Cher Ming Tan · Thong Ngee Goh *Editors*

Theory and Practice of Quality and Reliability Engineering in Asia Industry

Springer

Theory and Practice of Quality and Reliability Engineering in Asia Industry

Cher Ming Tan · Thong Ngee Goh
Editors

Theory and Practice of Quality and Reliability Engineering in Asia Industry

Springer

Editors
Cher Ming Tan
Center for Reliability Science
and Technologies
Chang Gung University
Taoyuan
Taiwan

Thong Ngee Goh
Department of Industrial and Systems
Engineering
National University of Singapore
Singapore
Singapore

ISBN 978-981-10-3288-2 ISBN 978-981-10-3290-5 (eBook)
DOI 10.1007/978-981-10-3290-5

Library of Congress Control Number: 2016958465

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #22-06/08 Gateway East, Singapore 189721, Singapore

Preface

The increasing instances of product recalls globally, with the high-profile cases of the Tanaka airbags in cars and Samsung Note 7 smartphones, signify the importance and necessity of measures for Quality and Reliability in industries. This is especially so in Asia as more and more manufacturing activities are taking place in this part of the world, with a corresponding increase in research and development initiatives.

It cannot be overemphasized that Quality and Reliability Management, which focuses on organization, planning and behavioral matters, with an orientation towards analytical tools as well as technical performance, is dynamic in nature as customer demand and market conditions evolve. The teaching of past "Quality Gurus" like Juran and Deming provides a bearing for business leaders and operations personnel, but as time passes, new thinking, techniques and analytical procedures will continue to develop to meet fresh needs and requirements as product variety, complexity and application domains increase. Thus globalization, big data, customization, sustainability, instant and wide communications—to name a few major items—are realities which traditional mainstream methodologies of Statistical Process Control and Reliability Testing, for example, have hardly addressed. Indeed professionals everywhere are constantly faced with expected or unexpected emerging needs and requirements. Thus there is an incessant stream of new interests and issues faced by researchers and practitioners, making Quality and Reliability studies such fascinating and important areas of endeavor within both academia and industrial sectors worldwide.

This book is contributed by researchers and practitioners of Quality and Reliability Management and Engineering in the recent past. The papers are based on the authors' presentations at the 2014 ANQ (Asian Quality Network) Congress in Singapore; they have been selected based on their relevance to the practice of Quality and Reliability, as well as the directions they have pointed out in the profession's development. The compilation is not meant to be a textbook on Quality and Reliability, for which there are already many titles in the market, but is a snapshot of areas within Quality and Reliability that are of current interest to professionals in this field, and also—as explicitly described in some of the papers—society in general. The contents of this book would constitute a very useful supplement to users of traditional textbooks on Quality and Reliability as there are many actual cases and examples reflecting the latest applications in areas such as services and the environment.

We would like to thank the ANQ Governing Board for its approval of and support for the publication of this book, and in particular to Mr Kenneth Liang, President of the Singapore Quality Institute (SQI), who had mobilized resources for the publication's successful preparation. We would also like to take this opportunity to thank all those who have in one way or another contributed to the compilation of the paper in this volume. In particular, we would like to mention Dr Ho Siong Lin, Mr Teo Lip Hong, Mr Ellson Boey, Ms Sharon Tay and Mr Kitson Lee who helped with the selection and editing of papers in Chapters 3, 4, 5, 7 and 8 respectively. Last but not least, we would like to thank Mr Too Meng Ken, SQI publication chairman, and all the reviewers who have helped in ensuring the high quality of the papers included in this book. Thanks are also extended to Mr. Vivek Sagawan, Research Assistant of Professor Tan, who assisted in liaising with Springer to ensure the accuracy of the contents of these papers.

Taoyuan, Taiwan Cher Ming Tan
Singapore, Singapore Thong Ngee Goh

Contents

Part VIII Reliability Practices

Contributors

Grittapat Akkaphuriwong Office of Strategy Management and Budgeting, Chulalongkorn University, Pathumwan, Bangkok, Thailand

Charles Aubrey Asia Pacific Quality Organization, Grand Junction, MI, USA; International Academy for Quality, Grand Junction, MI, USA; Aubrey Partners, Grand Junction, MI, USA; Anderson Packaging, Grand Junction, MI, USA

Piao Chen Department of Industrial and Systems Engineering (ISE), National University of Singapore (NUS), Singapore, Singapore

Chuen-Sheng Cheng Department of Industrial Engineering and Management, Yuan-Ze University, Chung-Li, Taoyuan, Taiwan, ROC

K.S. Chin City University of Hong Kong, Kowloon, Hong Kong

Francis Chua Master in Education, Coleman College, Singapore, Singapore

Thong Ngee Goh Department of Industrial and Systems Engineering, National University of Singapore, Singapore, Singapore

Takeyuki Goto Waseda University, Okubo, Shinjuku, Tokyo, Japan

Guanghua Han Department of Industrial and Systems Engineering, National University of Singapore, Singapore, Singapore

A.K.M.F. Hoque Physics Department, Tejgaon College, Farmgate, Dhaka, Bangladesh

M.S. Hossain Physics Department, National University, Gazipur, Bangladesh

PoTsang B. Huang Department of Industrial and Systems Engineering, Chung Yuan Christian University (Taiwan), Chung Li, Taiwan

Lu Jin Department of Informatics, University of Electro-Communications, Chofu, Tokyo, Japan

Ong Chee Bin Johnson AUN-QA Expert, Education Quality International, Singapore, Singapore

Uk Jung School of Business, Dongguk University, Seoul, Korea

Chisato Kajihara Waseda University, Okubo, Shinjuku, Tokyo, Japan

Wibool Khongdee Office of Strategy Management and Budgeting, Chulalongkorn University, Pathumwan, Bangkok, Thailand

Gye-Soo Kim Department of Business Administration, Semyung University, Jechon, Chungbuk, Korea

Chiharu Kumazaki Department of Informatics, University of Electro-Communications, Chofu, Tokyo, Japan; Department of Power Train Experiment, Hino Motors, Ltd., Hino, Tokyo, Japan

Sam Lam Master in Business Administration, Singapore Quality Institute, Singapore, Singapore

Song Lan Department of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, Singapore

Hung-Ting Lee Department of Industrial Engineering and Management, Yuan-Ze University, Chung-Li, Taoyuan, Taiwan, ROC

Feng Leng School of Electrical Electronics Engineering, Nanyang Technological University, Singapore, Singapore; TUM CREATE PTE LTD, Singapore, Singapore

Jing Li Mechanical Engineering Department, National University of Singapore, Singapore, Singapore

M.Y. Lin City University of Hong Kong, Kowloon, Hong Kong

Kenza Maher School of Materials Science and Engineering and Research Institute at Nanyang (ERIAN), Nanyang Technological University, Singapore, Singapore

Devender Malhotra Wipro IT Services, Bangalore, India

Masahiko Munechika Waseda University, Okubo, Shinjuku, Tokyo, Japan

Vimolvan Pimpan Chulalongkorn University, Pathumwan, Thailand

Thongchai Pitimanuskul Office of Strategy Management and Budgeting, Chulalongkorn University, Pathumwan, Bangkok, Thailand

Masataka Sano Tokyo University of Science, Tokyo, Japan

Kazuyuki Suzuki Department of Informatics, University of Electro-Communications, Chofu, Tokyo, Japan; JSQC, University of Electro-Communications, Chōfu, Japan

Cher Ming Tan Center for Reliability Science and Technologies, Chang Gung University, Taoyuan, Taiwan; Department of Electronic Engineering, Chang Gung University, Taoyuan City, Taiwan

Chin Hon Tan Department of Industrial and Systems Engineering, National University of Singapore, Singapore, Singapore

Sharon Tay Sengkang Health, Singapore, Singapore

Kuo-Chun Tsai Department of Industrial and Systems Engineering, Chung Yuan Christian University (Taiwan), Chung Li, Taiwan

K.L. Tsui City University of Hong Kong, Kowloon, Hong Kong

Robert Wang School of Electrical Electronics Engineering, Nanyang Technological University, Singapore, Singapore

Watalu Yamamoto Department of Informatics, University of Electro-Communications, Chofu, Tokyo, Japan; JSQC, University of Electro-Communications, Chōfu, Japan

Ching-Chow Yang Department of Industrial and Systems Engineering, Chung Yuan Christian University (Taiwan), Chung Li, Taiwan

Rachid Yazami TUM CREATE PTE LTD, Singapore, Singapore; School of Materials Science and Engineering and Research Institute at Nanyang (ERIAN), Nanyang Technological University, Singapore, Singapore

Zhi-Sheng Ye Department of Industrial and Systems Engineering (ISE), National University of Singapore (NUS), Singapore, Singapore

Masahiro Yokoyama JSQC, Polytechnic University, Kodaira, Tokyo, Japan

Masato Yoshida Tokyo University of Science, Tokyo, Japan

M.A. Zaman Physics Department, Jahangirnagar University, Savar, Dhaka, Bangladesh

Min Zhou Industrial and Systems Engineering Department, National University of Singapore, Singapore, Singapore

Quality Engineering Today

T.N. Goh

To many in the Quality profession, the terms Quality Management (QM) and Quality Engineering (QE) tend to be used interchangeably, or at least there has been little official attempt to distinguish them. As recent as 2014, ASQ (American Society for Quality) http://asq.org/glossary/q.html (seen on April 29, 2014) describes QE as "The analysis of a manufacturing system at all stages to maximize the quality of the process itself and the products it produces." (ASQ 2014). The fact is, Quality today is not just about manufacturing (as quality of service has gained considerable attention nowadays) and QE, unlike many traditional engineering disciplines, cannot be effective without management elements—and that is why one needs the wisdom of both Shewhart and Deming: is it a surprise that the P-D-C/S-A cycle for quality improvement is attributed to both Shwehart and Deming, arguably from the engineering and management perspectives, respectively?

While there is no commonly recognized definitive delineation of what QM is meant to be and what QE precisely is, it may be said that the former touches more on conceptual and behavioral matters as well as approaches to accomplishing business objectives, and the latter the application of generic and usually rigorous analytical tools comprising technically justifiable methodologies devoid of ideologies or subjective judgments. It may be noted that the recent years has seen an increased use of the term "engineering" with a connotation of planning and endeavor—as in "social engineering." Whatever the interpretation, the impetus for research and application of good techniques continues unabated, and developing economies of the world such as those in Asia, are where the growth of the Quality field is particularly noticeable.

T.N. Goh (✉)
National University of Singapore, Singapore, Singapore
e-mail: isegohtn@nus.edu.sg; tngoh2008@gmail.com

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_1

Generally, professional and academic conferences are the most common platform for noteworthy quality techniques to be presented and discussed. Thus, this volume contains selected papers from the 2014 Congress of the Asian Network for Quality (ANQ) that serve to reflect a wide variety of approaches to quality improvement. Readers, regardless of individual experience and maturity, would catch a glimpse of some latest thoughts and applications in the Quality field. This collection reflects yet another step, albeit coming mostly from Asia, in the ardent, global journey toward quality excellence in all fields of human endeavor.

Author Biography

Goh Thong Ngee (吴桐毅), PE, is Academician of the International Academy for Quality (IAQ), Fellow of the American Society for Quality (ASQ) and Honorary Member and Advisor of the Singapore Quality Institute (SQI). He is a recipient of IEEE Engineering Management Society "Educator of the Year" award in 2005, IAQ Masing Medal in 2006 (for the book "Six Sigma: Advanced Tools for Black Belts and Master Black Belts", Wiley UK), ASQ Statistics Division William G Hunter award in 2007, Asia-Pacific Quality Organization Harrington-Ishikawa Medal in 2010, and ASQ Eugene L Grant Medal in 2012.

The citation for Prof Goh's ASQ Grant Medal reads: "For an outstanding career dedicated to the education and research of concepts and techniques for quality excellence and performance improvement; and for exceptional leadership in promoting the learning and application of effective quality tools throughout the world."

Overview of Reliability Engineering

Cher Ming Tan

Abstract The evolution of reliability since the late 1940 till 2000s is described. This evolution mainly came from United States, and the evolution is from system level reliability and maintenance to components reliability through statistical methods. Exponential distribution was used initially, and it was found not realistic in the 60s. Physics of failure began in the 90s, and reliability evaluation was then shifted from statistical methods to physics of failure. Since 2000, the hybrid physics–statistical approach evolve due to the need to combine both of them in order to evaluate component and system reliability realistically. While reliability is evolving and maturing in USA, the concept of reliability was introduced to Asia only in the late 70s in the form of qualification as US manufacturers were shifting their production to Asia, and they needed to qualify the factories in Asia. Unfortunately, the growth of reliability methodology is very slow in Asia, and still many are using exponential distribution to evaluate reliability. On the other hand, physics of failure approach is growing fast in Asia, due probably to the need to troubleshoot failures manufactured in Asia. This chapter helps the readers to have a border view of reliability and the necessity for its evolution.

Keywords Historical evolution · Statistical approach · Physics of Failure approach · Hybrid approach · Monte Carlo modeling · System reliability · Eponential distribution

Before World War II (WWII), reliability as a word came to mean dependability or repeatability. Such meaning still exists to many in Asia today. The modern use was redefined by the U.S. military in the 1940s and evolved to the present. By the 1940s, reliability and reliability engineering still did not exist.

The current meaning connotes a number of additional attributes that span products, service applications, software packages or human activity. These attributes now

C.M. Tan (✉)
Center for Reliability Science and Technologies,
Chang Gung University, Taoyuan, Taiwan
e-mail: cmtan@mail.cgu.edu.tw

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_2

pervade every aspect of our present day technologically intensive world. Let us follow the journey of reliability engineering from the early days to present, where most of the information described up to year 2000s are from References (McLinn 2010; Denson 1998; Azarkhail and Modarres 2012) unless otherwise referenced.

1 The Late 1940s

The demands of WWII introduced many new electronics products into US military. At the onset of the war, it was discovered that over 50% of the airborne electronics equipment in storage was unable to meet the requirements of the Air Core and Navy. This was because the main electronic devices were still the vacuum tubes, whether it was in radar systems or other electronics. These systems had proved problematic and costly during the war. For shipboard equipment after the war, it was estimated that half of the electronic equipment was down at any given time. Vacuum tubes in sockets were a natural cause of system intermittent problems. They could not afford to have half of their essential equipment non-functional all of the time. The operational and logistics costs would become astronomical if this situation was not soon rectified.

