

In perpetual pavement design, it is widely considered that increasing the bitumen content can enhance the pavement's fatigue resistance. Therefore, in order to find out the effect of increasing the binder content, we chose two levels of asphalt content: 3.6% (the optimum asphalt content) and 4.1% (3.6% + 0.5%).

4 FATIGUE TEST RESULTS ANALYSIS

The data of beam fatigue tests can be analyzed by two methods. The first is according to AASHTO T321 [4]. By this method the fatigue life is defined as a loading cycle number when the beam's stiffness modulus has decreased to 50% of the initial stiffness modulus. The second method adopts "Reduced Energy Ratio (RER)" [2] as the index to analyze fatigue test data, which can account for developing stages of fatigue failure.

4.1 AASHTO T321

In AASHTO T321, the failure criterion of mixture fatigue test is defined as that the specimen's stiffness modulus decreases to 50% of the initial one [4]. The specimen's initial stiffness modulus is the stress–strain ratio at 50th loading cycles.

The beams' fatigue lives at several strain levels are presented in table 2 and Figure 3 (At the stain level of 100με and 70με，the fatigue life was obtained by extrapolation. First the fitting curve was drawn by the relationship of the testing stiffness modulus and the loading times, then the beam's fatigue life was acquired by extrapolating the fitting curve to 50% initial stiffness modulus).

Table 2 Average Fatigue Lives at Different Strain Levels(AASHTO T321)

Strain level /με	Optimal asphalt content 3.6%			Optimal asphalt content +0.5% (4.1%)		
	Test piece	average/times	C_V/%	Test piece	average /times	C_V/%
400	6	3675	16	6	3512	13
200	6	25,401	15	6	29,807	12
100	4	31,224,268	19	4	80,601,582	23
70*	4	85,705,664	22	4	108,254,987	21

note* : Fatigue lives at 100με and 70με are acquired by extrapolation

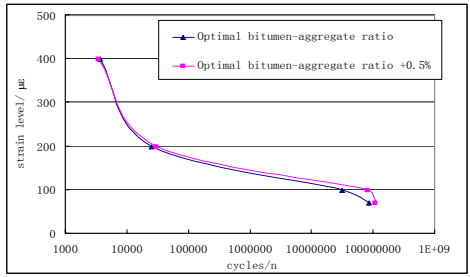
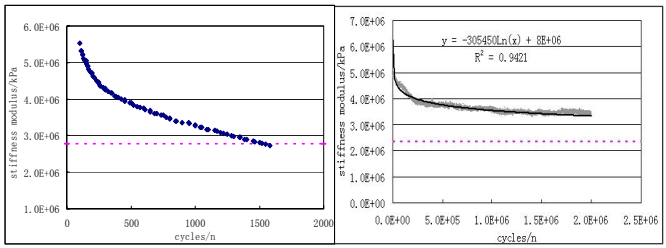


Figure 3 The Beams' Average Fatigue Lives at Different Strain Levels

Figure 4 presents the relation curves between the stiffness modulus and loading cycles. The strain levels are 400με and 70με respectively.



(a) controlling strain 400 $\mu\varepsilon$ (b) controlling strain 70 $\mu\varepsilon$

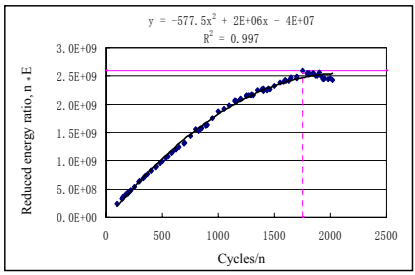
Figure 4 The Relationship of Loading Times And Stiffness Modulus

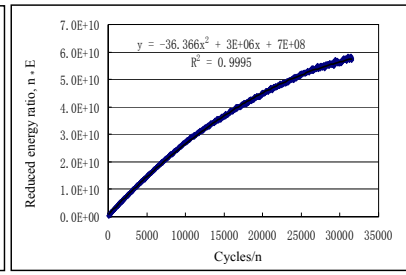
As shown in Figure.4a, at the strain level of 400με, the test beam experienced 1600 loading cycles when its stiffness modulus decreased to 50% of the initial. However, when the strain level was 70με, the test beam could undergo 2,000,000 loading cycles without failure, as Figure 4b shows. We can also conclude from Figure 4b that at the 70με strain level, the relation curve of stiffness modulus vs. loading cycles became flatter with the increasing of loading cycles and the last part of the curve tends to be horizontal. Therefore, controlling the strain level has a great effect on the test beam's fatigue life. When the strain level was very low (between 100με and 70με), it seemed that failure of the test beams was impossible. This phenomenon verifies the theory of asphalt mixtures' Fatigue Limit.

As table 2 and Figure 3 have shown, the test beams' fatigue lives increased with the decrease of strain level. When the strain level was below 200με, the strain-fatigue

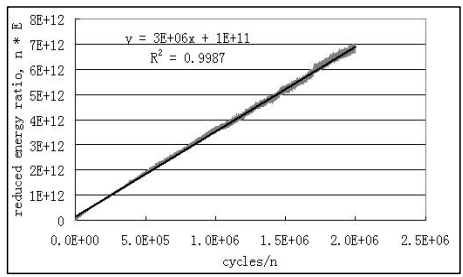
life curve became flat. This means that when the strain level is below 200με, reducing the strain level will be more efficient to increase the fatigue life of the asphalt beam . When the strain level was 70με or 100με, the test beam's fatigue life was between 3 and 10 hundred million times which is almost impossible to reach. Therefore we can conclude that LSM mixture's "Fatigue limit" is between 70με and 100με.

At the same time we can also see from table 2 that at the strain level of 400με or 200με, the bitumen content has little effect on the mixture fatigue characteristics and that when the strain level is low (between 100με and 70με), the test beams with high asphalt content exhibited better fatigue resistance. So at a low strain level, we can increase the asphalt content to enhance the asphalt mixture's fatigue resistance. But when the strain level is between 100με and 70με, the asphalt mixture's fatigue performance is completely satisfactory; then it is not necessary to increase the asphalt content to improve asphalt mixture's fatigue characteristics.





（a）asphalt content 4.1%， strain 400 $\mu\varepsilon$ （b）asphalt content 4.1% ,strain200 $\mu\varepsilon$



（c）asphalt content 4.1%， strain 70 $\mu\varepsilon$

Figure 5 Cycles To Failure Vs. Reduced Energy Ratio

4.2 Reduced Energy Ratio (RER)

Rowe and Bouldin found that "Reduced Energy Ratio (RER)" can be used to analyze the data of the asphalt mixture fatigue test [3]. The Reduced Energy Ratio was defined as the following:

$$R^* = n \times E_n^*$$

where R^* = reduced energy ratio of beam fatigue tests（strain controlling mode）;
n = cycle number;
E_n^* =Modulus at n-th cycle;

According to the research of Rowe and Bouldin，the fatigue failure process of a mixture can be divided into four stages：Internal Heating → Micro-crack Formation →Crack Formation → Sample Breakdown. Rowe and Bouldin defined "Crack Formation" as the fatigue criterion. The failure point can be easily found through a curve between RER and loading cycles which possess a quite visible peak (Figure 5a).

Figure 5 shows the curves of RER vs. the loading cycles of the asphalt mixture beam at different strain levels (one specimen at every strain level).

We can see from figure 5 that at the 400με, there is an apparent peak point at the curve of loading cycles vs. reduced energy ratio. But when the strain level is between 70με and 100με (Figure 5c), the trendline is almost a straight line that will never reach a peak point. Therefore we can conclude that at these low strain levels, asphalt mixture beams will never reach the "Crack Formation" stage (fatigue failure criterion defined by Rowe and Bouldin). Then, by the Reduced Energy Ratio analysis method, we further validate the theory of the "fatigue limit" of asphalt mixture.

Table 4 presents the cycle number as the fatigue tests reached the "Crack Formation" stage.

Table 4 Fatigue Test Beams' Cycles To Failure (Reduced Energy Ratio Analysis Method)

Strain level /με	Optimal asphalt content 3.6%			Optimal asphalt content +0.5% （4.1%）		
	The number of beams/n	cycles/n	C_V/%	The number of beams/n	cycles/n	C_V/%
400	6	5040	14	6	4275	12
200	6	29,626	18	6	41,350	15
100	4	/	/	4	/	/
70	4	/	/	4	/	/

As shown in table 2 and table 4, by the AASHTO T321 analysis method and the Reduced Energy Ratio analysis method, we can come to the similar conclusion that the LSM asphalt mixture has a"Fatigue Limit" at a low strain level（70με～100με）.

5. CONCLUSIONS

In order to verify asphalt mixtures' fatigue limit theory and find out the fatigue endurance limit of an LSM base, fatigue tests of two different asphalt contents and four strain levels were conducted, and two analysis methods were used to analyze the test results. From these tests and analyses, the following conclusions have been achieved:

1. At the low strain level（between 70με and 100με）, fatigue tests of LSM showed the "Fatigue Limit" phenomenon, which supports the asphalt mixtures' fatigue limit theory and offers an experimental basis for perpetual pavement design.

2. We can come to the similar conclusion by the AASHTO T321 analysis method and the Reduced Energy Ratio analysis method, that it is impossible for LSM to fail in fatigue tests when the strain level is about 70με. Consequently, in perpetual pavement design, it is proper to set the fatigue limit of the LSM base as 70με.

3. The relationship between asphalt content and fatigue life was analyzed. At high strain level, increasing asphalt content can not enhance the mixture's fatigue resistance distinctly. However, at a lower strain level, the mixture's fatigue performance can be improved significantly by increasing the asphalt content.

References

[1] Carpenter S. H. K. Ghuzlan, and S. Shen, Fatigue Endurance Limit for Highway and Airport Pavements. Transportation Research Record: Journal of the Transportation Research Board, No. 1832, TRB, National Research Council, Washington, D.C., 2003.

[2] G.M. Rowe, and M.G. Bouldin, Improved Techniques to evaluate the Fatigue Resistance of Asphaltic Mixtures. 2nd Euroasphalt and Eurobitume Congress, Book 1, Barcelona, 20-22 September, 2000, pp754-763.