On the other hand, much of the reliability work during this period had to do with testing new materials and fatigue of materials. M.A. Miner published the seminal paper titled “Cumulative Damage in Fatigue” in 1945 in an ASME Journal. B. Epstein published “Statistical Aspects of Fracture Problems” in the Journal of Applied Physics in February 1948. These were obviously insufficient to meet the reliability need on electronics, and IEEE formed the Reliability Society in 1948 with Richard Rollman as the first president. Z.W. Birnbaum founded the Laboratory of Statistical Research at the University of Washington in the same year, and through his long association with the Office of Naval Research, the lab served to strengthen and expand the use of statistics in reliability.

2 The 1950s

A study group was initiated in 1950. This group was called the Advisory Group on the Reliability of Electronic Equipment, AGREE for short. By 1952, an initial report by this group recommended the following three items for the creation of reliable systems:

1. There was a need to develop better components and more consistency from suppliers.
2. The military should establish quality and reliability requirements for component suppliers.

3. Actual field data should be collected on components in order to establish the root causes of problems.

In 1957, a final report was generated by the AGREE committee and it suggested the following:

1. Most vacuum tube radio systems followed a bathtub-type curve.
2. It was easier to develop replaceable electronic modules to quickly restore a failed system, and thus they emphasized modularity of design.
3. Recommend to run formal demonstration tests with statistical confidence for products.
4. Recommend to run longer and harsher environmental tests that included temperature extremes and vibration

The recommendations came to be known as AGREE testing and eventually turned into Military Standard 781. Another item provided by the AGREE report was the classic definition of reliability. The report stated that the definition is "the probability of a product performing a specified function without failure under given conditions for a specified period of time".

In parallel, Rome Air Development Center (RADC) was established in Rome, New York in 1951 to study reliability issues with the Air Force. In 1955, RADC issued "Reliability Factors for Ground Electronic Equipment". This was authored by Joseph Naresky. This decade ended with RCA publishing information in TR1100 on the failure rates of some military components. RADC picked this up and it became the basis for Mil-Std Handbook 217 (MH-217) in 1962.

The applications of the probabilistic notions of reliability around this time were widely represented by the exponential distribution. One of the main driving forces for this popularity was the simplicity of the corresponding reliability functions. Having limited computational resources, the early reliability practitioners were evidently seeking a simple reliability model with a straightforward mathematical representation. A combination of these factors made the exponential distribution the dominant model in early reliability assessments. This simplicity accelerated many improvements in traditional statistical and probabilistic approaches to measuring, predicting and testing of item reliability in the 1950s. On the other hand, Wallodi Weibull published his first paper for the ASME Journal of Applied Mechanics in English. It was titled "A Statistical Distribution Function of Wide Applicability" in 1951 in view of the limitations of the exponential distribution that he understood.

Later on, Weibull produced "Statistical Evaluation of Data from Fatigue and Creep Rupture Tests: Fundamental Concepts and General Methods" as a Wright Air Development Center Report 59-400 for the US military in 1959. Birnbaum also made significant contributions to probabilistic inequalities (i.e. Chebychev), non-parametric statistics, reliability of complex systems, cumulative damage models, competing risk, survival distributions and mortality rates during this decade. By 1956, ASQC was offering papers on reliability as part of their American Quality Congress. The radio engineers, ASME, ASTM and the Journal of Applied Statistics were contributing research papers. The IRE was already holding a conference and publishing proceedings titled "Transaction on Reliability and Quality Control in Electronics". This began in 1954 and continued until this conference merged with an IEEE Reliability conference and became the Reliability and Maintainability Symposium.

Another major report on "Predicting Reliability" in 1957 was that by Robert Lusser of Redstone Arsenal, where he pointed out that 60% of the failures of one Army missile system were due to components. He showed that current methods for obtaining quality and reliability for electronic components were inadequate and that something more was needed.

In 1955, a conference on electrical contacts and connectors was started, emphasizing reliability physics and understanding failure mechanisms. Other conferences began in the 1950s to focus on some of these important reliability topics.

This decade ended with a lot of promise and activities. It was the advent of the reliability engineering discipline, and several different methods were introduced to achieve the goal of higher reliability. Two branches of reliability began to emerge. One branch existed for investigation of failures and the other for predictions.

3 The 1960s

The demands of the military ranging from missiles to airplanes, helicopters and submarine applications drove a variety of technologies. The study of the effects of EMC on systems was initiated at RADC and this produced many developments in the 1960s.

By now, the reliability discipline was working under the tenet that reliability was a quantitative discipline that needed quantitative data sources to support its many statistically based techniques, such as allocation and redundancy modelling. Another branch of the reliability discipline focused on the physical processes by which components were failing. The first symposium devoted to this topic was the "Physics of Failure in Electronics" symposium sponsored by the RADC and IIT Research Institute (IITRI) in 1962. Richard Nelson of RADC produced the document "Quality and Reliability Assurance Procedures for Monolithic Microcircuits," which eventually became Mil-Std 883 and Mil-M 38510. This symposium later became the "International Reliability Physics Symposium (IRPS)". Unfortunately, the electronic industry continued using the lot tolerance percentage defective (LTPD) statistic to address failures of electronics. It was not until late 1980s that PoF was revisited as a serious alternative.

In 1962, G.A. Dodson and B.T. Howard of Bell Labs published "High Stress Aging to Failure of Semiconductor Devices" in the Proceedings of the 7th National Symposium of Reliability and Quality Control. This paper justified the Arrhenius model for semiconductors. Lots of other papers at this conference looked at other components for improvement.

1962 was a key year with the first issue of MH-217 by the Navy. Once issued, MH-217 quickly became the standard by which reliability predictions were performed, and other sources of failure rates gradually disappeared. Part of the reason for the demise of other sources was the fact that MH-217 was often a contractually cited document and defence contractors did not have the option of using other sources of data.

By the 1960s, the exponential distribution turned out to be not so practical for many applications and sensitive to departure from the initial assumptions. The application of this model for components with high reliability targets could result in unrealistic mean-time-to-failure (MTTF). Further, this model basically ignored any aging and degradation in the component and had no memory to keep track of the damage being accumulated in the item.

After such disappointments, reliability practitioners made an attempt to capture some of the physical characteristics of failure into their modelling by using other available traditional distributions, such as the Weibull and lognormal distributions. During this decade, a number of people began to use, and contribute to the growth and development of, the Weibull function, the common use of the Weibull graph, and the propagation of Weibull analysis methods and applications.

In 1969, Birnbaum and Saunders described a life distribution model that could be derived from a physical fatigue process where crack growth causes failure.

Human reliability had now been recognized and studied, which resulted in a paper by Swain on the techniques for human error rate prediction (THERP).

During the decade, a strong commitment of US Government to space exploration resulted in the establishment of NASA, and this provided a driving force for improved reliability of components and systems. The decade ended with a landing on the moon showing how far reliability had progressed in only 10 years.

However, the two branches of reliability engineering seemed to be diverging by the end of the decade, with the *system* engineers devoted to the tasks of specifying, allocating, predicting, and demonstrating reliability, while the physics-of-failure engineers and scientists were devoting their efforts to identifying and modelling the physical causes of failure. Both branches were integral parts of the reliability discipline.

4 The 1970s

In the early 1970s, the responsibility for preparing MH-217 was transferred to RADC, who published revision B in 1974. However, other than the transition to RADC, the 1970s maintained the status quo in reliability prediction. MH-217 was updated to reflect the technology at that time, and there was a shift in the complexity of the models being developed for MH-217.

In fact, there were several efforts to develop new innovative models for reliability prediction. The results of these efforts were extremely complex models that might have been technically sound, but were criticized by the user community as being too complex, too costly, and unrealistic. These models were never incorporated into MH-217.

The 1970s marked the birth of fault tree analysis, which was motivated by the need for safety assessment in the aerospace industry and later for nuclear power plants. Up to this point, most reliability engineering efforts were focused on reliability of components and devices. Nevertheless, there was intense interest in system-level safety, risk and reliability in different applications such as the gas, oil, and chemical industries, and above all, in nuclear power applications. These challenges were immensely appealing to reliability community in the 1970s.

The Navy Material Command brought in Willis Willoughby from NASA to help improve military reliability across a variety of platforms. During the Apollo space program, Willoughby had been responsible for making sure that the spacecraft worked reliably all the way to the moon and back. In coming to the Navy, he was determined to prevent unreliability. He insisted that all contracts contain specifications for reliability and maintainability instead of just performance requirements. Willoughby's efforts were successful because he attacked the basics and worked upon a broad front.

5 The 1980s

The 1980s demonstrated progress in reliability across a number of fronts from military to automotive and telecommunications to biomedical. This was because televisions had become all semiconductor. Automobiles rapidly increased their use of semiconductors with a variety of microcomputers under the hood and in the dash. Large air conditioning systems developed electronic controllers, as had microwave ovens and a variety of other appliances. Communications systems began to adopt electronics to replace older mechanical switching systems. RADC published their first Reliability Tool Kit.

While MH-217 was updated several times, other agencies were developing reliability prediction models unique to their industries. As an example, the automotive industry, under the auspices of the Society of Automotive Engineers (SAE) Reliability Standards Committee, developed a set of models specific to automotive electronics, SAE870050. The SAE committee believed that there were no existing prediction methodologies that applied to the specific quality levels and environments of automotive applications.

The Bellcore reliability prediction standard is another example of a specific industry developing methodologies for their unique conditions and equipment. It originally was developed by modifying MH-217 to reflect better the conditions of interest to the telecommunication industry. It has since taken on its own identity with models derived from telecommunication equipment and is now used widely within that industry.

During this decade, the failure rate of many components dropped by a factor of 10. Software became important to the reliability of systems; this discipline rapidly advanced with work at RADC. Complex software-controlled repairable systems began to use availability as a measure of success. Repairs on the fly or quick repairs to keep a system operating would be acceptable.

The Very High Speed Integrated Circuit (VHSIC) program was the US Government's attempt to leverage from the technological advancements of the commercial industry and at the same time produce circuits capable of meeting the unique requirements of military applications. From the VHSIC program came the Qualified Manufacturers List (QML)—a qualification methodology that qualifies an IC manufacturing line, unlike the traditional qualification of specific parts. The US Government realized that it needed a QML-like process if it were to leverage from the advancements in commercial technologies, and at the same time have a timely and effective qualification scheme for military parts. At this point, reliability concept was introduced to Asia as a form of qualification as some US electronics manufacturers were moving some of their manufacturing sites to some Asia countries such as Singapore and Taiwan since the 1970s.

As technology has advanced, the gate or transistor count became so high that it could no longer effectively be used as the measure of complexity in a reliability model. Furthermore, transistor or gate count data were often difficult or impossible to obtain. Therefore, the model developed for VHSIC microcircuits needed another measure of complexity on which to base the model. The best measures, and the ones most highly statistically correlated to reliability are defect density and the die area applicable to specific IC features, e.g., metalization and oxide. The failure rate (for small cumulative fraction failure) is directly proportional to the product of the area and defect density. Another factor that is also highly statistically correlated with defect density and area is the yield of the die, or the fraction of die that is functional upon manufacture.

However, the problem in using these factors in a model is that they are highly sensitive parameters from a market competition viewpoint and therefore are rarely released by the manufacturers, rendered inaccuracy of the models. The conflict between the usability of a model and its accuracy has always been a difficult compromise to address for model developers.

The traditional approach in developing a life model for such components was to collect as much field failure data as possible to build a statistical model for the component life. Because of the decreasing budget and resources, and also due to faster trends in mass production, great emphasis was placed on capturing the needed information with much less effort. As a result, design and assessment methodologies that addressed the root causes of failure and other operating conditions emerged as powerful cost saving techniques. The accelerated life modelling approach was a direct outcome of this movement.

Accelerated life models took into account some of the operational conditions and were a primary attempt by reliability practitioners to make the life models more flexible. In the first step of this approach a stress agent, which could be an aggregate effect of many physical and operational conditions, was introduced. In the next step this agent was added to the statistical distribution of TTF to form a robust and general life model. Such models had more flexibility, yet needed much less reliability (failure) data.

However, it is usually very complicated (if possible at all) to introduce one or two stress agents to replace the aggregate effect of all influential factors in accelerated life testing approach. Many assumptions and simplifications need to be made, which could unacceptably limit the relevance of the outcomes. Yet, this was not the only challenge, since the accelerated life models, like all other statistical-based approaches, needed data for validation, and data collection meant time and resources that were decisively tight for most start-ups and fast-growing businesses. In addition, there were no data available for a product in the design stage or for a highly reliable product that was hard to break. In such cases, if the modeller was lucky, reliability models could be constructed based on some generic data from the history of similar products. This data could be updated later with expert judgments or other soft data with statistical inference techniques. The uncertainty bounds of such predictions depended upon direct failure data, if available. Therefore, more dependable reliability techniques needed to be developed to address reliability challenges imposed by emerging mass manufacturing technologies.

This latter challenge led to the growing application of Bayesian method in probabilistic data analysis in the 1980s. Using this approach, engineers utilized data available in generic handbooks and from expert opinions and any previous experience with similar products to make a probability density referred to as a prior distribution. The Bayesian framework made it possible to update this prior knowledge later and with just a few available data, an upgraded posterior state of knowledge can be made. The applications of this approach, however, were originally limited to simple reliability models due to the mathematical complexity of involved algorithms. The integrals necessary for normalization at Bayesian conditional probability calculations can be very complex when dealing with multi-parameter reliability models. This remained one of the two most important constraining elements of this approach (the other was developing a proper likelihood function representing reliability data) until recently, when advanced computational tools and techniques became available after revolutionary improvement in the computational power of personal computers.

The 1980s also marked the development of initiatives for modelling dependencies at the system level. Most of these efforts tackled the common cause failures as frequent dependency problems in systems. The common cause failure (CCF), which is the failure of more than one component due to a shared root cause, is classified as a dependent failure. In the 1980s many implicit and explicit methods were developed to incorporate common cause into the system failure analysis (McLinn 2010). In early attempts to model CCF at the system level, a new independent failure event with a specific probability was usually added to the system model. The probability of this event was estimated using field data available on the dependent failures of components.

Contributions by William Meeker, Gerald Hahn, Richard Barlow and Frank Proschan in developing their statistical models for wear, degradation and system reliability made an impact on reliability in this decade and subsequent.

6 The 1990s

By the 1990s, the pace of IC development was picking up. New companies built more specialized circuits. Wider use of stand-alone microcomputers was common and the PC market helped keep IC densities following Moore's Law and doubling about every 18 months. It quickly became clear that high volume commercial components often exceeded the quality and reliability of the small batch specially screened military versions, and the move towards Commercial off the Shelf (COTS) components gained momentum. Many of the military specifications became obsolete and best commercial practices were often adopted. RADC updated their Reliability Tool Kit for COTS applications in this decade.

With the end of the cold war, the military reliability changed quickly. MH-217 ended in 1991 at revision F2. New research developed failure rate models based upon intrinsic defects that replaced some of the complexity-driven failure rates that dominated from the 1960s through the 1980s.

RAC issued a six set Blueprint for Establishing Effective Reliability Programs in 1996. The internet showed that one single software model would not work across the wide range of worldwide applications, which now includes wireless. Also, network availability goals became "five 9 s or 5 min annually" to describe the expected performance in telecommunication, raising the standard of reliability. New approaches were required such as software mirroring, rolling upgrades, hot swapping, self-healing and architecture changes.

With the spread of reliability requirements to commercial world, new reliability training opportunities and books became available to the practitioners. ASQ made a major update to its professional certification exam (i.e. certified reliability engineers certification) to keep pace with the changes evident. ISO 9000 added reliability measures as part of the design and development portion of the certification.

New technologies such as microelectromechanical systems (MEMS), hand-held GPS, and hand-held devices that combined cell phones and computers all represent challenges to maintain reliability. Product development time continued to shorten through this decade and what had been done in three years was now done in 18 months. This meant reliability tools and tasks must be more closely tied to the development process itself. Consumers have become more aware of reliability failures and the cost to them.

In this decade, several vocal critics of MH-217 have complained about its utility as an effective method for assessing reliability. While the faultfinders claim that it is inaccurate and costly, there was no viable replacement in the public domain then.

Table 1 Failure-cause distribution and definition for electronic systems

• **Parts (22%)**: Failure resulting from a part (e.g., microcircuit, transistor, resistor, connector) failing to perform its intended function
• **No defect (20%)**: Perceived failures that cannot be reproduced upon further testing. These may or may not be an actual failure; however, they are removals and therefore count towards the logistic replacement rate (which is often incorrectly referred to as failure rate)
• **Manufacturing (15%)**: Failures resulting from anomalies in the manufacturing process, (e.g., faculty solder joints, inadequate wire routing resulting in chafing, bent connector pins)
• **Induced (12%)**: Failures resulting from an externally applied stress such as electrical overstress and maintenance induced failures (e.g., dropping, bending pins)
• **Design (9%)**: Failures resulting from an inadequate design (e.g., tolerance stack-up, unanticipated logic conditions, a non-robust design for given environmental stresses)
• **Wearout (9%)**: Failures resulting from wearout-related failure mechanisms (e.g., electrolytic capacitors, solder joints, tubes (TWT), and switch and relay contacts)
• **Software (9%)**: Failure of system to perform its intended function due to the manifestation of a software fault
• **System management (4%)**: Failures to interpret system requirements, or failure to provide the resources required to design and build a reliable system

The fraction for each failure cause is given in ( ), and listed from most to least
Data are as of 2012

The premise of traditional methods, such as MH-217, is that the failure rate is primarily determined by components comprising the system. This was a not-unreasonable premise in the 1960s and 1970s when components had higher failure rates, and when systems were less complex than they are today. Also, MH-217 was originally created to support the weapon system acquisition process. As such, it was required that a reliability estimate be available to system program offices (SPO) at certain milestones in the DoD acquisition cycle.

In many cases, an estimate of reliability was required before the parts list was even complete. Model developers have long known that many of the factors which had a major influence on the reliability of the end product were not included in MH-217, but under the constraints of handbook users, no better solution was available.

Increased system complexity and component quality in the 1990s have resulted in a shift of system failure causes away from components to more system-level factors, including manufacturing, design, system requirements, interface, and software. Table 1 shows the failure cause distribution of electronic systems as

Table 2 Digital circuit board failure rates (failures per 10^6 part-hours)

	Ground benign				Ground fixed			
Temperature (°C)	10		70		10		70	
Stress (%)	10	50	10	50	10	50	10	50
ALCATEL	6.59	10.18	13.30	19.89	22.08	29.79	32.51	47.27
Bellcore Issue 4	5.72	7.09	31.64	35.43	8.56	10.63	47.46	53.14
Bellcore Issue 5	8.47	9.25	134.45	137.85	16.94	18.49	268.90	275.70
BrTelecom HDR4	6.72	6.72	6.72	6.72	9.84	9.84	9.84	9.84
BrTelecom HDR5	2.59	2.59	2.59	2.59	2.59	2.59	2.59	2.59
MH-217E Notice 1	10.92	20.20	94.37	111.36	36.38	56.04	128.98	165.91
MH-217F Notice 1	9.32	18.38	20.15	35.40	28.31	48.78	45.44	79.46
MH-217F Notice 2	6.41	9.83	18.31	26.76	24.74	40.15	73.63	119.21
RAC data		3.3						

Br ⇒ British

collected by RAC. Historically, these factors have not been explicitly addressed in prediction methods.

Therefore, a good model should reflects state-of-the-art technology, and it should includes the assessment of processes used in the design and manufacture of the system, including factors contributing to failure causes as follows:

- Parts
- Design
- Manufacturing
- System management
- Induced
- Wearout
- No defect found
- Software

The probability of failure is a complex interaction between the defect density, defect severity, and stresses incurred in operation. Failure rates predicted using empirical models are therefore typical failure rates and represent typical defect rates, design, and use conditions.

Table 2 contains the predicted failure rate of the various empirical methodologies for a digital circuit board. The failure rates in this table were calculated for each combination of environment, temperature and stress. These data show that there can be important differences between the predicted failure rate values, depending on the method used.

Differences are anticipated because each methodology is based on unique assumptions and data. The RAC data in the last row of the table is based on observed component failure rates in a ground benign application.

Furthermore, the empirical models are usually developed from the analysis of field data, which take time to collect. The faster the growth, the more difficult it is to derive an accurate model.

As a result, the 1990s marked the rebirth and widespread development of the PoF approach. While the roots of PoF methods can be found in fracture mechanics, which by itself started through the work of Alan A. Griffith just after World War I and was developed by George R. Irwin in the 1950s, and the PoF approach started through a series of four symposia from 1962 to 1966, but the concept lost its drive until the late 1980s. Enormous advancements in computational tools and faster personal computers on one side and emerging, advanced testing technologies in material science, on the other side, accelerated this trend. In this approach, facts from root cause physical/chemical failure processes are used to prevent the failure of the products by robust design and better manufacturing practices. Because of the competitive environment in production of consumer products and limited budget and resources, great emphasis was placed on capturing the needed information more quickly and with less effort. As a result, design and assessment methodologies that address the root causes of failure have emerged as powerful cost saving techniques.

In the early 1990s, the US Army and Air Force initiated two reliability physics related programs. In 1992, the Army authorized the Electronic Equipment PoF projects to promote a more scientific approach to reliability assessment of electronic equipment.

The PoF approach to reliability utilized scientific knowledge of degradation processes and the load profile applied to an item, its architecture, material properties and environmental conditions to identify potential failure mechanisms that individually or in combination lead to the item's failure. The PoF models, once developed and validated, would be used to estimate life expended and expected. Use of PoF reduced the need for a substantial amount of life data to arrive at a reliability model, since PoF employs the available well-developed knowledge about the process of failure. Such knowledge models how and why the item fails and reduces the need for large quantities of life data.

By the middle of the 1990s, criticism against the application of generic data in general and MH-217, in particular, became increasingly intense, and the basic idea of using failure rate data gathered in such databases was seriously questioned. However, the critics who wanted to abandon the data provided in handbooks for being irrelevant and useless in many applications had difficulties proving that the PoF approach could do any better in reliability predictions. In fact, most of the PoF models strongly depended on life or test data in one way or another. The question was, if there is enough data available to evaluate a PoF-based model, why not use the same data for statistical inference and take the traditional failure rate modelling path again?

To help identify the manner in which reliability predictions are used by reliability practitioners, a survey was issued at the end of the decade, and approximately 60 non-DoD companies responded. From these data, the 3 predominant purposes for performing reliability assessments, in order of importance were reported as:

1. Determining feasibility in achieving a reliability goal or requirement.
2. Aiding in achieving a reliable design, derating, component selection, environmental precautions, input to FMEA and Fault Trees.
3. Predicting warranty costs and maintenance support requirements.

Survey respondents were also asked to identify the methodologies they use when predictions are performed. MH-217 was reported to be by far the most universally applied failure rate prediction methodology, although its updated ceased in 1991.

7 The 2000s

In view of the critics against MH-217 and the need to prove the usefulness of the PoF approaches as mentioned earlier, the reliability community moved towards an integrated use of both approaches. Where the PoF-based approach could save time and money by addressing the root causes of failure and reduce the burden of gathering a substantial amount of data, the traditional statistical failure rates could be useful in probabilistic reliability predictions considering uncertainties involved. However, the uncertainty bounds were often so wide as to make the result almost worthless in decision-making processes. In order to better manage uncertainty and make practical engineering decisions, two factors needed to be considered. The first important element was indeed reliance on more data, for which accelerated life testing, step-stress testing, expert judgment and many different resources were exhausted. The second element, which was considered as important as the first, was an appropriate computational framework that allows new data to be easily added to the analysis. The classical maximum likelihood estimation (MLE) method introduced by Fisher was one of the possible choices.

The MLE method used the likelihood function of the available data for the model parameter estimation. This method provided no means to incorporate prior knowledge available for the model parameters. This method would mathematically collapse when no complete failure data was available. This was almost always the case with new, highly reliable components and systems. The uncertainty bounds provided by the local Fisher information matrix are not useful when dealing with small sample sizes.

In contrast with MLE methods, the Bayesian approach provided many useful features, including a powerful means to incorporate prior knowledge, dealing with the whole distribution of the likelihood function, fair coverage of uncertainties, and finally the possibility of using many different forms of data (exact, censored, fuzzy, partially relevant and expert judgments). Nevertheless, one of the limiting factors of the Bayesian inference methods in practical reliability analysis was the mathematical complexity of the problem.

Multidimensional joint distributions are generally hard to deal with. Later in the 2000s, the numerical and computational advancements in Bayesian statistical methods, such as Markov Chain Monte Carlo (MCMC) simulations and other sampling-based methodologies, combined with advancements in computational tools and development of powerful programming platforms, made the Bayesian inference techniques a common reliable practice.

Employing Bayesian analysis, reliability practitioners combined different types of data, including simple failure rates from traditional handbooks, engineering expert judgments, simulated results of sophisticated PoF models and direct test results, in a hybrid platform. With availability of fast computing, hybrid methodology became widely available and practical. These techniques could rely on the physical and, to a lesser extent, chemical phenomena that drive degradation and failures. Along with small (accelerated) tests and field or expert judgment data, such hybrid models became the source of industry-specific reliability data and analytical models needed to assess the life and safety of highly reliable consumer products and other complex engineering systems in the 2000s.

From the above description of the historical paths of reliability engineering, we can see that the advancement of reliability engineering is driven from US military and the weak reliability of electronics components. Over the decade, reliability engineering has expanded to commercial world outside US and reliability of electronic components are also getting much higher. Today the two primary purposes for performing a quantitative reliability assessment of systems are two folds, namely to

1. assess the capability of the parts and design to operate reliably in a given application;
2. estimate the 'number of field failures' or the 'probability of mission success'.

1945	1945-1950	1952	1953	1960s	1970s	1980s	1990s	2000s
V-1 missile is developed Quantitative measure for reliability is established "Weakest Link" reliability design concept is developed	U.S. Armed Forces Only 30% of the electronic devices are successful in missions Cost of maintenance and repair is 10 times the original cost Dec. 7, 1950 Ad hoc groups are established Reliability engineering for electronics begins	U.S. DOD AGREE (Advisory Group of the Reliability of Electronic Equipment) is formed System RMA is established to meet the government procurement needs RMA requirement specification is standardized	Exponential distribution is popular approach Infant mortalities are screened out to get almost constant failure rate Degradation was not an issue for electronic systems Analyses are simple	Exponential distribution declines in popularity because: It is not practical in many applications It is sensitive to departure from the initial assumptions It leads to high chance of accepting systems with poor mean time to failure Other statistical distributions are used to get more realistic hazard rate	Birth of fault tree analysis is motivated by nuclear safety considerations Monte Carlo methods for finding minimum cut sets are avoided Combinatorial algorithms are developed for analyzing very large fault trees Bayesian approach allows development of new state of knowledge form prior experience	Common cause failure analysis and modeling dependencies are implemented System-level reliability assessment is needed Bayesian statistics become more advanced Accelerated life testing is implemented and facts from real cause of failure is used in statistical models	Physics of failure approach is widespread Facts from root-cause failure processes are used to prevent the failure of the products Robust design approaches are implemented Better manufacturing practices are implemented	The age of hybrid physics-statistics approaches begins Simulations are performed instead or in addition to testing

Fig. 1 History of methods developing in reliability engineering

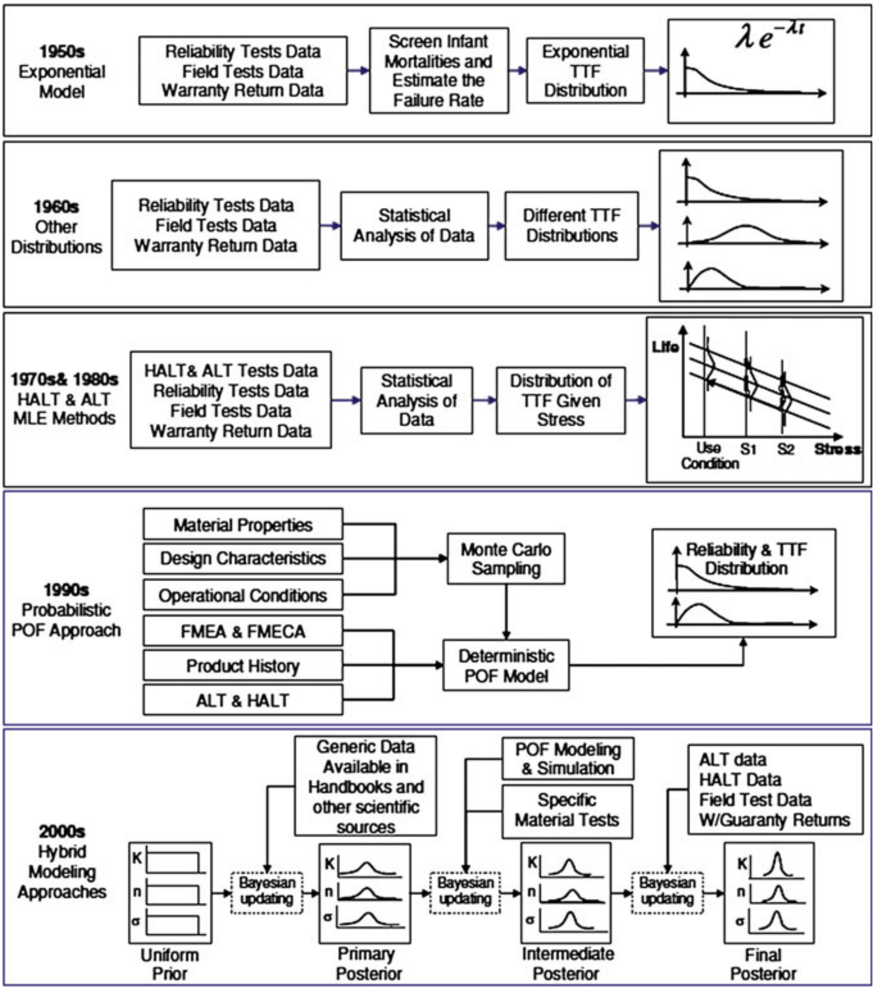


Fig. 2 Emerging PoF modelling approaches in reliability assesments

Purpose #1 does not require statistically based data or models, and physics-of-failure approaches have merit.