[3] Monismith, C. L., and F. Long. Overlay Design for Cracked and Seated Portland Cement Concrete (PCC) Pavement—Interstate Route 710. Technical Memorandum TM UCB PRC 99-3, Pavement Research Center, Institute for Transportation Studies, University of California, Berkeley, 1999.

[4] Standard Test Method For Determining the Fatigue Life Of Compacted Hot Mix Asphalt (HMA) Subjected to Repeated Flexural Bending. AASHTO Designation TP8-94(reapproved 1996),1995；

Compaction and Testing Method of Wind-blown-sand Subgrade

XUE, Ming[1]， SONG, Dandan[2], and ZHANG, Jun[2]
[1] Department of the road and airport, Tongji University, PH (021) 6598-3239; FAX (021) 6598-3239; email: xueming303@163.com
[2] Department of the road and airport, Tongji University, PH (021) 6598-3239; FAX (021) 6598-3239; email: doubledansongrz@163.com

Abstract: With the expansion of the West Development, the investment in western traffic infrastructure is increased; however, some important issues regarding road construction are thus gradually emerging. For instance, one of the problems is related to highway construction materials. There has been experimental research and highway engineering practice at Mao Wu Su Desert in inner Mongolia Autonomous Region. This paper briefly introduces the main physical and mechanical properties of wind-blown-sand, opens out the rule of its compacted density, proposes the approach for determining the standard density of wind-blown-sand, analyzes the construction procedures of compaction without water and the methods of its compacted density.
The keyword: wind-blown-sand, subgrade, compaction, testing

1. Highway engineering at the desert

On our planet, a large area is covered by desert. The reported area and distribution of desert all over the world is shown in Table.1.

Table 1. The Stat of area of desert over the world

Range of stat	Asia	Africa	Australia	America	Summation
Area of desert (ten thousand km^2)	261.7	165.7	113.6	14.5	555.5

With the expansion of human activities all over the world, the demand of highway engineering in the desert area keeps increasing. The total area of desert in our country (gravel desert, stone desert and salt desert not included) is about 700,000 km^2, which lies mostly in the provinces of Sinkiang, Neimeng, Gansu, and others. With the expansion of the West Development, the development of western traffic infrastructure is fast. Some special construction problems emerge which need to be solved urgently. One of the problems is the supplying of the materials for the construction. The special factor of the supplying of material in the desert is its huge quantity and singularity. There is sand all over the area which can never be used up; singularity means that there is no other material for construction except sand, and if it is necessary to choose other materials, the distance and cost for transporting them is unrecoverable. Therefore, the important task of highway construction in a desert is how to use sand to construct a

usable highway.

2. Highway engineering and Compaction

According to the classical theory, the carrying capacity of the layer is directly proportional to the density: with a certain material, the carrying capacity will increase when the density of the construction material in unit volume increases. Therefore, the task of compaction is accorded great importance during the construction of a highway. Financial efforts are made to achieve the utmost uniform compaction as possible; compaction is a very important chapter in the current construction criteria and the degree of compaction is a very important parameter for control of construction and the check for engineering.

To the compaction of normal material for construction, abundant experience has accumulated in the group of highway engineering, including the establishment of standard density, the implement of compaction with grind, and the testing of degree of compaction, etc.

3. The physical and mechanical properties of wind-blown-sand and its effect on compaction

Compared to the normal construction material, the very special physical and mechanical properties of wind-blown-sand have impacts on its compaction. The main factor infection of compaction of wind-blown-sand is the composition of the grain. The test results of the sample of the locale of the testing road for research is shown in Table 2 and Figure 1.

Table 2. Composition of the Wind-blown-sand grain from the place of highway for test

Aperture of sieve pore (mm)	First row		Second row	
	Quality of griddling (g)	Proportion	Quality of griddling (g)	Proportion
4.750	0.0	0.00%	0.0	0.00%
2.360	0.0	0.00%	0.0	0.00%
1.180	0.0	0.00%	0.0	0.00%
0.600	1.9	0.38%	1.9	0.38%
0.300	63.6	12.72%	62.7	12.54%
0.150	178.7	35.74%	180.6	36.12%
0.075	250.7	50.14%	249.4	49.88%
<0.075	5.5	1.10%	5.2	1.04%
Total weight	500.0	100.00%	500.0	100.00%

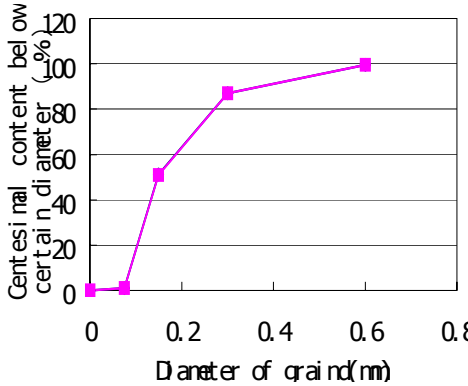


Figure 1. The accumulative curve of grain of the wind-blown-sand

According to the result, the diameters of the grain of wind-blown-sand mostly are between 0.3～0.075mm, and the grade is poor because the grain is too weak. There is no grain large enough for main grain, nor the muscovite grain for impelling material aggregating among the composition of the grain. It is difficult to form structure through compaction of this kind of material, and it is very loose while being used because of the poor intensity for stress.

4. The Standard Density

4.1 The Research Approach

The maximum dry density of material is one of the most important control indices for highway engineering. The best moisture content of the soil is one of the most important control indices to ensure the standard design dry density during construction, so it is necessary to establish the most dry density and the best moisture content during the design phase and seedtime for construction before construction. At present, the standard density of soil for highway engineering is determined through compaction tests, the standard density of the asphalt mixture through Marshal Test, and the standard density of material with more grains through Vibrating compaction. The research team adopts Proctor compaction test，Modified Proctor test，Marshal test and Vibrating compaction for comparison to chalk up the most appropriate approach to suit the specialty of wind-blown-sand.

Through Modified Proctor test, the relation between dry density and moisture content of wind-blown-sand is shown in Table.3:

Table 3. Relation between dry density and moisture content of Modified Proctor test

ω（%）	0.26	0.99	2.75	3.07	5.02
ρ（kg/m^3）	1.694	1.671	1.688	1.700	1.702

Through Marshal Test, the relation between dry density and moisture content of wind-blown-sand is shown in Table.4:

Table 4. Relation between dry density and moisture content of Marshal Test

ω(%)	0.20	1.02	2.56	3.00	5.04
ρ(g/cm^3)	1.737	1.718	1.710	1.717	1.722

Vibrating compaction is fulfilled at the standard librated stage. The main factors that affect the degree of density after librations include 1）the frequency and swing of the standard librated stage; 2）the time of vibration; 3）the load quality of the top face of the sands. The flat vibrator with frequency of 50HZ, swing of 0.35mm, power of 1.1kw is used during research. Blank -5 and Table.6 reflect the impact on dry density by the time of vibratory and the load quality on the top face of samples:

Table 5. Relation between density and the time of vibratory without press block

time of vibratory (s)	30	60	90	2×15	2×30	2×45
ρ（g/cm^3）	1.571	1.621	1.633	1.593	1.621	1.649

Table 6. Relation between density and the time of vibratory with press block

Quality of press block（g）	627	936	1232	1574	627	936	1232	1534
time of vibratory（s）	60	60	60	60	2×30	2×30	2×30	2×30
ρ（g/cm^3）	1.680	1.687	1.694	1.710	1.671	1.683	1.695	1.695

4.2 The Main conclusion of research

Two main conclusions are gained through the tests above-mentioned and research:

4.2.1 There are two specialties of compaction of wind-blown-sand

Following the data of Table.3 and Table.4, the rule of compacting of wind-blown-sand is different from that of normal soils. There are two specialties of compaction of wind-blown-sand, and there are two peak values during the proceeding of compaction of wind-blown-sand. One is during the phase of lower moisture content and the other is during the phase of best moisture content that also exists within

normal soils and could be the maximum when the moisture content is very low (shown in Figure.2). The normal soil has only one peak value at the point of optimum moisture content which is also the maximum dry density of the soil.

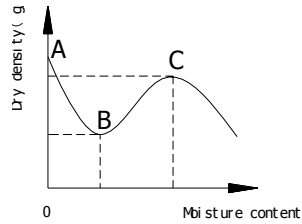


Figure 2. Relation between dry density and moisture content of wind-blown-sand

4.2.2 The maximum dry density through several different compaction methods is shown in Table.7:

Table 7. Comparison of densities through several compaction methods

Compaction test	Dry density (g/cm^3)
Proctor compaction test	1.652
Modified Proctor test	1.715
Vibrating compaction	1.71
Marshal test	1.735

According to the research, it is obtained that the Marshall test has a large contact area with sand, and the sand is loaded equally. The harmer and the tube for compaction both have a confining force on the sand effectively, which could settle the problem of hump of wind-blown-sand during compaction, and then the compaction work is mostly for conquering the frictional resistance between the grains of sand. On the condition of equality of the total compaction work, the test of harmer of Marshall for compaction is better than standard heavy duty for compaction. Also, the test of harmer of Marshall for compaction is better than the vibration method.

5. Compaction of wind-blown-sand without water

Following the analysis for the composition of wind-blown-sand and other mechanical tests, the grains of wind-blown-sand lack adhesion, which causes serious lack of shear strength of wind-blown-sand. When construction vehicles pass by, the surface of compaction will become displaced and loose owing to the backward horizontal force posed by the vehicles.

The main factors that affect the compaction of the wind-blown-sand include the surface characteristics (grain composition, grain shape, and so on), mechanical

characteristics and combination of machines for compaction. There are two segments of the rule for compaction of wind-blown-sand on the job site for the best effect of compaction, and they are shown in Table.7 and Figure.3.