Purpose #2, requires empirical data and models derived from those data; this is because field component failures are predominantly caused by component and manufacturing defects which can only be quantified through the statistical analysis of empirical data.

For system acquisition, there is clearly a timeliness element. The prediction must be available early enough to influence the design and/or selection of the design for the system. Yet, the earlier the prediction is needed, the less detailed technical information is available to support the prediction itself. Also, there is a cost/benefit tradeoff that must be considered. Usually, no one is willing to spend more money to define accurately the system reliability than the system itself is worth, and few people are interested in identifying the system reliability after they own it and it is too late to influence its reliability. On the other hand, the emerging of the predictive maintenance in the 2000s used the components' reliability data to predict and schedule future maintenance time is gaining acceptance. Examples of the predictive maintenance can be found in (Tan and Raghavan 2007, 2008; Le and Tan 2013).

Figure 1 summarizes the history of methods in reliability engineering and Fig. 2 illustrates the gradual paradigm shift to applications of PoF modelling methods in reliability assessment of components and systems in the last 50 years.

As shown in Fig. 2, the early reliability models, which were originally developed to address the probability of the occurrence of a failure event, gradually became more complicated to address the real cause of failure. This important change was first realized through statistical representation of the TTF. Applications of other distributions were made by reliability engineers to at least consider some of the real-life characteristics by introducing variable hazard rate models.

The next step was the accelerated life modelling approach, in which the aggregate effect of operational conditions (or stress) was added to the life model to be used as a link between the failure data available in different operational conditions. To introduce a stress agent for an accelerated life model, a complete understanding of the progressing failure mechanisms was needed. The acquisition of this type of knowledge, alongside advancements in materials testing, helped the development of an inclusive library of deterministic PoF models that was later used as a basis for the Monte Carlo-based simulations in probabilistic PoF modelling approaches in the 1990s. A comprehensive description of the progression and status of PoF modelling for integrated circuit interconnections can be found in (Tan 2010).

Those with the view that the deterministic PoF would be sufficient to address reliability and could be used to assure that no failure would occur during the expected life of the item soon realized that there were many uncertainties associated with the PoF models. This forced further testing and data collection and accounting of the associated uncertainties. Generally speaking, the uncertainty of TTF in real data is wider than that predicted by the PoF model. To address this issue, reliability engineers are currently utilizing hybrid approaches within a Bayesian data assessment framework, meaning that a prior TTF distribution is first developed utilizing the PoF model/s; this prior is updated later, using appropriate field or test data to make the final posterior TTF distribution.

The PoF models are deterministic by nature, since they are usually predicting the basic behaviour of the materials in a controlled condition. These models therefore need a separate stochastic process, such as Monte Carlo simulation, to generate the statistical-based probability measures of reliability. Limited computational resources were the other restrictive factor for the further development of PoF modelling approach into the system-level in early stages. However, this constraint was overcome by the overwhelming advancements in personal computers and their truly inexpensive computational power, related operating systems and computational tools. Today, the hybrid physics–statistics approach is the most popular approach for assessing the reliability of components and subsystems. This is mainly because the Bayesian inference data analysis platform allows integration of many different independent sources of knowledge into the reliability assessment. The restrictions in the available system level reliability methods, however, still remain as the main barrier limiting the versatility of the PoF models.

8 Present Status

Since the middle of 2000s, the consumer culture has been dramatically altered from the past few decades. Today, most consumers anticipate shorter product lives, but expect higher dependability of their products. The product is discarded not because it is no longer functional, but rather because it is simply obsolete or because more affordable, newer versions with extra features are available. Consumer tolerance of defective products has also dramatically decreased. Only two decades ago it could take years before consumer reviews could sufficiently damage a product's reputation in the market. Today, however, it can be a matter of only a few days before customer reviews in Internet forums that can isolate a brand name and may lead to financial misfortune for the company. A combination of these factors, plus ever-increasing demand for better after-sale services, has created a very competitive environment and has left a tiny allowable margin for error and negligence.

Today, reliability engineers are asked to develop reliability assessment platforms that are capable of integrating diverse sources of knowledge, from simple field returns to the results of the most sophisticated physics-of-failure and finite-element stress analyses of the component.

From the safety regulation perspective, the diversity of design and alternative safety features, the cost saving requisite of the projects, and environmental concerns have all forced experts to abandon traditional conservative views and adopt new, promising, risk-informed, risk-based and/or performance-based decision-making approaches. This was mostly because even the conservative approaches to design, manufacturing, and operation could not eliminate the possibility of accidents. The risk assessment techniques that were originally developed to support conservatism in design, construction, manufacturing, and operation were increasingly used to develop comprehensive probabilistic risk assessment models to support critical decisions about the alternatives.

Traditional reliability methods and concepts should be revised in order to address the fast-growing demand for highly reliable and quickly evolving engineering systems, structures, and components. Traditional reliability assessment techniques have long been criticized for their shortfalls. The popular constant hazard rate failure model is not practical in many applications and is sensitive to departure from the initial assumptions and has also been obsoleted.

Technologies are evolving at a pace much faster than the time needed to generate enough field data or to perform a large amount of reliability tests economically. This poses challenges to reliability prediction and assessment.

While reliability engineering has been advancing significantly in the US over the decades, its progress in Asia is unfortunately lagging behind. In this last part of the Chapter, let me share my personal experiences in my industrial involvement in reliability activities, as reliability engineer in the 1980s and later as industrial consultant and trainer on reliability engineering in Singapore and Taiwan. Although it is not exhaustive, it is typical in Asia.

9 Reliability Engineering in Asia

Many US electronics companies set up their manufacturing plants in Taiwan and Singapore in the late 1970s. Following their practices to ensure reliability, all these manufacturing plants in Asia had to go through qualification before they were allowed to mass produced. However, design and development were not transferred over to Asia, and thus the reliability activities that these manufacturers seen were Reliability qualification and periodic reliability monitoring.

With the introduction of the concept of QML in the 1980s as mentioned earlier, the suppliers in Singapore and Taiwan were also needed to be qualified, and hence reliability qualification concept was extended to other manufacturers in Asia. As qualification tests are designed with no failure, and the computation of the product reliability was based on exponential distribution in the 1980s, the reliability knowledge for the engineers then were limited to zero failure represents good reliability and the formulae for reliability computation were those derived based on exponential distribution.

As reliability training is seriously lacking in Asia, and the trainers are mostly from industrial where they were involved in reliability in the early days, the reliability knowledge that passed down to the new engineers are again limited to the case of zero failure and exponential distribution, even up to 2000s. This is also partly due to the fact that present days reliability engineering has been evolved to become mathematically intensive, and engineers in industrial tend to avoid mathematics in Asia. Hence, many are still stick to MH-217 for reliability evaluation/prediction, without knowing that MH-217 has no longer been updated after 1990s. In fact, in some cases, companies are using the quality concept such as LTPD and Mil-Std 105 to evaluate product reliability. In a book chapter written by me, I listed 10 misconceptions in reliability that are found to be common in Asia, they are (Tan 2009):

1. High MTTF represents good reliability
2. Zero failure is better than having failures in reliability test.
3. Exponential distribution is all that I need for the analysis of test data
4. All test data are valid

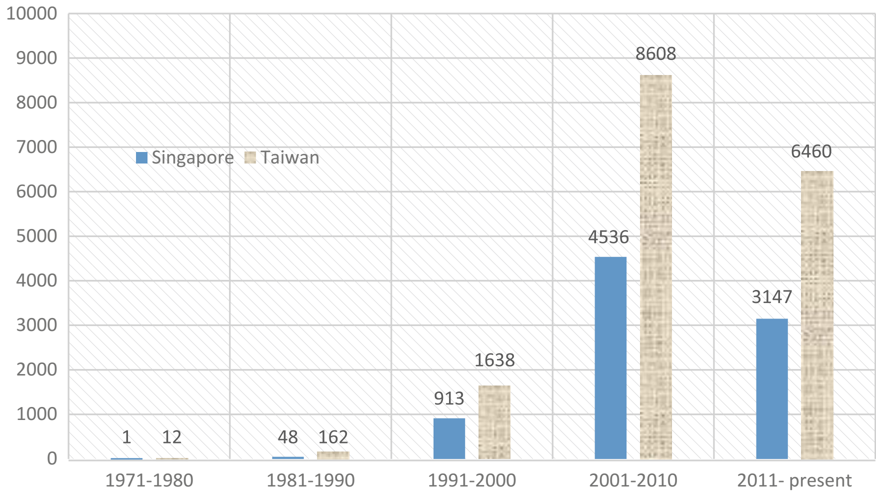


Fig. 3 Number of reliability related papers contributed from Singapore and Taiwan over the various decades. The *y*-axis represents the number of papers, and the actual number of papers are on *top of the bars*. The *x*-axis represents the various decades

5. There is only one failure mechanism in the failure data
6. Select the right suppliers through accelerated stress test.
7. To save time in reliability, simply increase the stress level.
8. Mil-Std handbook provides the basis for calculation

On the other hand, with the advancement in technology in Asia in the late 1990s, and the fact that manufacturers in Asia need to troubleshoot their low yield and customer field returns as part of their responsibilities, the knowledge of physics of failure is improving, and thus the reliability branch on PoF is progressing well in Asia.

With the evidence of good technical competency in Asia countries, such as Singapore and Taiwan, US companies are also shifting their R&D activities to these countries for cost reduction as well as to have their development activities close to their manufacturing sites and markets in Asia. This shift of R&D to Asia surface out the inadequate reliability knowledge needed for R&D activities.

On the other hand, universities in Asia are performing research in reliability engineering, both in the PoF as well as statistical approaches. Figure 3 shows the trend of Journal and conference papers on reliability engineering from Singapore and Taiwan since 1990. The research works done in universities are generally so much more advanced that engineers from industrials are either not able to understand the works or to apply the works in their workplaces. The gap in the reliability engineering knowledge between academic and industrial is large in Asia, and at the same time, reliability engineering training is scare even till now. There is no specialization on reliability engineering in almost all the universities in Asia.

To meet the need on reliability engineering evaluations for industry as more R&D activities are moving to Asia, various non-academic institutions in Asia are offering basic reliability training, where again the use of MH-217 and exponential distribution for reliability computation are taught. Some national institutes are offering certified reliability engineers program in order to meet the need. However, from the feedback of the trained engineers, they are not able to put what they have learnt into practice as their managements still holding the early concept of zero failure imply good reliability and the use of closed form formulae derived based on exponential distribution of time to failure for estimation of their product reliability. Management in Asia also tends to be put cost far above product reliability. On the other hand, as reliability is not being designed in the product and process during the R&D activities, products designed and manufactured in Asia tends to have poorer reliability as people feel generally. Of course, many exceptions are present in Asia also.

Due to the increasing need for reliability, reliability testing activities are getting popular. However, most of the testing laboratories simply perform the reliability tests according to the instruction from the clients or international standards, and no analysis of the failure and data are being done due to the complexity natures of the needed analyses.

While the industrial is still struggle with the basic reliability engineering in Asia, the complexities of current products and technologies present need new challenges in reliability engineering that require more research in it.

In summary, there is a need for the senior management in Asia to have a paradigm shift in their understanding of reliability, and I am seeing a promising change as I delivered several talks for industrial in several countries in Asia. Specialization in reliability engineering should also be provided in universities, and if not, by private institutions. But this is possible only if there are good markets for the trainees, and this again depends on the senior management. Unlike in US where an engineer can be a reliability engineer provided he or she has passed his or her certified reliability engineer program, this is not practice in Asia. Industrial should work with Universities to address the need for reliability knowledge for their engineers and to address the challenges in their product and process reliability.

References

James McLinn, “History of Reliability”, Originally published in Reliability Review of ASQ, March 2010.

William Denson, “The history of reliability prediction”, IEEE TRANSACTIONS ON RELIABILITY, VOL. 47, NO. 3-SP 1998 SEPTEMBER, pg. 321–328.

M. Azarkhail and M. Modarres, “The evolution and history of reliability engineering: Rise of mechanistic reliability modeling”, International J. of Performability Engineering, 8(1), pp. 35–47, Jan 2012.

C. M. Tan and N. Raghavan. “A framework to Practical Predictive Maintenance Modeling for Multi-State Systems,” *Reliability Engineering & System Safety*, vol. 93, pp. 1138, 2008.

C. M. Tan and N. Raghavan. "Root cause analysis based maintenance policy," *International Journal of Quality and Reliability Management*, 24(2), 203–228, 2007.
M. D. Le and C. M. Tan. "Optimal maintenance strategy of deteriorating system under imperfect maintenance & inspection using mixed inspection scheduling," *Reliability Engineering & System Safety*, V. 113(1), p. 21–29, Jan. 2013.
C. M. Tan. *Electromigration in ULSI Interconnections*. New York: World Scientific, 2010.
Cher Ming Tan, "Reliability Assessment of Integrated Circuits and Its Misconception", p. 45–68, in "Integrated Circuits, Photodiodes and Organic Field Effect Transistors", Ed. R. McIntire and P. Donnell, Nova Science Publishers, Inc. 2009.

Author Biography

Cher Ming Tan is the Director of Centre for Reliability of Sciences and Technologies (CReST) in Chang Gung University, Taiwan and also a Professor in her Electronic Department under the College of Engineering. He is also a Chair Professor in Ming Chi University of Technology, Taiwan. He was a faculty member of the Nanyang Technological University, Singapore for more than 19 years before he joined the current position. He started his reliability work in 1984 with Fairchild Semiconductor where the reliability and failure analysis works were needed for the US Defence. He later pursued his master and Ph.D. in University of Toronto, Canada. After graduation, he joined Lite-On Power Semiconductor, Taiwan, continued his work on quality, reliability and failure analysis. He is still active in providing training and consultation to companies that need reliability worldwide. He is an Editor of IEEE Transaction on Device and Materials Reliability, Editor of Scientific Report, Chief Editor of SpringerBrief on Reliability, Advisory board member of Microelectronic Reliability, and IEEE Distinguished lecturer of EDS in Reliability. He was the course coordinator of certified reliability program of Singapore Quality Institute.

Part I
Applications of Statistics in Industry

Dr. Siong Lin Ho, Ngee Polytechnic, Singapore

In today's business world and industry, Statistics play an important role in design, research & development, and also for monitoring, controlling, and improving the performance of products and processes. Over the past decades, as a result of advances in computer technology and statistical methodology—coupled with the propagation and advocates of quality improvement initiatives such as Six Sigma—the use and relevance of statistical tools and techniques are now more widespread in a myriad of industries, as compared to the early applications that focused primarily on manufacturing processes and quality control.

The papers presented at the ANQ2014 Congress provided useful insights and further demonstrated the use of statistical methods in a wide spectrum of applications—with techniques spanning from Statistical Process Control (SPC), Design of Experiments (DOE), nonparametric statistics, ANOVA, multivariate analysis, regression, Taguchi methods, chi-square goodness of fit test, to management science methods such as linear programming, decision tree, etc. Some interesting findings and key highlights of selected papers are summarised as follows:

SPC charts are important tools for monitoring process stability in manufacturing and service industries. In Uk Jung's paper titled "Nonparametric Wavelet-based Multivariate Control Chart for Rotating Machinery Condition Monitoring," he describes a new feature extraction technique that overcomes the high dimensionality problem of implementing multivariate SPC when the quality characteristic is a vibration signal from bearing system. Numerical simulation of a bearing system and the Average Run-Length (ARL) performance evaluation of the multiscale wavelet scalogram-based bootstrap control chart demonstrated that the proposed method has satisfactory fault-discriminating ability without any distributional assumptions.

An interesting application of multivariate process control is found in Lee and Cheng's paper titled "Identifying the Source of Variance Shifts in Multivariate Statistical Process Control using Ensemble Classifiers." The authors proposed a useful diagnostic aid in monitoring shifts in process dispersion. Out-of-control signals in multivariate control charts may be caused by one or more variables. In quality control, in addition to quick detection of process change, it is also critical to diagnose the change and identify which variable(s) has changed after an out-of-control signal occurred. Using a decision tree (DT)-based ensemble classifier in diagnosing out-of-control signal, simulation results found that the proposed model is successful in identifying the source of variance change.

Computer experiments usually treat large number of factors with wide range of levels. Consequently, there is a need to explore the large region spanned by large number of factors. The paper by Shu Yamada on “Design Criteria for Exploring Feasible Region of Factors in Multi-Dimensional Space” discusses the application of DOE and strategies for exploring the feasible region of factors in computer simulations. An example of flight simulator on the design and analysis of glider’s performance is illustrated.

About the author

Dr. Ho Siong Lin is currently a Senior Manager and Course Chair in the School of Engineering, Ngee Ann Polytechnic, Singapore. He received his Ph.D. in Industrial and Systems Engineering from the National University of Singapore, where he also obtained his B.Eng. (Hons) & M.Sc. in Electrical Engineering, M.Sc. in Industrial and Systems Engineering. He was awarded the Motorola Electronics Prize for achieving the best overall academic performance in his postgraduate specialization in Quality and Reliability Engineering and has also received the NP Academic Teaching Awards for his excellence in teaching. Dr. Ho is a Certified Reliability Engineer of the American Society of Quality. His research interests include six sigma, quality and reliability engineering, time series models, forecasting, neural networks, and he has published technical papers in several international conferences and journals. Dr. Ho is the lead author of two books on “Statistical Literacy for Industry and Business.”

Email: hsl@np.edu.sg

Identifying the Source of Variance Shifts in Multivariate Statistical Process Control Using Ensemble Classifiers

Chuen-Sheng Cheng and Hung-Ting Lee

Abstract Statistical process control charts are important tools for monitoring process stability in manufacturing and service industries. Using multivariate control charts to monitor two or more related quality characteristic has become increasingly popular in today's environment. Out-of-control signals in multivariate control charts may be caused by one or more variables or a set of variables. In the practice of quality control, in addition to the quick detection of process change, it is also critical to diagnose the change and identify which variable(s) has changed after an out-of-control signal occurred. In this paper, we propose a decision tree (DT)-based ensemble classifier to approach the problem of diagnosing out-of-control signal. The focus of this paper is on the variance shifts of a multivariate process. Simulation studies indicate that the proposed ensemble classification model is a successful method in identifying the source of variance change. A comparative study also reveals that DT using extracted features as input vector has slightly better classification performance than using raw data as input. The proposed method could be a useful diagnostic aid in monitoring shifts in process dispersion.

Keywords Multivariate control chart · Variance shift · Ensemble classifications · Decision tree

1 Introduction

Statistical process control charts have been widely used to control various industrial processes. In many industrial applications, the quality of a product or process can be attributed to several correlated variables. In this situation, a multivariate chart is needed to determine whether the process is in control. Many types of multivariate

C.-S. Cheng (✉) · H.-T. Lee
Department of Industrial Engineering and Management,
Yuan-Ze University, 135 Yuan-Tung Road,
Chung-Li, Taoyuan, Taiwan, ROC
e-mail: ieccheng@saturn.yzu.edu.tw

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_3

control charts have been proposed to monitor process shifts in the mean or in the process dispersion. The Hotelling T^2 control chart is commonly used to monitor the shifts in the mean of a multivariate process. Detecting shifts in the process variability is also of great importance in process control. In multivariate process control, process variability is characterized by the $p \times p$ covariance matrix $\mathbf{\Sigma}$, where p is the number of variables. The main diagonal elements of this matrix are the variances of the individual variables and the off-diagonal elements are the covariances. For detecting changes in the covariance matrix, a widely used approach is the generalized variance, $|\boldsymbol{S}|$, control chart (Alt 1985). What follows is a brief introduction of this control chart, for further details the reader is referred to Montgomery (2013). Consider a multivariate process with p correlated quality characteristics of interest. It is also assumed that the quality distribution of the p quality characteristics is the $p-$variate, normal distribution with the in-control mean vector $\boldsymbol{\mu}_0$, and covariance matrix $\mathbf{\Sigma}_0$. The control limits of the control chart for $|\boldsymbol{S}|$ can be expressed as $\mathrm{UCL} = |\mathbf{\Sigma}_0|\left(b_1 + 3\sqrt{b_2}\right)$, $\mathrm{CL} = b_1|\mathbf{\Sigma}_0|$, $\mathrm{LCL} = |\mathbf{\Sigma}_0|\left(b_1 - 3\sqrt{b_2}\right)$, where the coefficients b_1 and b_2 are given by

$$b_1 = \frac{1}{(n-1)^p} \prod_{i=1}^{p} (n-i) \tag{1}$$

$$b_2 = \frac{1}{(n-1)^{2p}} \prod_{i=1}^{p} (n-i) \left[\prod_{j=1}^{p} (n-j+2) - \prod_{j=1}^{p} (n-j) \right] \tag{2}$$

The lower control limit is replaced with zero if the calculated value is negative. Usually $\mathbf{\Sigma}_0$ is unknown, and it can be derived by a sample covariance matrix $\boldsymbol{S}$, based on a set of historical samples. In this situation, one would replace $|\mathbf{\Sigma}_0|$ by an unbiased estimator $|\boldsymbol{S}|/b_1$, where $|\boldsymbol{S}|$ is the determinant of the in-control covariance matrix. The chart statistic is $|\boldsymbol{S}_i|$, where $\boldsymbol{S}_i$ is the sample covariance matrix for sample i. The process variability is monitored by plotting $|\boldsymbol{S}_i|$ on the chart. If a value of $|\boldsymbol{S}_i|$ exceeds the UCL or LCL, then the process variability is considered to be out of control indicating that there has been a change in $\mathbf{\Sigma}$. Surtihadi et al. (2004) considered various possible patterns of shifts in the covariance matrix and developed control chart to detect these process change. Costa and Machado (2009) proposed a control chart for detecting shifts in the covariance matrix of multivariate process. Machado et al. (2009) proposed a new statistic to control the covariance matrix of bivariate process. When an out-of-control event is identified as a signal in a control chart, a diagnostic procedure is needed for locating the specific variables that are the cause of the problem. Although control charts are efficient in detecting shift in the variance, they do not reveal which variable (or combination of variables) caused the signal. Due to this reason, identifying the responsible variable for the out-of-control process has become one of the most important topics in the field of multivariate process control. For monitoring process mean shift, the most effective method that addresses this problem is a decomposition of the control statistic

(Runger et al. 1996). However, it is evident from the literature that little work has been directed toward the identification of the source of process variability.

Various soft computing techniques have recently been proposed to identify the responsible variables when the process is considered to be out of control (Low et al. 2003; Aparisi et al. 2006; Guh 2007). To identify the source of variance shift, Cheng and Cheng (2008) considered two classifiers based on artificial neural networks (ANN) and support vector machine (SVM) to interpret the signals issued by a generalized variance chart. Cheng et al. (2011) proposed a model based on a least squares support vector machine (LS-SVM) to diagnose abnormalities in the bivariate processes of a covariance matrix.

Lately, ensemble classification techniques have received much attention and have demonstrated promising capabilities in improving classification accuracy over single classifier (Hansen and Salamon 1990; Kim and Kang 2010; Yu et al. 2010). The ensemble techniques have been used in the area of multivariate process control to identify the source(s) of out-of-control signals. Recent examples are reported in the literature (Alfaro et al. 2009; Yu and Xi 2009a, b; Yu et al. 2009; Cheng and Lee 2014).

In this paper, we propose using the decision tree (DT) ensemble as a diagnostic aid to determine which variable(s) in the process that has changed after the chart triggered a signal. The present work considers the identification of the source(s) of the variance shifts. The proposed system includes a variance shift detector and a classifier. The traditional $|\boldsymbol{S}|$ chart is employed as a variance shift detector. An ensemble learning approach is used to construct the classifier. The ensemble classifier includes a number of DT-based models. The proposed system is depicted in Fig. 1. The ensemble classifier, which determines the responsible variable(s) for the variance shift, starts after the $|\boldsymbol{S}|$ chart sends out an out-of-control signal. This study attempts to extend previous research in several directions. First, ensemble model is created by considering the statistical properties of dataset. Second, a set of features is proposed to further improve the classification accuracy.

We will illustrate the proposed approach using simulated multivariate data. This paper demonstrates that the DT ensemble can improve the classification accuracy greatly than using a single DT. The rest of this paper is organized as follows. In Sect. 2, we briefly describe the fundamental theory of DT. This section also provides the basic concept of ensemble-based classification model. Section 3 presents the details of building the proposed ensemble classifier. Next, experimental results are given in Sect. 4. Finally, in Sect. 5, we summary the main conclusions from this study and provide some directions for future research.

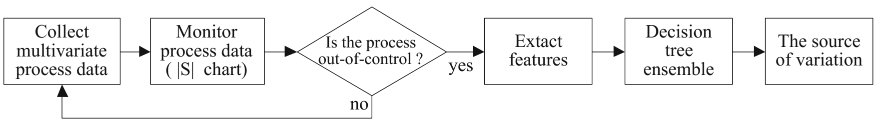


Fig. 1 The proposed diagnosis system

2 Theoretical Background

2.1 Decision Tree

Decision tree (DT) is a non-parametric supervised learning method used for classification and regression. It is composed of root, internal, and leaf nodes. DTs have attracted wide popularity due to their efficiency and simplicity in solving a wide range of problems in various fields.

There are several algorithms available to construct a DT. The research described in this paper considers the CART (Classification and Regression Trees) algorithm developed by Breiman et al. (1984). The goal of CART is to explain the variation of a single response (dependent) variable using a set of explanatory (independent) variables, via a binary partitioning procedure. Both the response and the explanatory variables can be either categorical or numerical. Depending on the data type of response variable, CART can create either classification tree or regression tree. A classification tree is produced when the response variable is categorical while a regression tree is obtained for a numerical response variable. This paper focuses on using CART to build a classification model.

The model building process of CART involves three basic steps (Breiman et al. 1984). In the first step a tree is built using a recursive partitioning technique to select variables and split points based on a splitting criterion (or splitting function). Several criteria are available for determining the splits, including gini, and towing. For a more detailed description of the mentioned criteria one can refer to Breiman et al. (1984).

Once a large tree is identified, the second step of the CART algorithm uses a pruning method that incorporates a minimal cost complexity measure. The pruning procedure converts a large tree into a less complex tree making it easier to understand. Properly pruned tree will result in a faster implementation and avoid over-fitting. The last step of the methodology is to select the optimal tree. The selection is usually based on the evaluation of cross-validation error.

2.2 Ensemble-Based Classification Model

Ensemble learning is the process of creating multiple models and combining the results of individual models in order to solve a particular problem. This section focuses on classification-related applications of ensemble learning. An ensemble classification model is defined as a set of basic classifiers whose individual outcomes are combined (or fused) in some manner to form a final decision. It has been shown that ensemble classification models outperform single classification models, whose performance is limited by the imperfection of feature extraction, learning algorithms, and the inadequacy of training data (Hansen and Salamon 1990).

In general, an effective ensemble-based model consisting of diverse models with much disagreement is more likely to obtain a good generalization performance.

In other words, the ensemble members in the ensemble model must possess different classification properties. Therefore, how to generate the diverse models is a crucial factor for constructing an effective ensemble classification system. There are various methods suggested for the creation of an ensemble of classifiers (Dietterich 1997). Strategies that can be used to create an ensemble of classifiers include (1) using different subsets of training data with a single learning algorithm; (2) using different training parameters with a single learning algorithm; and (3) using various types of learning algorithms. The above methods are usually referred to as diversity strategies.