Table 8. Degree of compaction changing with depth

Depth (cm)		0	10	20	30	40	53	68	80
Dry density(g /cm^3)	K109+650 Over lapping twice 1/3	1.64	1.67	1.69	1.69	1.695	1.71	1.69	1.68
	K109+67 Over lapping twice 1/4	1.65	1.695	1.715	1.71	1.72	1.72	1.695	1.66

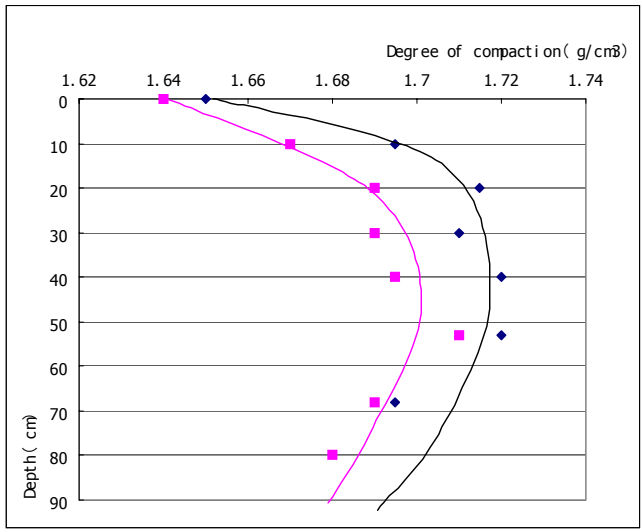


Figure 3. Degree of compaction changing with depth

Conclusions that arise from the analysis of the chart above:

□. The density doesn't increase but declines between the range of depth of 10—15cm, because of the horizontal force bacwards of the vehicles for construction. But the density all increases below the depth of 15cm.

□. The depth of work by the compaction machine could reach 90cm, and the optimum depth is 20~70cm.

□. The way for compaction of over lapping 1/4 is better than 1/3 for the increase of density, and the way for compaction of over lapping 1/4 for four times is much higher than 1/3, so the way for compaction of over lapping 1/4 for four times is adopted in this test.

The following formula is gained through the rules above for calculating the maximum gauge for filling the field:

H = γ(HO - H′)

H — the maximum gauge for filling every layer（cm）

HO — the effective gauge of compaction by the compaction machine during construction（cm）

H′— the turbulent gauge by horizontal force backwards of the vehicles for construction（cm）

γ — the index for pavement decided by test

6. Test for density of the structure of wind-blown-sand

The density is one of the most important control indices for the construction of a road base. The wind-blown-sand lacks viscosity and main aggregate, so it can't be used in the normal way as used in normal highway engineering ——cutting ring method, filling sand method, and filling water method to test the density of wind-blown-sand without water. As it can't keep its initial form when sampling from the structure, the sides of a tunnel for sampling may collapse easily. The research team has devised an approach to test the on-site density.

To be specific, nail in a metal confining tube into the compaction structure for testing, then draw out the sand in the tube and test the wet weight of the sand, the weight of moisture content, the dry weight as normal method; and test the volume of the sand from the tube through a filling sand method or a filling water method. Calculate the density of the dry sand g:

Formula: $g = \frac{m_0/(1+\omega_0)}{V} = \frac{m_0}{(1+\omega_0)\bullet V}$

m_0 ——the weight of the sand drawn out of the tube（g）;

ω_0 ——the natural moisture content of the sand drawn out of the tube（g）;

V ——the volume of the sand drawn out of the tube (cm^3)。

Conclusion The Marshall test has a large contact area with sand, the sand is loaded equally, the harmer and the tube for compaction both have effective confining force on the sand, which could settle the problem of humping of wind-blown-sand during compaction, and then the compaction work is almost for conquering the frictional resistance among the grains of the sand. On the condition of equality of the total compaction work, the Marshall test for compaction is not only better than Proctor compaction test but also better than the vibration test. Through the test highway engineering improvement, the maximum dry density of the Marshal test

mostly occurs in the engineering.

(2) The compaction of wind-blown-sand is affected by the variations of the grains of the wind-blown-sand and the compaction machine. There is a loose layer owing to the horizontal force backwards of the vehicles for construction. The thickness of the layer usually is 15cm and there is a transition layer for the weak compaction below the depth of 15cm.The efficiency depth with compaction of the CA-30 vibration road roller used in the test road will reach 90cm, and the best compaction layer will be during 30~70cm.

(3) On the condition of using pure wind-blown-sand to fill the road base, the procedures of compaction mentioned above could be adopted which could ensure the achievement of the design demand of the filling layer.

(4) During the test for density of an on-site road, when the moisture content of the sands $\omega>1.2\%$, the agglomerating force of the sand is enough to ensure the stability for a sample with the cutting ring method. Then the normal cutting ring method could be adopted. When the moisture content of the sand is $\omega<1.2\%$, the agglomerating force of the sand is enough to ensure the stability for the sample with the cutting ring method, and then the confining method should be adopted.

References:

LI, Lihan (1999). “Construction material of road”. Tongji University Press.

MINISTRY OF COMMUNICATIONS OF PRC (1999). “Test Methods of Soil for Highway Engineering”. China Communication Press.

MINISTRY OF COMMUNICATIONS OF PRC (2004). “Technical Specification for Construction of Highway Subgrades” .China Communication Press.

YAO, Zukang(1987). “Subgrade and pavement engineering.” Tongji University Press.

YAO, Honglin (1998). “Theory of desert.” Neimenggu People Press.

Micromechanical Finite Element Models for Micro-Damage and Complex Constitutive Behavior of Asphalt Mixes

Qingli Dai [1] and Zhanping You [2]

[1] Research Assistant Professor, Department of Mechanical Engineering - Engineering Mechanics, Michigan Technological University, Houghton, MI 49931; PH: (906) 487-2620; e-mail: qingdai@mtu.edu
[2] Assistant Professor, Department of Civil and Environmental Engineering, Michigan Technological University, Houghton, MI 49931. PH: (906) 487-1059; e-mail: zyou@mtu.edu

Abstract: This paper presents a comprehensive review of recently developed micromechanical modeling approaches for asphalt mixture in pavement engineering by the authors. These approaches include the Micro-Frame Finite Element Network Model (FENM) and microstructure-based Continuum Finite Element Model (CFEM) for predicting micro-damage behavior and constitutive properties of asphalt mixtures. Both models consider the multi-phase and heterogeneous microstructure and aggregate-to-aggregate interaction of asphalt mixture. Both finite element modeling schemes and their advanced features were described and addressed. Based on these developed models, the virtual laboratory simulation was conducted to compare with the laboratory specimen test results. Simulation samples were generated from surface photographic data of actual laboratory specimens by using image processing. Model simulation results of the overall sample behavior and evolving micro-damage patterns compared favorably with experimental data collected on the test specimens.
Keywords: a*sphalt mixes, micromechanical modeling, finite element method, damage mechanics, constitutive behaviors, image analysis*

Objectives and Roadmap

The objectives of this paper are to: 1) comprehensively review the micromechanical FE modeling of asphalt mixture done by authors and 2) discuss the model's applications for numerical simulation of micro-damage and complex constitutive behavior.

To help clarify the objectives of this study, a brief background description on the asphalt mixture and micromechanical FE models is provided. A Finite Element Network Model (FENM) is demonstrated for damage simulation of 2D and 3D asphalt specimens. After this, a microstructure-based Continuum Finite Element Model (CFEM) for asphalt mixture is described to predict the complex constitutive behavior of asphalt mixtures. Other advances in asphalt mixture modeling including the modeling of three-dimensional (3D) microstructure of asphalt mixture are then briefly discussed.

Introduction

Asphalt mixtures are complex heterogeneous materials that consist of graded aggregates, asphalt binder, additives and void space. Some research has been conducted on the micromechanical behaviors of the particulate, porous and

heterogeneous materials. For example, studies on cemented particulate materials by (Dvorkin et al. 1994) provide information on the load transfer between particles which are cemented by an elastic binder, and studies on inter-particle behavior for two elastic particles bonded by a viscoelastic binder can be found in the work by (Chang and Meegoda 1997; Zhu 2001). Such elastic and viscoelastic mechanics provide details on the normal and tangential interparticle load transfer, and would be fundamental in developing a micromechanical theory for load distribution and failure of such materials.

Finite Element Modeling (FEM) studies has been conducted to investigate asphalt mixture behavior including (Bahuguna et al. 2006; Buttlar et al. 2006; Dai et al. 2006a; Dai and You 2007; Dai et al. 2006b; Guddati et al. 2002; Kim et al. 2005; Panneerselvam and Panoskaltsis 2005; Papagiannakis et al. 2002). A common finite element approach to simulate heterogeneous particulate materials employs the idea of representing the interparticle behavior using an equivalent lattice network system. This type of microstructural modeling has been used by (Bazant 1990; Budhu et al. 1997; Mora 1992; Sadd et al. 1992). In addition, some research has been reported on the three-dimensional (3D) finite element model of asphalt pavement including (Al-Qadi et al. 2003; Dessouky et al. 2006; Elseifi et al. 2006; Novak et al. 2003; Yoo et al. 2006). X-ray techniques were also applied to capture the 3D microstructure of asphalt mixture such as (Masad 2004; Wang et al. 2004).

The damage behavior of asphalt material is extremely complex due to its microstructure. The interior binder is composed of microcracks, microvoids and defects that result in strength loss. As the load is increased, local cracks would start to form near existing cracks and thus would lead to the interior failure and interfacial debonding. Continuum damage mechanics is based on the thermodynamics of irreversible processes to characterize elastic-coupled damage behaviors, (Chaboche 1988). In general, a Continuum Damage Model (CDM) consists typically of three major components: (a) selection of damage variable(s); (b) definition of strain energy density (as a function of damage variables and other state variables); and (c) a damage evolution law. (Schapery 1999) developed a micro-mechanics based constitutive model for viscoelastic composite solids with growing damage. This constitutive modeling work was extended by (Hinterhoelzl and Schapery 2004; Park et al. 1996; Park and Schapery 1997; Schapery 2002).

Micro-Frame Finite Element Network Model

Bituminous asphalt concrete can be described as a multi-phase material containing aggregates, mastic cement (including asphalt binder and fine aggregates) and air voids. The load transfer between the neighboring aggregates plays a primary role in determining the load carrying capacity and failure of such complex materials. Micro-frame FENM assumes that the micro-mechanical load transfer occurs through the effective mastic zone between neighboring aggregates. The 2D equivalent micro-frame network approach was applied to asphalt mixtures, where the local interaction between neighboring particles was modeled with a special frame-type finite element (Dai 2004; Dai et al. 2005; Sadd et al. 2004a; Sadd et al. 2004c). The element stiffness matrix is constructed by considering the normal, tangential and rotational behaviors between cemented particles, and this is accomplished using an approximate elasticity solution from (Dvorkin et al. 1994) for the stress distribution within the

cementation interface. Details of the calculation of the various stiffness terms have been reported previously by (Dai et al. 2005), and the final result is given by

$$[K]=\begin{bmatrix} K_{nn} & 0 & K_{nn}e & -K_{nn} & 0 & -K_{nn}e \\ \cdot & K_{tt} & K_{tt}r_1 & 0 & -K_{tt} & K_{tt}r_2 \\ \cdot & \cdot & K_{tt}r_1^2+\dfrac{K_{nn}}{3}\left(w_2^2-w_1w_2+w_1^2\right) & -K_{nn}e & -K_{tt}r_1 & K_{tt}r_1r_2-\dfrac{K_{nn}}{3}\left(w_2^2-w_1w_2+w_1^2\right) \\ \cdot & \cdot & \cdot & K_{nn} & 0 & K_{nn}e \\ \cdot & \cdot & \cdot & \cdot & K_{tt} & -K_{tt}r_2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & K_{tt}r_2^2+\dfrac{K_{nn}}{3}\left(w_2^2-w_1w_2+w_1^2\right) \end{bmatrix} \tag{1}$$

where $K_{nn}=(\lambda+2\mu)w/\bar{h}$, $K_{tt}=\mu w/\bar{h}$, λ and μ are the usual elastic moduli, $\bar{h}$ and w are the average cementation thickness and cementation width, r_1 and r_2 are the radial dimensions from each aggregate center to the cementation boundary, w_1 and w_2 are left and right width of cementation, and eccentricity $e=(w_2-w_1)/2$. This procedure establishes the elastic stiffness matrix for each binder element, which is a function of the local material microstructure and binder moduli.

To simulate the inelastic damage behaviors observed in asphalt materials, continuum damage mechanics was incorporated within the inter-particle cementation model. The damage stiffness matrix $[D_s]$ can be expressed in terms of the initial elastic stiffness matrix $[D_0]$ using continuum damage principles,

$$[D_s]=([I]-[\Omega])[D_0] \tag{2}$$

and a *damage tensor* $[\Omega]$ is defined by considering the reduction of the effective area of load transfer within the continuum.

For the uniaxial normal and tangential behavior of micro-frame element, an exponential damage evolution law was chosen to describe one-dimensional damage behavior. This damage modeling scheme was incorporated into the finite element network model by modifying the micro-frame element stiffness matrix given in Equation (1). Using the uniaxial relation, the incremental normal and tangential damage stiffness terms for the inelastic behavior can be written as

$$\left(K_{nn}\right)_s=K_{nn}e^{-b(\Delta u_n/\Delta U_n)},\ \left(K_{tt}\right)_s=K_{tt}e^{-b(\Delta u_t/\Delta U_t)} \tag{3}$$

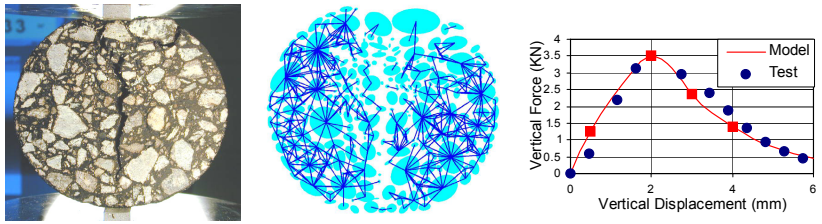
When the element displacements are larger than the softening displacements, the elements will have softening behavior. The corresponding incremental normal and tangential damage softening stiffness are given by

$$\begin{aligned}\left(K_{nn}\right)_s&=-(K_{nn}m/b)(1-e^{-b})e^{m(1-\Delta u_n/\Delta U_n)}\\ \left(K_{tt}\right)_s&=-(K_{tt}m/b)(1-e^{-b})e^{m(1-\Delta u_t/\Delta U_t)}\end{aligned} \tag{4}$$

where b and m are material parameters in the damage evolution laws, Δu_n and Δu_t are the normal and tangential accumulated relative displacements and ΔU_n and ΔU_t are the normal and tangential displacement softening criteria. Thus the micro-frame element incremental damage stiffness matrix $[K_s]$ is constructed from Equation (1) by replacing K_{nn} and K_{tt} with $\left(K_{nn}\right)_s$ and $\left(K_{tt}\right)_s$. Displacement-based softening criteria for tension, compression and shear are defined with accumulated relative displacements between particle pairs, where softening displacement = multiplication of softening factor and micro-frame element geometries. The stress-based failure

criteria for three behaviors are determined with the multiplication of the failure factor and maximum stress. When the softening stress is less than the failure stress, the element is failed. Failed elements still remain in the computational model and have very low loading supportability.

The micro-frame finite element model was applied to simulate the damage behavior of an indirect tension and a compression sample obtained from the scanned specimens as shown in the Figures 1 and 2. These indicate that the micro-frame damage model has good ability to predict the sample fracture behaviors.



(a) Damaged specimen (b) Model simulation results (c) Overall load-deformation behavior

Figure 1. Fracture patterns and load-deformation behavior comparisons for indirect

(a) Damaged specimen (b) Model simulation with removal of failed element

Figure 2. Fracture patterns in compression test and simulation

Following these approaches, the 3D micro-frame element network models were developed for the 3D behavior simulation of asphalt mixture. The continuum damage mechanics principles were also incorporated into 3D micro-frame network model to simulate the microstructure-based damage behavior. The 3D load transfer between each aggregate pair includes resultant normal and tangential forces and moments (six components). The 3D micro-frame element stiffness matrix was developed based on unit normal, tangential and rotational particle motion cases. Damage mechanics principles were also applied to the 3D micro-frame element behavior for 3D micro-damage simulation. This part of the work is still ongoing and the details will be published soon.

Micromechanical Continuum Finite Element Model

As mentioned previously, asphalt mixtures consisting of very irregular aggregates and complex distributed mastic were divided into different aggregate and mastic sub domains. The finite element mesh was generated within the sub domains (aggregates and mastic) and along the sub domain boundaries. Figure 3 shows the

finite element meshes in the aggregates and mastic sub domains of a specimen surface. Finite elements in the neighboring sub domains share the nodes on irregular boundaries, and therefore the displacements of neighboring sub domains were connected through the shared nodes. Then the linear/damage-coupled viscoelastic mastic with specified properties defined in an ABAQUS user subroutine was combined with elastic aggregates to simulate global linear and damage-coupled viscoelastic behavior of asphalt mixtures.

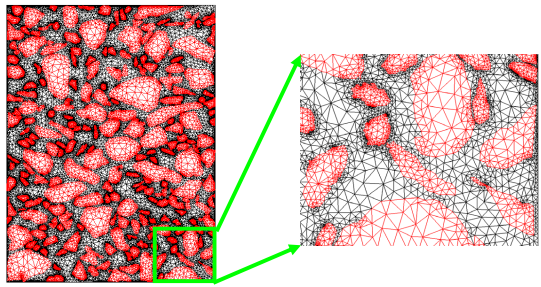


Figure 3. The FEM meshes for the aggregate and mastic

The linear constitutive behavior for Maxwell-type viscoelastic model can be expressed as a hereditary integral

$$\sigma_{ij} = E_{\infty}\varepsilon_{ij} + \int_0^t E_t \frac{d\varepsilon_{ij}(\tau)}{d\tau} d\tau \,, \tag{5}$$

where E_t is the transient modulus changing with time, and expressed with a Prony series

$$E_t = \sum_{m=1}^{M} E_m e^{-\frac{(t-\tau)}{\rho_m}} \text{, and } \rho_m = \frac{\eta_m}{E_m} \tag{6}$$

In these equations, E_m, η_m and ρ_m are the spring constant, dashpot viscosity and relaxation time respectively for the m^{th} Maxwell element.

The reduced time (effective time) is defined by using time-temperature superposition principle as

$$\xi(t) = \int_0^t \frac{1}{\alpha_T} d\tau \tag{7}$$

where the term $\alpha_T = \alpha_T(T(\tau))$ is a temperature-dependent time-scale shift factor.

The damage-coupled viscoelastic behavior was obtained from the original Schapery nonlinear viscoelastic model (Schapery 1969), given as

$$\sigma(\xi) = h_e(\varepsilon_{\max})E_{\infty}\varepsilon(\xi) + \int_0^{\xi} h_1(\varepsilon_{\max})E_t(\xi - \xi')\frac{d(\varepsilon(\xi'))}{d\xi'} d\xi' \tag{8}$$

where h_e and h_1 are the elastic and viscoelastic damage variables for the rate-independent failure behavior. These variables are functions of the maximum strain $\varepsilon_{\max}$.

The elastic damage variable $h_e = 1 - \Omega$ measures the relaxed elastic stiffness reduction, and can be described by using the inelastic damage evolution law in (Sadd et al. 2004a; Sadd et al. 2004b),

$$h_e(\varepsilon_{\max}) = e^{-b\frac{\varepsilon_{\max}}{\varepsilon_0}} \tag{9}$$

where the material parameters ε_0 and b are related to the softening strain and damage evolution rate respectively.