Using data diversity to produce ensemble classifier relies on varying the data in some way. Commonly used methods of varying the data include different sampling methods, use of different data sources, use of different preprocessing methods, distortion, and adaptive resampling (Kotsiantis and Pintelas 2004). Among them, we put focus on the method of using different training data subsets. We consider the most popular methods such as bagging (Breiman 1996) and boosting (Freund 1995). They are briefly described below.

Bagging (stands for bootstrap aggregating) is used to create and combine multiple classifiers. Diversity of classifiers in bagging is obtained using bootstrapped replicas of the training data. The bootstrap data is created by resampling samples randomly with replacement from the original training set. Each classifier is created by training on corresponding bootstrap replicate. Individual classifiers could be trained in parallel and combined by taking a simple majority vote of their decisions. The class chosen by most number of classifiers is the final decision.

Boosting constructs an ensemble classifier by iteratively training classifiers. Boosting algorithm assigns weights to the training samples, and these weight values are changed depending on how well the sample is learned by the classifier. Typically, the weights for misclassified samples are increased. Thus boosting attempts to produce new classifiers that are better able to classify samples for which the current ensemble's performance is poor. Boosting combines classifications of ensemble of classifiers with weighted majority voting by giving more weights on more accurate classifications. The most common implementation of boosting is the AdaBoost algorithm (Freund and Schapire 1997).

The following section describes an ensemble classification model using different subsets of training data with a single training method.

3 Proposed Decision Tree-Based Ensemble Classification Model

The section describes the proposed ensemble classification model using different subsets of training data within a single training method. First, we describe the construction of a single classifier, including the training dataset, input vector, and model training. Next, we present the development of ensemble classifiers.

3.1 Building Single DT-Based Classification Model

3.1.1 Generation of Training Data Set

Preparing an appropriate training dataset is a critical step in constructing an effective classifier. Due to extensive demands for training data, Monte Carlo simulation is the often-recommended method to generate training data. A p-dimensional multivariate normal process was simulated by generating pseudo-random variates from a multivariate normal distribution whose mean was $\boldsymbol{\mu}$ and whose covariance matrix was $\boldsymbol{\Sigma}$. By manipulating the value of covariance matrix $\boldsymbol{\Sigma}$, an in-control or an out-of-control process can then be simulated. This paper considers the case of increasing in process dispersion, which corresponds to deterioration of process or product quality. We consider the patterns of shifts in the covariance matrix described in Surtihadi et al. (2004). Using the notation of Surtihadi et al. (2004), a description of the matrices follows.

Let $\boldsymbol{\Sigma}_0$ denote the known in-control covariance matrix. For simplicity, it is assumed that the covariance matrix $\boldsymbol{\Sigma}_0$ is known. In practice it will be necessary to collect sufficient data over a substantial amount of time when the process is in control to estimate $\boldsymbol{\Sigma}_0$:

$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix} \tag{3}$$

The entry σ_{ij} $(i \neq j)$ is the covariance between variables i and j, when $i = j$ the diagonal entries of the covariance matrix are equal to the variances of the individual variables. When abnormal disturbance affects the process variance, the in-control covariance matrix $\boldsymbol{\Sigma}_0$ is changed to $\boldsymbol{\Sigma}_1$. The following change schemes are considered. If the monitored process involves a change in one variable, then the out-of-control covariance matrix $\boldsymbol{\Sigma}_1$ is written as follows:

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} \theta^2\sigma_{11} & \theta\sigma_{12} & \cdots & \theta\sigma_{1p} \\ \theta\sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \theta\sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix} \tag{4}$$

When the dispersions of all variables change simultaneously, each element of the covariance matrix gets multiplied by an unknown constant θ. The out-of-control covariance matrix $\boldsymbol{\Sigma}_1$ can be expressed as

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} \theta\sigma_{11} & \theta\sigma_{12} & \cdots & \theta\sigma_{1p} \\ \theta\sigma_{12} & \theta\sigma_{22} & \cdots & \theta\sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \theta\sigma_{1p} & \theta\sigma_{2p} & \cdots & \theta\sigma_{pp} \end{bmatrix} = \theta\,\boldsymbol{\Sigma}_0 \tag{5}$$

The above change patterns were considered in the work of Cheng and Cheng (2008). The present work considers the identification of the source(s) of the variance shifts after the control chart issues an out-of-control signal. Due to this reason, each selected sample data must exceed the UCL of the $|\boldsymbol{S}|$ chart. This study considered the case that process dispersion increases. Shifts were investigated for inflation factor θ from 1.5 to 5.0 in steps of 0.25. In this study, the simulation was implemented using MATLAB. For convenience, all variables were scaled such that each process variable has mean zero and variance one. With this approach, the covariance matrix is in correlation form; that is, the main diagonal elements are all one and the off-diagonal elements are the pairwise correlation (ρ) between the process variables. Here we assume that the process mean vector remains constant as the process dispersion shifts. The current study considers the case of equi-correlation (pairs of variables are equi-correlated). The values of ρ were from 0.1 to 0.9 in steps of 0.2. It will also be assumed that the value of ρ is stable with respect to shifts in the process dispersion. Table 1 summarizes the shift patterns considered for each ρ.

3.1.2 Selection of Input Vector

Determining the input feature vector to the classifier is an important task in building a classification model. In the present research, there are many different elements that could be used to construct an appropriate feature vector. The first feature vector (henceforth vector-I) consists of sample data observed in the kth sample with $|\boldsymbol{S}_k| > \text{UCL}$. An alternative is to use computed sample statistics as the input components (hereafter called vector-II). The feature vector can be written as $(S_{1k}, S_{2k}, \ldots, S_{jk}, \bar{X}_{1k}, \bar{X}_{2k}, \ldots, \bar{X}_{jk}, D_{1k}, D_{2k}, \ldots, D_{lk})$, where S_{jk} is the standard

Table 1 Shift patterns

Variable	Shift type	Class label
$p = 2$	The 1st variable shifts	1
	The 2nd variable shifts	2
	All variables shift	3
$p = 3$	The 1st variable shifts	1
	The 2nd variable shifts	2
	The 3rd variable shifts	3
	All variables shift	4

deviation on the jth variable in the kth sample, $\bar{X}_{jk}$ is the sample mean of the jth variable, and D_{lk} is the Mahalanobis distance obtained from the kth sample data, with subscript l being the index of class l. In the present research, Mahalanobis distance measures the distance between an observation and a given class (shift type). Obviously, this feature vector provides a compact and efficient representation of the input vector. A similar feature vector was proposed by Cheng and Lee (2014). Later in this paper, we will compare the performance of these two alternatives.

3.1.3 Building Tree Models

The tree models were built using MATLAB classregtree procedure. The class label was used as the response variable and input features described in Sect. 3.1.1 as explanatory variables. The "splitcriterion" was set to be Gini's diversity index.

3.2 Proposed DT Ensemble Model

This section presents the specific design details of the ensemble model. Building ensemble classifier involves two parts: generation of classifiers and design of ensemble strategy. In this study, using different training datasets is selected as the diversity strategy to create ensemble members. This approach is referred to as data diversity. Under this approach, multiple classifiers are trained using different training subsets. This is in contrast to traditional approaches which try to train a classifier using a single training set (Aparisi et al. 2006; Cheng and Cheng 2008). In previous studies, samples corresponding to various magnitudes of process changes must be considered simultaneously in a single training set. In other words, the training set involves a multiple-shift distribution. This approach complicates the training process.

If the dataset could be functionally or structurally divisible into different training subsets then different subsets will be a feasible approach to create data diversity. In this paper, we propose using the shift size for preparing different training data subsets. We create multiple classifiers, each of which is trained using the same learning algorithm (DT) with a training set corresponding to a particular shift size, θ. Using this setting, the number of classifiers of the ensemble model (i.e., ensemble size) is equal to 15.

After training, each classifier has generated its own result. The next step is to combine these classifiers into an aggregated classifier using an ensemble strategy. In the present study, a majority voting-based ensemble strategy is adopted due to its ease of implementation. The decision of the ensemble classifier is obtained from the decision results of DTs via a majority voting-based ensemble strategy. Figure 2 illustrates the implementation of the DT ensemble.

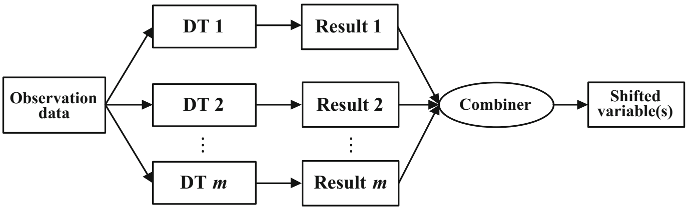


Fig. 2 The structure of DT ensemble

4 Results and Discussion

This section presents the comparisons of classification accuracy (%) of various classifiers, including single DT, bagging, AdaBoost, and the proposed DT ensemble. To demonstrate the power of the ensemble classifier and the benefits of proposed features, we used single DT method and vector-I as a baseline. The test data were generated in same way as training data. Shifts were investigated for θ from 1.5 to 5.0 in steps of 0.25. The sample size considered in this study is $n = 5$. Bagging and AdaBoost ensemble models were created using "fit ensemble" function of MATLAB. For bagging ensemble model, the input argument "NLearn" was set to be 15. This setting was made so as to have a fair comparison between bagging and the proposed DT ensemble. For AdaBoost ensemble classifier, the input argument "Method" was set to be "AdaBoostM2." Various values of "NLearn" ranging from 1 to 100 were tried. Only the best result was reported.

In each simulation run, 1000 test samples (five times the size of training samples) were created for each θ. For each ρ, classification rate was approximated based on 15000 test samples for each variance shift pattern. Based on this simulation setting, each classifier has been performed 10 independent runs of simulation and the average performance is reported. Tables 2 and 3 summarize the overall performances of classification models in terms of the classification accuracy. The results are reported for $\rho = 0.1, 0.3, 0.5, 0.7, \text{ and } 0.9$. It can be observed from Tables 2, 3 that models built with vector-II perform better than those with vector-I. It can also be observed that DT ensemble uniformly performs better than single DT.

Table 2 Comparison of the classification accuracy ($p = 2$)

ρ	Vector-I				Vector-II			
	Single DT	AdaBoost	Bagging	DT ensemble	Single DT	AdaBoost	Bagging	DT ensemble
0.1	76.66	79.96	80.89	84.44	80.30	80.82	85.56	90.83
0.3	77.88	81.57	82.15	85.71	80.95	81.56	85.77	90.88
0.5	79.80	83.42	84.34	87.19	83.29	84.38	87.60	91.19
0.7	84.37	85.88	87.91	89.13	86.54	87.82	90.45	92.90
0.9	90.61	91.60	93.36	94.12	93.32	94.73	95.73	96.42

Table 3 Comparison of the classification accuracy ($p = 3$)

ρ	Vector-I				Vector-II			
	Single DT	AdaBoost	Bagging	DT ensemble	Single DT	AdaBoost	Bagging	DT ensemble
0.1	82.19	83.99	86.17	87.29	83.74	84.16	88.11	92.06
0.3	82.31	84.56	86.63	87.69	84.55	84.63	88.43	92.38
0.5	84.18	86.03	87.60	88.52	85.72	86.03	89.20	92.86
0.7	85.55	88.08	90.45	90.59	88.46	90.16	91.54	94.33
0.9	93.57	94.50	95.73	96.52	94.22	95.08	96.16	97.47

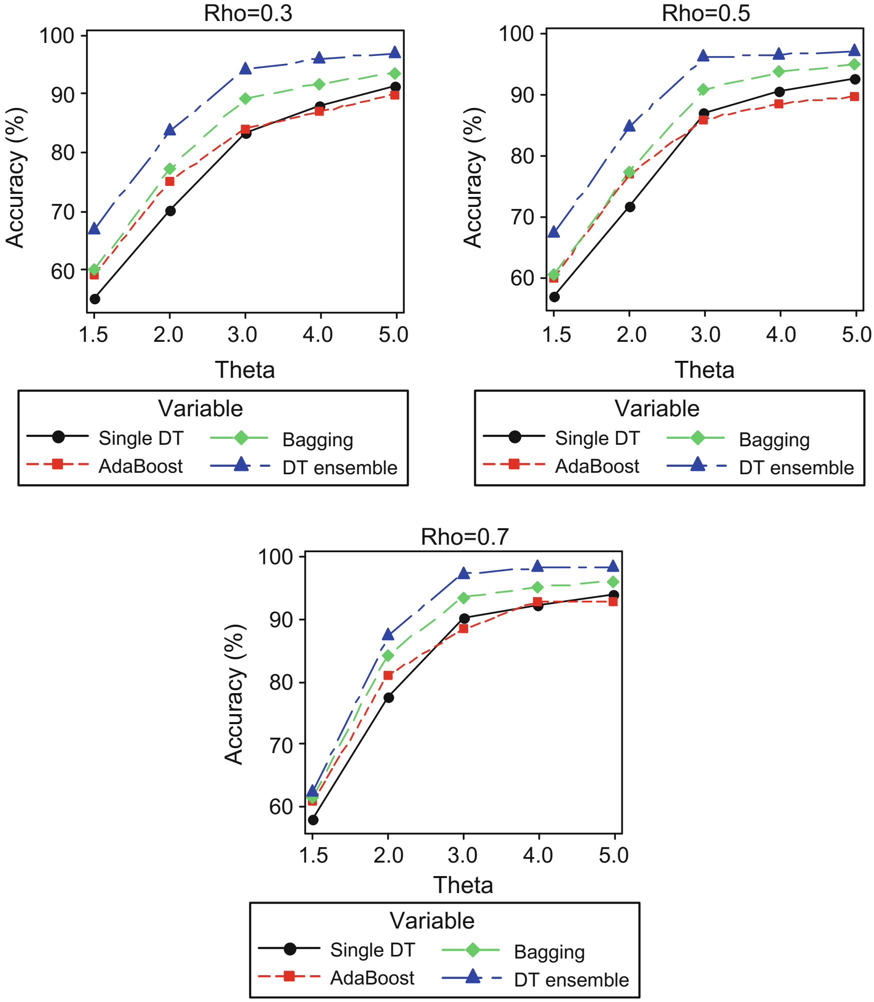


Fig. 3 Comparison of classification accuracy ($p = 2$)

Generally the proposed method has the best performance, followed by bagging and AdaBoost. Figure 3 presents the classification accuracy in the graphical form, as a function of θ. The results are reported for $p = 2$ and $\rho = 0.3, 0.5,$ and 0.7. It can be seen that DT ensemble consistently performs better than other classification models irrespective of the size of θ.