The viscoelastic variable h_1 measures the damage effect in the transient modulus, and is chosen with the following exponential form by (Simo and Ju 1987),

$$h_1(\varepsilon_{\max}) = \beta + (1-\beta)\frac{1 - e^{-\left(\frac{\varepsilon_{\max}}{\varepsilon_0}\right)}}{\varepsilon_{\max}/\varepsilon_0},\ \beta \in [0,1] \tag{10}$$

The variable h_1 will reduce from 1 to β as the maximum strain $\varepsilon_{\max}$ increases, and ε_0 is also the softening stain.

Three-dimensional behavior can be formulated with uncoupled volumetric and deviatoric stress-strain relations. A displacement-based incremental finite element modeling scheme with constant strain rate over each increment was developed in the following format:

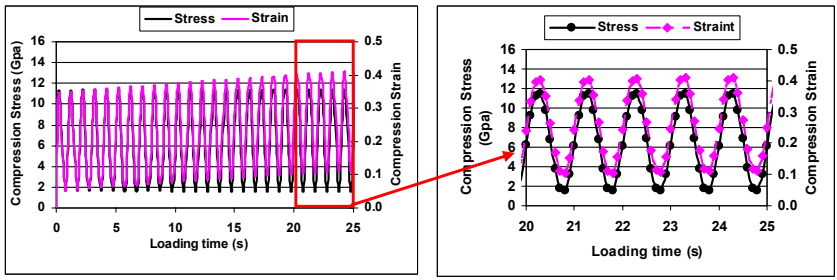
$$\Delta\sigma = K \cdot \Delta\varepsilon + \Delta\sigma^R \tag{11}$$

Where $\Delta\sigma$ and $\Delta\varepsilon$ are incremental stress and strain, K is the incremental stiffness and $\Delta\sigma^R$ is the residue stress vector. Detailed information on incremental finite element formulation can be checked from the work by Dai et al. (2006a).

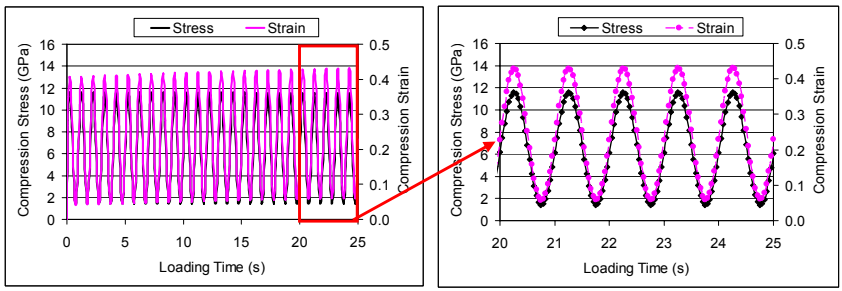
Sinusoidal cyclic force loading was imposed to the computational specimen shown in Figure 3 to conduct uniaxial compression simulation under the different frequencies (0.1 Hz, 1 Hz, 5 Hz and 10 Hz). The laboratory tests of elastic aggregate (rock) and linear viscoelastic mastic were conducted to provide material input parameters for the FE models. The aggregate modulus takes a typical modulus of 55.5GPa for the limestone. The evaluated Maxwell model for reclaimed asphalt mastic includes one spring and four Maxwell elements in parallel ($E_\infty = 59.7MPa$, $E_1 = 5710.6MPa$, $\tau_1 = 26.2s$, $E_2 = 2075.1MPa$, $\tau_2 = 311.9s$, $E_3 = 1449MPa$, $\tau_3 = 1678.8s$, $E_4 = 734.9MPa$, $\tau_1 = 19952.6s$). Model damage parameters were chosen as b =1, $\beta = 0.3$, and softening strain $\varepsilon_0 = 0.2$ for the simulation.

Figure 4 shows the cyclic simulation under the frequency of 1 Hz. The simulation was conducted with elastic aggregate and linear/damage-coupled viscoelastic mastic properties. In the figures, the black curve is the imposed compression stress loading with left-hand axial scales, and the magenta color indicates the strain response with right-hand axial scales. For better illustration, the final several cycles were magnified in the right-side figures for each frequency. The figures show the increasing creep compression strain along the loading time for the constant force loading magnitude. Comparing results (a) and (b) using linear and damage-coupled viscoelastic mastic properties, it shows viscoelastic damage behavior increases the sample creep deformation under the constant cyclic force loading. From the results of different loading frequencies, it was found that the uniaxial compression strain

responses decease with the loading frequencies due to the less relaxation time. 3D continuum finite element model are developed based on the scanned images of asphalt mixture specimens. The 3D FE model will provide more aggregate interactions and yield better solution than 2D. This work is still actively ongoing.



(a) Linear viscoelastic mastic



(b) Damage-coupled viscoelastic mastic

Figure 4. The FEM simulation results under sinusoidal loading at a frequency of 1Hz

Conclusions

The micromechanical finite element network model has been used to simulate the micro-damage behavior of asphalt concrete samples. Both indirect tension and compression samples have been investigated under typical laboratory loading conditions. The comparison indicates that the FENM has the ability to predict the typical damage behavior in asphalt mixture under the loading. The micromechanical Finite Element (FE) model by incorporating elastic aggregates and linear/damage-coupled viscoelastic mastic has been used to simulate viscoelastic behavior of asphalt mixture. Simulation results under a loading frequency showed a viscoelastic creep behavior under the constant loading magnitude. The damage-coupled viscoelastic mixtures generated larger creep deformation comparing the linear behavior. Three-dimensional micromechanical modeling and viscoplastic constitutive behavior are actively investigated.

References

Al-Qadi, I. L., Elseifi, M., and Leonard, D. (Year). "Development of an overlay design model for reflective cracking with and without steel reinforcing nettings." Association of Asphalt Paving Technologist, White Bear Lake, MN 55110, United States, Lexington, KY., United States, 388-423.

Bahuguna, S., Panoskaltsis, V. P., and Papoulia, K. D. (2006). "Identification and modeling of permanent deformations of asphalt concrete." *Journal of Engineering Mechanics*, 132(3), 231-239.

Bazant, Z. P., Tabbara, M.R., Kazemi, Y., Pijaudier-Cabot, G. (1990). " Random particle simulation of damage and fracture in particulate or fiber-reinforced composites. ." In: *Damage Mechanics in Engineering Materials*, Trans. ASME, AMD.

Budhu, M. S., Ramakrishnan, and Frantziskonis, G. (Year). "Modeling of Granular Materials: A Numerical Model Using Lattices, Mechanics of Deformation and Flow of Particulate Materials." *McNu Conference,Trans. ASCE*, Northwestern Univ.

Buttlar, W. G., Paulino, G. H., and Song, S. H. (2006). "Application of graded finite elements for asphalt pavements." *Journal of Engineering Mechanics*, 132(3), 240-249.

Chaboche, J. L. (1988). "Continuum damage mechanics: part I - general concepts." *Journal of Applied Mechanics, Transactions ASME*, 55(1), 59-64.

Chang, G. K., and Meegoda, J. N. (1997). "Micromechanical Simulation of Hot Mixture Asphalt." *ASCE J. Engng. Mech.*, 123(5), 495-503.

Dai, Q. (2004). "Micromechanical Modeling of Constitutive and Damage Behavior of Heterogeneous Asphalt Materials," University of Rhode Island.

Dai, Q., Sadd, M. H., Parameswaran, V., and Shukla, A. (2005). "Prediction of damage behaviors in asphalt materials using a micromechanical finite-element model and image analysis." *Journal of Engineering Mechanics*, 131(7), 668-677.

Dai, Q., Sadd, M. H., and You, Z. (2006a). "A micromechanical finite element model for linear and damage-coupled viscoelastic behaviour of asphalt mixture." *International Journal for Numerical and Analytical Methods in Geomechanics*, 30(11), 1135-1158.

Dai, Q., and You, Z. (2007). "Prediction of Creep Stiffness of Asphalt Mixture with Micromechanical Finite Element and Discrete Element Methods." *Journal of Engineering Mechanics, ASCE.*, 133(2), 163-173.

Dai, Q., You, Z., and Sadd, M. H. (2006b). "A Micromechanical Viscoelaso-Plastic Model For Asphalt Mixtures." In: *Geotechnical Special Publication (GSP): Asphalt Concrete: Simulation, Modeling, and Experimental Characterization. American Society of Civil Engineers (ASCE)* 12-20.

Dessouky, S., Masad, E., Little, D., and Zbib, H. (2006). "Finite-element analysis of hot mix asphalt microstructure using effective local material properties and strain gradient elasticity." *Journal of Engineering Mechanics*, 132(2), 158-171.

Dvorkin, J., Nur, A., and Yin, H. (1994). "Effective Properties of Cemented Granular Materials." *Mech.of Materials*, 18, 351-366.

Elseifi, M. A., Al-Qadi, I. L., and Yoo, P. J. (2006). "Viscoelastic modeling and field validation of flexible pavements." *Journal of Engineering Mechanics*, 132(2), 172-178.

Guddati, M. N., Feng, Z., and Kim, R. (2002). "Toward a micromechanics-based procedure to characterize fatigue performance of asphalt concrete." *Transportation Research Record*(1789), 121-128.

Hinterhoelzl, R. M., and Schapery, R. A. (2004). "FEM implementation of a three-dimensional viscoelastic constitutive model for particulate composites with damage growth." *Mechanics Time-Dependent Materials*, 8(1), 65-94.

Kim, Y.-R., Allen, D. H., and Little, D. N. (2005). "Damage-induced modeling of asphalt mixtures through computational micromechanics and cohesive zone fracture." *Journal of Materials in Civil Engineering*, 17(5), 477-484.

Masad, E. (2004). "X-ray computed tomography of aggregates and asphalt mixes." *Materials Evaluation*, 62(7), 775-783.

Mora, P. (Year). "A Lattice Solid Model for Rock Rheology and Tectonics." *The Seismic Simulation Project Tech.*, Institut de Physique du Globe, Paris, 3-28.

Novak, M., Birgisson, B., and Roque, R. (2003). "Tire Contact Stresses and Their Effects on Instability Rutting of Asphalt Mixture Pavements: Three-Dimensional Finite Element Analysis." *Transportation Research Record*(1853), 150-156.