Based on the above results, it can be concluded that ensemble method can improve the accuracy in the classification of shifted variable(s). The results show that the proposed method can be a useful diagnostic tool in multivariate process control.

5 Conclusion

Identifying the variable(s) responsible for the variance shifts is an important topic in multivariate process control. In this paper, a DT-based ensemble classification model has been proposed for the interpretation of variance shift signal in multivariate processes. The focus is on using the data diversity strategy to create an effective ensemble classifier.

The results of extensive simulation studies indicate that the proposed DT-based ensemble classifier provides better performance than single classifier. The proposed DT ensemble model can also consistently outperform the other comparable models including bagging and boosting. Comparative studies also reveal that substantial improvement in classification accuracy can be obtained when DT was built using relevant features. It is believed that the proposed DT-based ensemble classification system can be used as a useful diagnostic tool in the interpretation of out-of-control signal in multivariate process control. The main contributions of this paper include a set of features useful for classification and a data diversity strategy based on shift size.

Further studies are needed to address some important issues. First, selecting appropriate number of ensemble members is an important issue to be studied in future work. Second, further improvements of ensemble classifiers could be achieved using new ensemble strategies. Moreover, future research could investigate and evaluate ensemble of different classifiers.

References

Alfaro, E., Alfaro, J. L., Gámez, M., and García, N. (2009): "A boosting approach for understanding out-of-control signals in multivariate control charts", *International Journal of Production Research*, 47, pp 6821–6831.

Alt, F. B. (1985): Multivariate quality control. In *Encyclopedia of Statistical Science*, vol. 6. Johnson, N. L., and Kotz, S. (eds.), John Wiley and Sons, New York.

Aparisi, F., Avendaño G., and Sanz, J. (2006): "Techniques to interpret T^2 control chart signals", *IIE Transactions*, 38, pp 647–657.

Breiman, L. (1996): Bagging predictors. *Machine Learning*, 24, pp 123–140.

Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J., (1984): *Classification and Regressing Tree*, Wadsworth, Pacific Grove, CA.

Cheng, C. S., and Cheng, H. P. (2008): "Identifying the source of variance shifts in the multivariate process using neural networks and support vector machines", *Expert System with Applications*, 35, pp 198–206.

Cheng, C. S., and Lee, H. T. (2014): "Applying decision tree-based ensemble classifiers for diagnosing mean shift signals in multivariate control chart", *Journal of Quality*, 21, pp 91–103.

Cheng, Z. Q., Ma, Y. Z., and Bu, A. J. (2011): "Variance shift identification model of bivariate process based on LS-SVM pattern recognizer", *Communication in Statistics: Simulation and Computation*, 40, pp 286–296.

Costa, A. F. B., and Machado, M. A. G. (2009): "A new chart based on sample variance for monitoring the covariance matrix of multivariate processes", *The International Journal of Advanced Manufacturing Technology*, 41, pp 770–779.

Dietterich, T. G. (1997). Machine learning research: Four current directions. *AI Magazine*, 18, 97–136.

Freund, Y. (1995): "Boosting a weak learning algorithm by majority", *Information and Computation*, 121, pp 256–258.

Freund, Y., and Schapire, R. E. (1997): "A decision theoretic generalization of online learning and an application to boosting" *Journal of Computer and System Science*, 55, pp 119–139.

Guh, R.S. (2007): "On-line identification and quantification of mean shifts in bivariate processes using a neural network-based approach", *Quality and Reliability Engineering International*, 23, pp 367–385.

Hansen, L., and Salamon, P. (1990): "Neural network ensembles", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12, pp 993–1001.

Kim, M. J., and Kang, D. K. (2010): "Ensemble with neural networks for bankruptcy prediction", *Expert Systems with Applications*, 37, pp 3373–3379.

Kotsiantis, S. B., and Pintelas, P. E. (2004): "Combining bagging and boosting", *International Journal of Computational Intelligence*, 1, pp 324–333.

Low, C., Hsu, C. M. and Yu, F. J. (2003): "Analysis of variations in a multi-variate process using neural networks", *International Journal of Advanced Manufacturing Technology*, 22, pp 911–921.

Machado, M. A. G., Costa, A. F. B., and Rahim, M. A. (2009): "The synthetic control chart based on two sample variances for monitoring the covariance matrix", *Quality and Reliability Engineering International*, 25, pp 595–606.

Montgomery, D. C. (2013): *Introduction to Statistical Quality Control* (7th ed.), Wiley and Sons, New York.

Runger, G. C., Alt, F. B., and Montgomery, D. C (1996): "Contributors to a multivariate statistical process control signal", *Communications in Statistics-Theory and Methods*, 25, pp 2203–2213.

Surtihadi, J., Raghavachari, M., and Runger, G. (2004): "Multivariate control charts for process dispersion", *International Journal of Production Research*, 42, pp 2993–3009.

Yu, J. B., and Xi, L. F. (2009a): "A neural network ensemble-based model for on-line monitoring and diagnosis of out-of-control signals in multivariate manufacturing processes", *Expert Systems with Applications*, 36, pp 909–921.

Yu, J. B., and Xi, L. F. (2009b): "A hybrid learning-based model for on-line monitoring and diagnosis of out-of-control signals in multivariate manufacturing processes", *International Journal of Production Research*, 47, pp 4077–4108.

Yu, J. B., Xi, L.F., and Zhou, X. J. (2009): "Identifying source(s) of out-of-control signals in multivariate manufacturing processes using selective neural network ensemble", *Engineering Applications of Artificial Intelligence*, 22, pp 141–152.

Yu. L., Yue, W., Wang, S., & Lai, K. K. (2010): "Support vector machine based multiagent ensemble learning for credit risk evaluation", *Expert Systems with Applications*, 37, pp 1351–1360.

Authors Biography

Chuen-Sheng Cheng is Professor and Chair in the Department of Industrial Engineering and Management at Yuan Ze University, Taiwan. He received his Ph.D. degree in industrial engineering from Arizona State University (U.S.A.) in 1989, an M.S. degree in industrial engineering from Texas Tech. University (U.S.A.) in 1985, and a B.S. degree in industrial engineering from National Tsinghua University (Taiwan) in 1980. His research interests include statistical quality control, intelligent process monitoring, and applications of artificial neural networks. He has authored several books in the areas of quality management and statistical quality control. Professor Cheng has extensive experience in training and consultancy in the areas of statistical quality control, quality management, and Six Sigma. In 2000, Professor Cheng obtained the Individual Award of the 11th National Quality Awards of the Republic of China. Also, as the key facilitator, he led Yuan Ze University in winning the "Organization Award" of the 14th National Quality Award in 2003. Professor Cheng is the Editor-in-Chief of Journal of Quality, and he is also a Fellow and a council member of the Chinese Society for Quality (CSQ).

Hung-Ting Lee is a senior engineer at Unimicron Technology Corporation. She received her Ph. D. degree in industrial engineering from Yuan Ze University in 2014, an M.S. degree in industrial engineering and management from National Formosa University in 2008. Her primary research interests include multivariate statistical process control and intelligent process control.

Nonparametric Wavelet-Based Multivariate Control Chart for Rotating Machinery Condition Monitoring

Uk Jung

Abstract Signal processing of vibration signals from rotating machinery has been an active research area for recent years. Especially, discrete wavelet transform (DWT) is considered as a powerful tool for feature extraction in detecting fault on rotating machinery. However, the number of retained DWT features can be still too large to be used for standard multivariate statistical process control (SPC) techniques although DWT significantly reduces the dimensionality of the data. Even though many feature-based SPC methods have been introduced to tackle this deficiency, most of methods require a parametric distributional assumption that restricts their feasibility to specific problem of process control and thus limits their applications. This study introduced new feature extraction technique to alleviate the high dimensionality problem of implementing multivariate SPC when the quality characteristic is a vibration signal from bearing system. A set of multiscale wavelet scalogram features was generated to reduce the dimensionality of data, and is combined with the bootstrapping technique as nonparametric density estimation to set up an upper control limit of control chart. Our example and numerical simulation of a bearing system demonstrated that the proposed method has satisfactory fault-discriminating ability without any distributional assumption.

Keywords Multivariate control chart · Fault detection · Wavelet transformation · Bootstrap technique · Vibration signal

1 Introduction

Evaluating structural integrity of engineering structures has been a critical research topic for last few decades in conjunction with ever-increasing demands on the longevity of service life. Most of engineering structures are easily exposed to unforeseen defects resulting from inside and outside of their environment such as

U. Jung (✉)
School of Business, Dongguk University, Seoul, Korea
e-mail: ukjung@dongguk.edu

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_4

fatigue crack, thermal degradation, impact, overloading, and corrosion during their normal operating condition. Because overall cost of repair or downtime could be enormous for severely damaged social infrastructures, the importance of early detection and prognosis of damage has been greatly increased. The oldest and the simplest measure of damage detection such as visual inspection, however, has been a very limited option only for examining human-accessible areas, which becomes impossible in a certain circumstance. For these reasons, structural health monitoring systems using the measurements of vibration characteristics of the structure have gained attention in recent years (Palacz and Krawczuk 2002; Sohn et al. 2003).

Many of vibration-based damage detection methods exploit the change of dynamic properties such as modal parameters (natural frequencies, mode shapes, and frequency response functions) due to stiffness or mass variations in a structure, so that structural integrity can be remotely and indirectly assessed without physically involving human inspection. However, one of the main stumbling blocks for vibration-based approach is inherently low sensitivity of modal properties toward small size of structural defect (Swamidas and Chen 1995). It is very difficult to discriminate the variation of damage-caused modal properties until the severity of damage becomes significant.

Rolling element bearings are the most popular locations of potential defects in rotating machinery which is one of engineering structures. Because the essence of bearing condition monitoring is a prognostic warning of incipient fault, advanced signal processing and fault diagnosis are very critical. In this context, signal processing of vibration measurement has been an active research area in recent years (Lin and Zhang 2005; Lei et al. 2011; Koh et al. 2008). Typically, a bearing system with a defect experiences periodic impulses in its vibration signal, due to repeated collisions between the defect and rolling components of the bearing.

The typical condition monitoring process is composed of two steps: first, acquisition of data from sensor measurements around the bearing system. Then, extracting damage-sensitive features from the measured data, and classifying the pattern for further diagnosis, such as allowing a decision for maintenance (Jardine et al. 2006). Over the last three decades, feature extraction from vibration signal has been widely investigated for monitoring bearing conditions, in conjunction with advanced signal processing techniques (Randall and Antoni 2011). Considering the nonstationary nature of bearing dynamics, many studies are devoted to the time-domain and/or time–frequency analysis. In the time domain, the simplest fault detection method is to evaluate the crest factor or kurtosis value from measurements (Dyer 1978). Also, empirical mode decomposition (Yu et al. 2006), wavelet transform (Prabhakar et al. 2002), and autoregressive filter (Wang 2002) have been successfully employed for monitoring fault in rotating machinery.

Among many other fields, condition monitoring for rotating machinery using statistical process control (SPC) combined with statistical pattern recognition methodology, such as feature extraction, has become an area of promising research. Recent literature has presented several variants of a wavelet-decomposition-based multiscale data analysis method for vibration-based fault detection, which are, in concept, superior to those of the traditional single scale methods, such as the time series method (Jung and Koh 2009; Jung et al. 2006). A discrete wavelet transformation (DWT) uses appropriate low-pass and high-pass filters, to map the signal from the time domain into the time-scale domain. While DWT significantly reduces the dimensionality of the data, using denoising techniques, the number of retained wavelet coefficients can still be significantly large. Hence, additional dimensionality reduction effort is necessary to summarize the wavelet coefficients. Consequently, a DWT is only a start, and further analysis of the coefficients is needed for multivariate monitoring process.

Thus, in this present study, the multiscale wavelet scalogram (MWS) is selected as a set of defect-sensitive features, especially in very low dimension, for constructing a control chart, to avoid a high false alarm rate (FAR). MWS provides measures of signal energy at various frequency bands, and is commonly used in decision making tasks, including signal and image processing, astronomy, and metrology (Rioul and Vetterli 1991; Scargle 1997; Peng et al. 2002; Jeong et al. 2003). MWS represents the scale-wise distribution of energies of the signal. These energies at different scales have distinct characteristics of different signals. Therefore, intuitively, it is possible that MWS will be useful in monitoring process changes, with data collected in time sequences.

Traditionally, a control chart assumes that the monitoring statistics follow a certain probability distribution. The most widely used multivariate control chart has been Hotelling's T^2 (T^2 chart) (Hotelling 1947). The T^2 charts monitor T^2 statistics that measure the distance between an observation and the scaled-mean, estimated from the in-control data. Assuming that the observed process data follow a multivariate normal distribution, the control limit of a T^2 control chart is proportional to the percentile of an β distribution in Phase I (retrospective analysis), and an F distribution in Phase II (future monitoring) (Mason and Young 2002). However, when the normality assumption of the data does not hold, a control limit based on the β and F distribution may be inaccurate, because the control limit determined this way may increase the false alarm rate (Chou et al. 2001). To tackle the limitation posed by the distributional assumption underpinning traditional control charts, nonparametric (or distribution-free) control charts have been developed. In particular, many studies have focused on the construction of nonparametric control charts using a bootstrap procedure (Bajgier 1992; Seppala et al. 1995; Liu and Tang 1996; Lio and Park 2008; Polansky 2005; Phaladiganon et al. 2011). This procedure is favored over the traditional distribution-assuming procedures, because of its proven capabilities to effectively manage process data, without posing assumptions about their distribution.

In this study, multivariate control chart analysis, which is the most commonly used SPC technique, and very suitable for automated condition monitoring, is applied to evaluate vibration signals, for further diagnosing the incipient fault in a bearing system. In our case, even though MWS composes a set of good features with high discriminating power in very low dimension, suitable for SPC application, the probability distributional information of MWS is theoretically intractable, due to the nature of the nonparametric approach, so that the traditional multivariate statistical control chart, such as Hotelling's T^2 control chart, is not a good choice. To overcome the limitation posed by the parametric assumption, this study proposes a nonparametric multivariate control chart method, using a bootstrap approach based on MWS-based monitoring statistic for fault detection. Before the proposed method is introduced, we first summarize a discrete wavelet transformation and multivariate SPC techniques.

The organization of this paper is as follows. The theoretical background for discrete wavelet transformation and multivariate SPC is outlined in Sect. 2. In Sect. 3, the proposed method for monitoring a high-dimensional vibration signal is explained. In Sect. 4, the illustrative example of a bearing system and vibration signal acquisition is summarized. In Sect. 5, the performance of the proposed method is evaluated, through simulation of the bearing signal. Finally, Sect. 6 contains the conclusion and future research work.