Panneerselvam, D., and Panoskaltsis, V. P. (2005). "Numerical Implementation of a Hyperelastic-Viscoplastic Damage Model for Asphalt Concrete Materials and Pavements." In: *Asphalt Concrete, Simulation, Modeling, and Experimental Characterization,* , E. Masad, V. P. Panoskaltsis, and L. B. Wang, eds., ASCE, 61-79.

Papagiannakis, A. T., Abbas, A., and Masad, E. (2002). "Micromechanical analysis of viscoelastic properties of asphalt concretes." *Transportation Research Record*(1789), 113-120.

Park, S. W., Kim, Y. R., and Schapery, R. A. (1996). "Viscoelastic continuum damage model and its application to uniaxial behavior of asphalt concrete." *Mechanics of Materials*, 24(4), 241-255.

Park, S. W., and Schapery, R. A. (1997). "Viscoelastic constitutive model for particulate composites with growing damage." *International Journal of Solids and Structures*, 34(8), 931-947.

Sadd, M. H., Dai, Q., and Parameswaran, V. (2004a). "Microstructural simulation of asphalt materials: Modeling and experimental studies." *Journal of Materials in Civil Engineering*, 16(2), 107-115.

Sadd, M. H., Dai, Q., Parameswaran, V., and Shukla, A. (2004b). "Microstructural Simulation of Asphalt Materials: Modeling and Experimental Studies." *Journal of Materials in Civil Engineering*, 16(2), 107-115.

Sadd, M. H., Dai, Q., Parameswaran, V., and Shukla, A. (2004c). "Simulation of Asphalt Materials Using Finite Element Micromechanical Model with Damage Mechanics." *Transportation Research Record*(1832), 86-95.

Sadd, M. H., Qiu, L., Boardman, W. G., and Shukla, A. (1992). "Modelling wave propagation in granular media using elastic networks." *International Journal of*

Rock Mechanics and Mining Sciences & Geomechanics Abstracts, 29(2), 161-170.

Schapery, R. A. (1969). "On the characterization of nonlinear viscoelastic materials." *Polymer Engineering and Science*, 9(4), 295-310.

Schapery, R. A. (1999). "Nonlinear viscoelastic and viscoplastic constitutive equations with growing damage." *International Journal of Fracture*, 97(1-4), 33-66.

Schapery, R. A. (2002). "Homogenized constitutive equations for linear viscoelastic unidirectional composites with growing transverse cracks." *Mechanics Time-Dependent Materials*, 6(2), 101-131.

Simo, J. C., and Ju, J. W. (1987). "Strain- and Stress-Based Continuum Damage Models - I. Formulation." *International Journal of Solids and Structures*, 23(7), 821-840.

Wang, L., Paul, H., Harman, T., and D'Angelo, J. (2004). "Characterization of Aggregates and Asphalt Concrete Using X-ray Computerized Tomography." *J. Assoc. Asphalt Paving Tech.*, 73(467-500).

Yoo, P. J., Al-Qadi, I. L., Elseifi, M. A., and Janajreh, I. (2006). "Flexible pavement responses to different loading amplitudes considering layer interface condition and lateral shear forces." *International Journal of Pavement Engineering*, 7(1), 73-86.

Zhu, H. (2001). "A contact mechanism based theory of Maxwell matrix composites." *International Journal of Solids and Structures*, 38(26-27), 4477-4488.

Analysis of strain in asphalt pavement using FRP-OFBG sensors

CHEN, Fengchen[1]，TAN, Yiqiu[2]，LIU, Hao[3] and WANG, Baoxin[4]

[1]School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin, China; PH (86) 451-82291543; FAX (86) 451-86282118; email: cfc_0@sina.com

[2]School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin, China; PH (86) 451-86283090; FAX (86) 451-86282118; email: yiqiutan@163.com

[3]School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin, China; PH (86) 451-86283090; FAX (86) 451-86282118; email: liuhao19696@126.com

[4]Luxing Material Center of Beijing Road and Bridge Company, Beijing, China; PH (86) 10-60281436; FAX (86) 10-60281416; email: mantou139@163.com

Abstract：FRP -OFBG is a smart sensor in civil engineering due to its advantages of corrosion resistance, high strength, fatigue resistance, electric magnetic resistance, small size,, etc. The experiment of FRP-OFBG sensors was conducted to test and analyze the structure strain of asphalt pavement in a Beijing freeway. The result shows that vehicle loads do more damage to the middle of the surface course. Therefore, FRP-OFBG sensors have stable characteristics and could be widely used to measure structure strain of asphalt pavement.

Key words：*OFBG, FRP, asphalt pavement, strain*

1.Introduction

The total kilometrage of freeways in China has been over 30,000 km by 2005, and about 95% of it is asphalt pavement. But with the rapid development of asphalt pavement, exposed technical problems are very serious. Premature damage of asphalt pavement has attracted more and more road researchers. In order to effectively combat this grim situation, it is necessary to analyze the initial mechanism and development law of premature damage.

At present, due to the lack in detection and trial technology, the real mechanics situation under vehicle loads is not explicit because of a lack of understanding the initial mechanism of damage. The appearance of an OFBG sensor makes internal mechanics monitoring in asphalt pavement structure come true, especially stress-strain parameters. In this paper, analysis of strain in asphalt pavement is accomplished by using FRP-OFBG sensors.

2. Basic Characteristics of FRP-OFBG Sensors

FRP (Fiber Reinforced Polymer) as a popular engineering material has been applied more and more due to its advantages of corrosion resistance, high strength, nonmagnetism, fatigue resistance, and so on. Due to its advantages of electric

magnetic resistance, small size, distributing measurement, durability, etc., OFBG (Optical Fiber Bragg Grating) is now widely accepted as a smart sensor in many fields, such as civil engineering, aerospace engineering, ship engineering, power industry, petroleum and chemical industries, etc. FRP-OFBG is a type of OFBG sensor which is encapsulated with FRP; it contains integrated advantages of both.

According to the uniqueness of asphalt pavement structure, HIT (Harbin Institute of Technology) developed FRP-OFBG sensors which can measure three-dimensional strains in the same location of a road's internal structure , as shown in Figure 1.

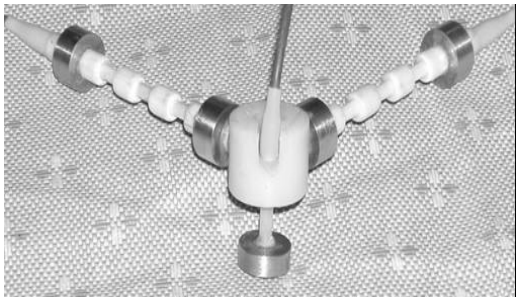


Figure 1. Three-dimensional FRP-OFBG sensors

3. Field Trials of FRP-OFBG Sensors

In August 2006, three-dimensional FRP-OFBG sensors were embedded in a built freeway to monitor the internal strain variation of pavement structure under real traffic loads in Beijing.

Two sensors are united as one group throughout the test to test the strain in the middle and bottom of the surface course. The test selected eight sets of three-dimensional FRP-OFBG sensors to ensure that one group can have a full survival; the specific location can be seen in Figure 2. Such a layout avoids an inadequate compaction degree caused by sensors. The numbering is as follows: the number of sensors for testing the middle of surface course is NO.A, for testing the vertical strain is A-1, for testing the transverse strain is A-2 and for testing the longitudinal strain is A-3; similarly, the sensors for testing the bottom of the surface course are NO.B.

Since the sensors were embedded in the freeway which was built 3 years ago, it was necessary to dig out some areas for construction operation. The sensors require a larger area to reduce the influences of the digging border, and the larger area will bring more damage to the road. The area should be reasonably determined in the actual situation. The determined area is 80*80cm2, 15cm in depth. The detailed process of the construction is shown in Figure 3.

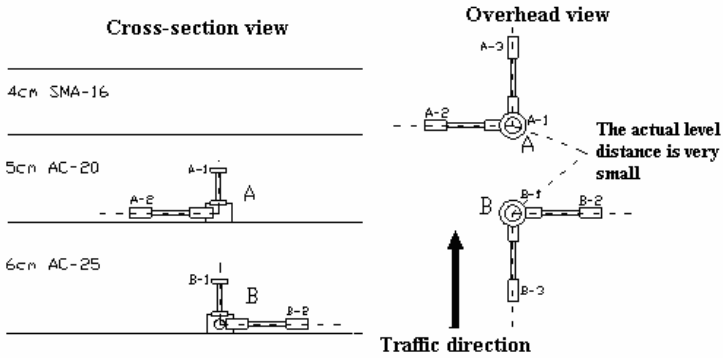


Figure 2. Field location of FBG

Figure 3. Process of FBG's embedding

4. Measure and Analysis of Structure Strain

4.1 Testing program

（1）Measure the axle-load of vehicle, calculate the tire ground pressure;

（2）When the tire ground pressure is the same and the speed is different, measure the strain of the pavement structure and analyze the speed's effect on the strain;

（3）When the speed is the same and the tire ground pressure is different, measure the strain of the pavement structure and analyze the effect on the strain of the tire ground pressure.

4.2 Testing of tire ground pressure

In order to simulate the actual effects of different loads, the same trucks loaded with different weights were regarded as overweight trucks and standard trucks for the test. Calculate the tire ground pressure by using carbon paper; the test results are shown in Table 1.

Table 1. Value of tire grounding pressure

Vehicle type	Vehicle load (Kg)	Axle-load (Kg)	Tire ground area (cm^2)	Tire ground pressure (MPa)
Standard truck	29920	6500	977.08	0.65
		11700	1979.17	0.58
		11720	1947.92	0.59
Overweight truck	49660	6760	1354.16	0.49
		21700	2195.83	0.97
		21200	2656.25	0.78

4.3 Analysis of strain

4.3.1 Same tire ground pressure and different speed

This paper takes overweight trucks as an example to introduce the trend of strain change, and the standard truck has the same regularity. Overweight trucks run over the measured point at the different speeds as17km/h, 30km/h, 44km/h, 56km/hand 69km/h, and the sensors record the strain. Right after the corresponding data processing, it can be found as follows:

（1）With the increase of speed, the strain of every sensor shows a declining trend. This illuminates that the damage caused by vehicle loads decreases along with the increase of speed, and it is consistent with theoretical results. Detailed results are in Figure 4;

（2）Strain in different depths of the surface course varies with the speed. Horizontal strain in the middle and bottom of the surface course is more stable with different speedsof change, but the vertical strain has an obvious variety. At the heavy-load level (referring to the second and third axles) and high speed, the value of vertical strain is closer; with the increasing of speed, strain in the middle of surface course increases much faster than at the bottom. That indicates that the middle of a surface course undertakes the main vertical strain at low speed, and an overweight truck will do greater damage to it. Figure 5 is the strain graph of an overweight truck at the speed of 17km/h and 69km/h.

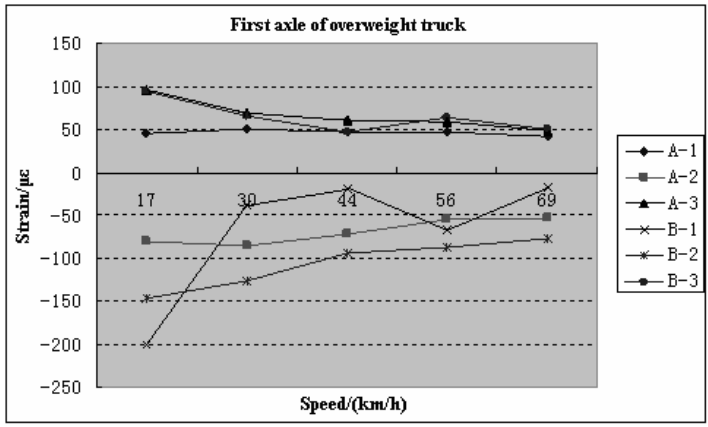
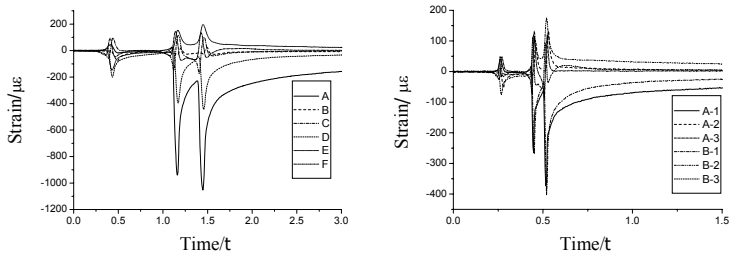


Figure 4. Strain graph of different speeds



17km/h 69km/h

Figure 5. Strain graph of an overweight truck at different speeds

4.3.2 Same speed and different tire ground pressure

Testing was carried out by using the standard truck and overweight truck for strain measurement at the same speed. High-speed and low-speed results were selected as an example to investigate the strain in asphalt pavement; the data are listed in Table 2.

Table 2. Strain value of different grounding pressure

Speed	Tire ground pressure		No. of sensors					
			A-1	A-2	A-3	B-1	B-2	B-3
17	Overweight truck	1th axle 0.49	44.17	-80.00	95.83	-200.83	-146.67	93.33
		2th axle 0.97	-1053.3	25.83	100	-444.17	197.50	135.00
		3th axle 0.78	-940.00	19.17	112.5	-395.83	153.33	145.83
16	Standard truck	1th axle 0.65	-911.67	267.50	75	-1175.00	185.83	186.67
		2th axle 0.58	-649.17	-135.00	100.00	-339.17	-111.67	128.33
		3th axle 0.59	-271.67	-108.33	100.00	-388.33	-175.83	115.00
69	Overweight truck	1th axle 0.49	41.67	-53.33	48.33	-17.50	-76.67	50.00
		2th axle 0.97	-373.33	133.33	120.00	-402.50	175.00	91.67
		3th axle 0.78	-264.17	93.33	115.83	-268.33	132.50	113.33
72	Standard truck	1th axle 0.65	-139.17	65.00	106.67	-401.67	-42.50	96.67
		2th axle 0.58	-345.83	-39.17	81.67	-95.00	73.33	82.50
		3th axle 0.59	-472.50	-68.33	64.17	-63.33	42.50	76.67

After corresponding data processing, it shows as follow:

（1）With the increase of tire ground pressure, the strain of sensors increases too. However, it is difficult to guarantee that the two trucks run at the same speed and act on the same measuring point. There are certainly some inevitable errors. Based on the data measured, it is difficult to get a numerical conclusion. This paper tries to classify the type of strain to analyze the effects of different grounding pressures on asphalt pavement structure. According to Table 2, the type of strain is analyzed whose results are shown in Table 3.

Table 3. Tire analyses of strain

Tire ground pressure（MPa）		No. of sensors					
		A-1	A-2	A-3	B-1	B-2	B-3
Six axles of two trucks at the same speed	0.49	T	C	T	C	C	T
	0.58	C	C	T	C	C	T
	0.59	C	C	T	C	C	T
	0.65	C	T	T	C	T	T
	0.78	C	T	T	C	T	T
	0.97	C	T	T	C	T	T

Remark: C stands for compression strain; T stands for tensile strain.

It can be seen from table 3 that in addition to the top of surface course where tensile strain exists at lower ground pressure, the rest of vertical strain is compression strain. All the longitudinal direction is tensile strain, and along with the enhancement of tire ground pressure, it has a certain increase, but not so marked. Horizontal strain is

sensitive to the change of tire ground pressure, from compression strain at low pressure change to tensile strain at high pressure, and the transformation of the type of strain is among 0.59MPa to 0.64MPa. Therefore, the value of tire ground pressure has an important impact on the type of strain.

（2）Strain in different depths of surface course varies with the tire ground pressure, especially the vertical strain. The vertical strain at the bottom of the surface course is larger than in the middle at low pressure; but along with the enhancement of tire ground pressure, the vertical strain in the middle of the surface course brought a gradual increment and exceeded the value at the bottom. The results illuminate that the middle of the surface course would undertake more strain along with the increase of ground pressure. This is consistent with effects at low speed and also matches some conclusions of theoretical research.

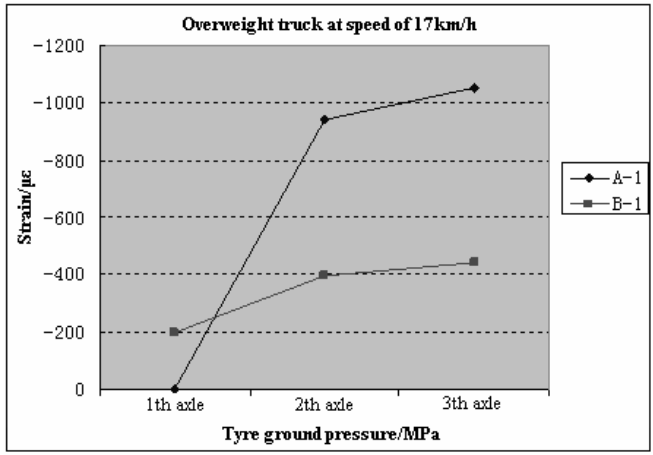


Figure 6. Strain graph of different grounding pressures

5. Conclusion

This paper has done some tentative research on the application of FRP-OFBG sensors to analyze the strain in asphalt pavement structure. Test results indicate that the strain, especially vertical strain in the middle of surface course, will increase a great deal along with enhancing the ground pressure and decreasing speed. The conclusion of the research is consistent with mechanics analysis, which confirms that the application of FRP-OFBG sensors is practicable in research for strain measuring in asphalt pavement. FRP-OFBG technology would have a vast application prospect in road engineering.

References

Alavie A.T., et al. Application and characterization of intracore grating sensors in CFRP prestressed concrete girder. SPIE, Vol.2191, 1994.

Frank A., Nellen M.. Fiber optical Bragg Grating sensors embedded in GFRP Rockbolts. SPIE, Vol.3670, 1999:497-504.

Habel W. R.. Fiber optic sensors for deformation measurements: criteria and method to put them to the best possible use. SPIE, Vol.5384, 2004:158-169.

Meltz G., Overview of Fiber Grating Based Sensors. SPIE, Vol. 2838, 2000.

Morey W. W., Meltz G., Glenn W.H.. Fibre optic Bragg grating sensors. SPIE, Vol.1169,1989.

Nellen, P.M., Rolg Bronnimann, Fank A.. Structurally embedded fiber bragg gratings: Civil engineering applications. SPIE, Vol.3860, 1999.

Ou Jinping, Zhou Zhi. Techniques of optical fiber Bragg grating smart sensors and intelligent monitoring systems of infrastructures. The first international workshop on advanced smart materials and smart structures technology. Havaii ，USA，June 23-23, 2003.

Ou Jinping, Zhou Zhi, Wang Bo. Smart FRP-OFBG bars and their application in reinforced concrete beams. Chinese High Technology Letters. 2005(4): 22-28.

Rao Y.J.. Recent progress in applications of in fiber Bragg grating sensors. Optics and lasers in Eng. Vol.31, 1999.

This page intentionally left blank

Subject Index

Page number refers to the first page of paper

Aging, 752
Air pollution, 825
Air transportation, 134
Airports, 134, 833
Algorithms, 52, 476, 512, 542
Asphalt pavements, 723, 833, 841, 877
Asphalts, 733, 752, 776, 786, 799, 867

Bicycles, 18, 433
Buses, 254, 605

Canada, 159, 192
China, 1, 18, 43, 58, 67, 76, 100, 111, 121, 134, 159, 170, 213, 222, 243, 433, 492, 500, 520, 566, 575, 616, 637, 673, 723, 760, 776, 811
Communication, 371
Commute, 90
Computation, 235
Computer programming, 419, 445
Computer software, 30, 596
Containers, 170
Coupling, 371

Damage, 867
Data analysis, 235, 723
Data collection, 328
Databases, 58
Decision support systems, 43
Degradation, 825
Delay time, 702
Driver behavior, 341, 360, 408
Dynamic models, 317

Emissions, 825
Environmental issues, 825
Equilibrium, 90, 583
Estimation, 702
Evaluation, 301, 352, 445, 712, 786

Fatigue, 341, 851
Finite element method, 867
Forecasting, 30, 147
Freight, 276
Freight transportation, 170, 530
Fuzzy sets, 147, 184

Geographic information systems, 7, 58
Germany, 520
Graphic methods, 352

Harbors, 170
Highways, 67, 276, 301, 308, 381, 465, 760
Human factors, 360

Imaging techniques, 867
Information management, 43, 288, 317, 637, 673
Information systems, 408
Information technology (IT), 637
Internet, 268
Intersections, 301, 308, 389, 520, 549, 575, 683, 702

Japan, 328, 520

Laboratory tests, 752, 786, 851
Land usage, 159
Loads, 811

Magnetic levitation trains, 222
Measurement, 451
Mixtures, 733, 752, 786, 851, 867

Models, 760
Monte Carlo method, 294

Optimization, 254, 328, 500, 583, 605, 660

Parking facilities, 673
Passengers, 67, 476, 596
Pavement design, 833, 841, 851
Pavements, 712, 811
Pedestrians, 294, 512
Performance characteristics, 451, 712, 733
Planning, 30
Polymers, 799
Predictions, 192, 465, 649, 723
Pricing, 213, 268, 276
Probability, 627
Production management, 799
Public transportation, 352, 492

Rail transportation, 401, 476, 530, 660
Railroad stations, 660
Railroad trains, 401
Research, 76, 203, 235, 243, 294, 360, 371, 381, 389, 401, 596
Road construction, 743
Road design, 743
Roads, 451, 733
Routes, 627, 660
Rubber, 776, 786

Sand, 858
Sensors, 877
Simulation, 1, 90, 294, 317, 360, 419, 512, 596, 649, 683
Soil compaction, 858
Spacing, 760
Statistics, 723
Strain, 877
Streets, 76
Subgrades, 858
Surveys, 111, 408
Switzerland, 549

Technology, 235
Temperature, 723
Tensile strength, 799
Terminal facilities, 67
Tests, 858
Tires, 776
Tolls, 695
Traffic accidents, 294
Traffic assignment, 52, 542, 660
Traffic congestion, 7, 500, 627
Traffic flow, 76
Traffic management, 18, 147, 184, 192,203, 235, 243, 288, 317,328,408, 433, 445, 512, 616
Traffic models, 30, 203, 616
Traffic safety, 301, 308, 341, 360, 381, 389, 465, 649, 695, 760
Traffic signals, 76, 389, 419, 520, 549, 575, 583, 605, 616, 683
Traffic signs, 288, 465
Traffic speed, 76, 465
Traffic surveillance, 1
Traffic volume, 743
Transportation management, 1, 30, 43, 52, 58, 111, 121, 159, 213, 254, 268, 371, 500, 637
Transportation models, 276
Transportation networks, 52, 90, 134, 243, 288, 308, 566
Travel demand, 30
Travel time, 451
Trip distribution, 100, 111
Tunnels, 381

Uncertainty principles, 147
Urban areas, 1, 52, 67, 76, 100, 213, 243, 254, 352, 492, 566
Urban development, 121
Utilities, 627

Vehicles, 294, 433

Water transportation, 492

Author Index

Page number refers to the first page of paper

Abdel-Aty, Mohamed, 360, 389, 683
Abraham, John Edward, 159
Adhikari, Sanjeev, 841
Al-Hadidy, Ai, 799

Bai, Hai-Ying, 825
Bai, Qifeng, 786, 851
Bai, Yu, 605, 637
Bai, Zhu, 408
Bar-Gera, Hillel, 52
Bin, Shuai, 445
Bin, Zhuang, 352
Bressette, Terrie, 776

Cai, Guoqiang, 371
Cao, Rongji, 786
Chang, Chaofan, 213
Chang, Edmond Chin-Ping, 616
Chen, Fengchen, 877
Chen, Shaojun, 288
Chen, Shuai, 530
Chen, Shuangyan, 67
Chen, Tao, 170
Chen, Tuansheng, 213
Chen, Xiaohong, 1, 18, 43, 389, 433, 683
Chen, Yuren, 288
Chen, Zhangxi, 445
Cheng, Qi, 492
Chiao, Kuo-Ann, 121

Dai, Leilei, 408
Dai, Qingli, 867
Deng, Wei, 243
Deng, Yi, 222
Ding, Jiping, 695
Dong, Hong-Li, 760

Du, Meng, 184
Du, Xiaoli, 465
Du, Yu-Chuan, 566
Du, Zhigang, 381

Fang, Fang Clara, 419
Fang, Shouen, 294
Feng, Zhongliang, 752

Gao, Jian-Zhi, 100
Gao, Mourong, 184
Gao, Peng, 512
Gao, Ziyou, 52
Goh, Shuwei, 841
Gu, Wang, 203
Guan, Juxiang, 147, 760
Guan, Qiang, 825
Guan, Zhichao, 1
Guo, Xiucheng, 67, 203
Guo, Zhongyin, 465

Han, Baoming, 476
Harb, Rami, 360
Hicks, R. Gary, 776
Hong, Wangjiang, 445
Huang, Hainan, 67
Huang, Zhaoyi, 18
Huang, Zhongxiang, 500
Huber, Gerald A., 811
Hunt, John Douglas, 159

Jia, Lu, 723
Jiang, Guiyan, 408
Jiang, Jinyong, 575
Jie, Yang, 203
Jun, Fan, 67

Kondo, Akiko, 492

Kondo, Akio, 492

Lei, Deming, 308
Li, Gao, 222
Li, Haiyuan, 243
Li, Xi, 371
Lin, Hangfei, 1, 43
Lin, Qun, 1
Ling, Jianming, 833
Liu, Hao, 877
Liu, Kai, 328
Liu, Qiang, 134
Liu, Rongfang (Rachel), 222
Liu, Wusheng, 111
Liu, Yan-Hu, 542
Liu, Zhaobin, 192
Lomax, T., 7
Lu, Huapu, 134
Lu, Jian, 276, 301, 695
Luk, James Y. K., 317, 451
Luo, Dayong, 500

Ma He-Ling, 100
Ma, Wanjing, 605
Mao, Baohua, 213
Mao, Zhe, 341
Mei, Tong, 235
Mo, Yikui, 673
Mohammad, Louay N., 712
Morikawa, Taka, 328

Nakamura, Hideki, 520

Pan, Xiaodong, 381
Peng, Wuxiong, 147, 760

Qiao, Rui-Jun, 542

Radwan, Essam, 360, 683
Ren, Yingwei, 276

Santos, Cristina Dos, 360
Shao, Chun, 702
Shao, Minhua, 76
Sharma, Satish, 192
Shimada, Hideaki, 492
Song, Dandan, 858
Stonex, Anne, 776
Sun, Lijun, 76, 723
Sun, Qinmei, 760
Sun, Youwang, 530
Sun, Li-Jun, 566

Tan, Jaubert Jianan, 549
Tan, Yi-Qiu, 752, 799, 877
Tang, Keshuang, 520
Tang, Shoupeng, 627
Teng, Jing, 476
Timm, David H., 811

van Zuylen, Henk J., 649

Wang, Baoxin, 877
Wang, Jia-ni, 752
Wang, Jinkai, 268
Wang, Junhua, 294
Wang, Xuesong, 389, 683
Wang, YueRu, 288
Wang, Zhan-Yu, 825
Wang, Zhengwu, 500
Wen, Ya, 111
Wong, Ka-Fai, 30
Wong, Winggun, 476
Wu, Chao-Zhong, 341
Wu, Ning, 583
Wu, Shie-Shin, 743
Wu, Shourong, 276
Wu, Zhizhou, 235
Wu, Zhong, 712

Xiang, Jinsong, 401
Xiang, Qiaojun, 276, 301
Xiong, Wen, 18, 433
Xu, Huining, 752
Xu, Meng, 52
Xu, Ruihua, 512, 660
Xu, Xingfang, 401
Xu, Yuan, 147, 760

Xu, Zhi-hong, 733
Xue, Kun, 637
Xue, Ming, 858

Yamamoto, Toshiyuki, 328
Yan, Kefei, 111, 673
Yan, Xinping, 308, 341
Yan, Xuedong, 360, 683
Yang Zhen, 381
Yang, Chao, 43, 317
Yang, Dongyuan, 1
Yang, Fei, 121
Yang, Jiawen, 58
Yang, Peikun, 575
Yang, Rui-hua, 733
Yang, Xiaoguang, 352
Yang, Xiongbin, 596
Yang, Yongshun, 811
Yang, Zhiqing, 465
Ye, Fan, 695
Yi, Ping, 702
Yin, Zhihong, 530
You, Zhanping, 841, 867
Yu, Chaowei, 111
Yu, Yonghua, 723
Yuan, Li, 301
Yuan, Wen-Ping, 566
Yun, Meiping, 627

Zhang, Baoyu, 673
Zhang, Chao, 352, 733
Zhang, Guoqiang, 301
Zhang, H. M., 90
Zhang, Jun, 858
Zhang, Ming, 7
Zhang, Rong, 170
Zhang, Xiaoning, 76, 90
Zhang, Xishu, 811
Zhang, Zhi-Xiang, 851
Zhang, Zhongjie, 712
Zhao, Fang, 254
Zhao, Hongduo, 833
Zhao, Jianyou, 184
Zhao, Jiaqi, 408
Zhao, Yao, 100
Zheng, Zuduo, 408
Zhong, Ming, 159
Zhou, Haiping, 776
Zhou, Lixin, 596
Zhou, Xi, 121
Zhu, Shunying, 147, 760
Zhu, Xiaoning, 371
Zhu, Xiao-Ning, 542
Zi, Haibo, 67
Zou, Bo, 134
Zou, Xiaolei, 512, 660
Zou, Zhi-Yun, 100, 841