2 Theoretical Background: DWT, SPC, and Bootstrapping

2.1 Discrete Wavelet Transformation (DWT)

This section provides a brief review of DWT (Mallat 1989; Vidakovic 1999). A wavelet is a function $\psi(t) \in L^2(R)$ with the following basic properties: $\int_R \psi(t)\,\mathrm{d}t = 0$ and $\int_R \psi^2(t)\,\mathrm{d}t = 1$, where $L^2(R)$ is the space of square integrable real functions defined on the real line R. Wavelets can be used to create a family of time–frequency atoms, $\psi_{s,u}(t) = s^{1/2}\psi(st - u)$, via the dilation factor s, and the translation u. We also require a scaling function $\phi(t) \in L^2(R)$ that satisfies $\int_R \phi(t)\,\mathrm{d}t \neq 0$ and $\int_R \phi^2(t)\,\mathrm{d}t = 1$. Selecting the scaling and wavelet functions as $\{\phi_{L,k}(t) = 2^{L/2}\phi(2^L t - k); k \in Z\}$, $\{\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k);\ j \geq L,\ k \in Z\}$, respectively, one can form an orthonormal basis to represent a signal function $f(t) \in L^2(R)$, as follows: $f(t) = \sum_{k\in Z} c_{L,k}\phi_{L,k}(t) + \sum_{j\geq L}\sum_{k\in Z} d_{j,k}\psi_{j,k}(t),$, where Z denotes the set of all integers $\{0, \pm 1, \pm 2, \ldots\}$, and the coefficients $c_{L,k} = \int_R f(t)\phi_{L,k}(t)\,\mathrm{d}t$ are considered to be the approximation level coefficients characterizing smoother data patterns, and $d_{j,k} = \int_R f(t)\psi_{j,k}(t)\,\mathrm{d}t$ are viewed as the detail level coefficients describing (local) details of data patterns. In practice, the following finite version of the wavelet series approximation is used: $\tilde{f}(t) = \sum_{k\in Z} c_{L,k}\phi_{L,k}(t) + \sum_{j=L}^{J}\sum_{k\in Z} d_{j,k}\psi_{j,k}(t),$, where $J > L$ and L correspond to the coarsest resolution level. Consider a sequence of data $\mathbf{y} = (y(t_1), \ldots, y(t_N))^{\mathrm{T}}$ taken from $f(t)$. The DWT of $\mathbf{y}$ is defined as $\mathbf{d} = (\mathbf{d}_{(1)}, \mathbf{d}_{(2)}, \ldots, \mathbf{d}_{(I)})^{\mathrm{T}}$ where $\mathrm{I} = J - L + 2$.

In Fig. 1, the plot on the top shows the original vibration signal on time–amplitude representation, and the one on the bottom represents a plot on which the horizontal axis represents position along the time, the vertical axis represents scale

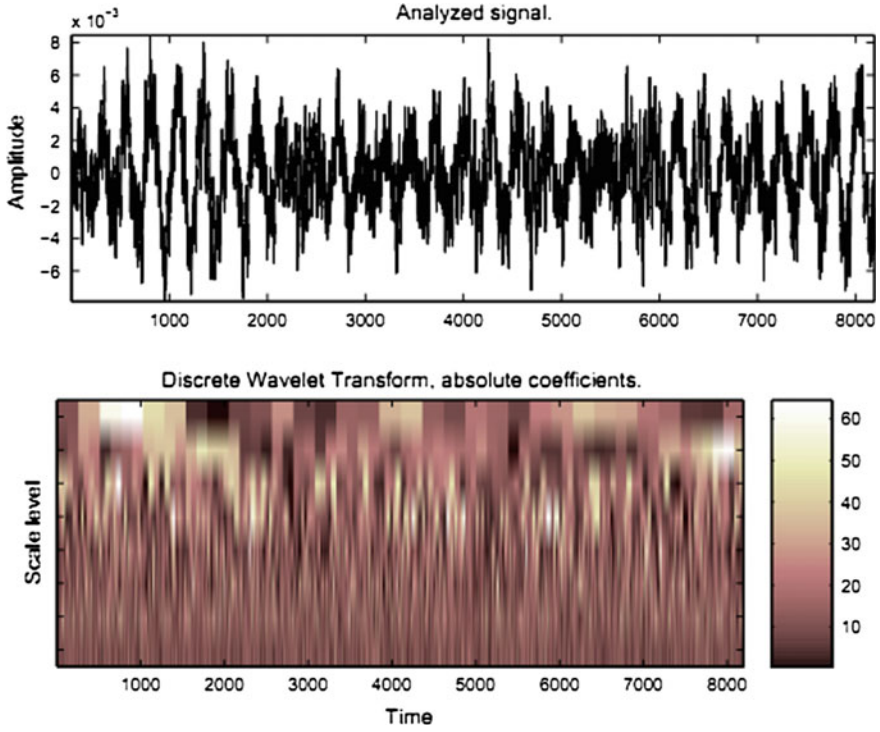


Fig. 1 Time–frequency plot of discrete wavelet coefficients

(resolution), and the color at each horizontal–vertical point represents the magnitude of the wavelet coefficient.

The main advantage of exploiting wavelets is performing local analysis of a highly localized signal, such as spikes. On the other hand, wavelets can be contained in a finite interval. Hence, they are well suited to represent or approximate signals containing discontinuities. This functionality is particularly important for fault detection using vibration signals. Moreover, due to the availability of a fast transform version, the computational effort to perform the signal processing is reduced. Because of these features, the wavelet transform has recently been considered as a powerful tool for fault detection and condition monitoring for rotating machinery (Prabhakar et al. 2002). Wavelet-based examples showing the time-scale nature of damage characteristics in vibration responses can be found in (Law et al. 2005, Li et al. 2006, Li et al. 2007).

2.2 Control Chart in Statistical Process Control (SPC)

In general, control chart problems in SPC can be divided into two phases (Woodall and Montgomery 1999; Woodall 2000). Phase I analysis tries to isolate the in-control (baseline) data from an unknown historical data set, and establish the control limits for future monitoring (Zhang and Albin 2007). Phase II analysis monitors the process using control charts derived from the "cleaned" in-control data set from the Phase I analysis. With a simple plot of the set of monitoring statistics derived from the original samples, the control chart can effectively determine whether or not a process is in an in-control state. In addition to the monitoring statistics, another important component of control charts is determining the control limits, which are often calculated based on the probabilistic distribution of the monitoring statistics (Sukchotrat et al. 2010). Common characteristics used for comparing the performance of control charts are first an in-control average run length (ARL0), and second an out-of-control average run length (ARL1). ARL0 is defined as the expected number of observations taken from an in-control process, until the control chart falsely signals out-of-control. ARL0 is regarded as acceptable, if it is large enough to keep the level of false alarms at an acceptable level. ARL1 is defined as the expected number of observations taken from an out-of-control process, until the control chart correctly signals that the process is out-of-control. Ideally, the ARL1 should be as small as possible.

Hotelling's T^2 chart is a multivariate control chart that can monitor a multivariate process efficiently (Hotelling 1947). Mason and Young give the basic steps for the implementation of multivariate SPC using the T^2 statistic, and they published a textbook on the practical development and application of multivariate control techniques using the T^2 statistic (Mason and Young 2002). Although Hotelling's T^2 chart is one of most popular multivariate control charts, it is difficult to apply this chart to the non-normal data that is often found in many applications, due to the distributional assumption of the T^2 chart. A number of nonparametric control charts have been developed to address the limitation of the distributional assumptions (Chakraborty et al. 2001; Liu et al. 2004; Bakir 2006; Kim et al. 2007; Qiu 2008). However, no consensus exists about which of them best satisfies all the conditions encountered in modern process systems. A detailed review of nonparametric control charts is beyond the scope of this study.

2.3 Nonparametric Approach for Non-normality Situations: Bootstrapping

In traditional control charts, the control limits are determined based on the underlying distribution of the monitoring statistic, with the user-specified value of type I error rate (α). Here, type I error rate represents the probability of a false alarm. In contrast, the distribution of some nonparametric monitoring statistics of new control charts is unknown, due to its nonparametric nature. This provides the motivation to develop an appropriate nonparametric procedure to establish the control limits (Chou et al. 2001; Sukchotrat et al. 2010). In particular, many studies have focused on the construction of nonparametric control charts, using a bootstrap procedure. The bootstrap, originally proposed and named by Efron (1979), is a computational technique that can be used to effectively estimate the sampling distribution of a statistic. One can use the nonparametric bootstrap to estimate the sampling distribution of a statistic, while the only assumptions are that the sample is representative of the population from which it is drawn, and that the observations are independent and identically distributed. Bajgier (1992) introduced a univariate control chart whose lower and upper control limits were estimated using the bootstrap technique, and Seppala et al. (1995) proposed a subgroup bootstrap chart, to compensate for the limitations of Bajgier's control charts. Liu and Tang (1996) proposed a bootstrap control chart that can monitor both independent and dependent observations. Note that several control charts based on the bootstrap are compared in the study from Jones and Woodall (1998). We apply the concept of the bootstrap procedure in Sect. 3.2 to construct a control limit for our test statistic, with a more explicit description.

3 The Proposed Control Chart: MWS-Based Bootstrap Control Chart

3.1 Multiscale Wavelet Scalogram (MWS)

When the ith component of discrete wavelet transform $\mathbf{d}$ in Sect. 2 is defined as $\mathbf{d}_{(i)} = (d_{(i)1}, d_{(i)2}, \ldots, d_{(i)m_i})^{\mathrm{T}}$, where m_i is the number of wavelet coefficients at the (i)th resolution level, the element of multiscale wavelet scalogram (MWS) at the (i)th resolution level is defined as $s_{(i)} = \sum_{k=1}^{m_i} d_{(i)k}^2, \quad i = 1, 2, \ldots, I$ (Vidakovic 1999, p. 289; Jeong et al. 2003) which is the sum of the squares of all wavelet coefficients at the (i)th resolution level, where $I = J - L + 2$. Then, a vector of MWS that represents a single vibration signal is defined as $S = (s_{(1)}, s_{(2)}, \ldots, s_{(I)})$, where $s_{(1)}$ is for the approximation level, and $s_{(i)}, 2 \leq i \leq I$ is for the detail levels. MWS provides measures of signal energy at various frequency bands (Rioul and Vetterli 1991). MWS represents the scale-wise distribution of the energies of a signal. Thus, a vector of MWS, $S = (s_{(1)}, s_{(2)}, \ldots, s_{(I)})$, is in very low dimension, depending on the predetermined depth of decomposition (decomposition level $= I - 1$). An adequate depth of decomposition is required, to avoid the risk of eliminating important features, and over-smoothing of the signal. For a signal of length N in an off-line mode, or a window of length N in an on-line mode, empirical evidence suggests that the depth of decomposition should be half the maximum possible depth.

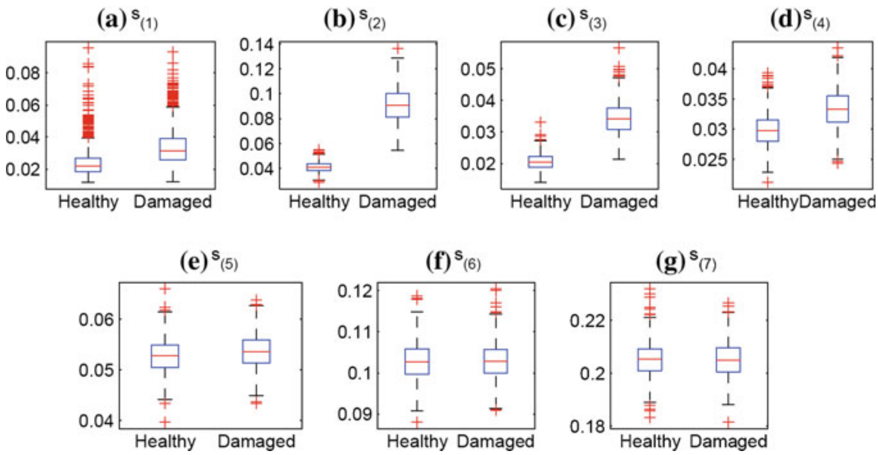


Fig. 2 Boxplot analysis based on MWS at different scales with two different classes (healthy and damaged condition)

According to Ganesan et al. (2004), for a dyadic data length $N = 2^j$, the recommended depth of decomposition is $j/2$ (i.e., $I = j/2 + 1$). Figure 2 shows the result of boxplot analysis based on MWS, for the bearing signal having N = 4096 datapoints and 6 decomposition levels ($S = (s_{(1)}, s_{(2)}, \ldots, s_{(7)})$), toward the healthy and damaged cases. Here, a thousand sets of bearing signals for both healthy and damaged cases are generated for demonstration purpose (each set has N = 4096 datapoints). As shown in the figure, the approximation level $s_{(1)}$, and relatively high frequency scales, such as $s_{(4)}$, $s_{(5)}$, $s_{(6)}$, and $s_{(7)}$, are less sensitive to the bearing defect, while low frequency scales, such as $s_{(2)}$ and $s_{(3)}$, are quite sensitive to the defect. For example, the highest frequency scale $s_{(7)}$ in Fig. 2g has almost the same interquartile ranges between the health and damaged case, while the lowest frequency scale $s_{(2)}$ and $s_{(3)}$ in Fig. 2b has no overlapping interquartile ranges. Thus, in this present study, MWS is regarded as the set of extracted features that are damage-sensitive, in terms of scale-wise energy representation of vibration signals. MWS is also, in very low dimension, suitable for constructing a multivariate control chart.

3.2 *MWS-Based Bootstrap Control Chart (MWSB Chart)*

The proposed monitoring statistic using MWS, $T^2_{(S)}$, measures the distance between an observed MWS vector of the vibration signal, and the scaled-mean MWS vector, estimated from the in-control data. The monitoring statistic for the ith individual vibration signal observation can be calculated by the following equation, which is similar to Hotelling's T^2 statistic: $T^2_{(S)i} = (S^{(i)} - \bar{S})^T \boldsymbol{\Sigma}_S^{-1}(S^{(i)} - \bar{S})$, where $S^{(i)}$ is a vector of MWS for the ith observation, and $\bar{S}$ and $\boldsymbol{\Sigma}_S$ are a sample mean vector and a sample covariance matrix of those vectors of MWS for vibration signals from an in-control process, respectively. Then, the procedure for constructing our multiscale wavelet scalogram-based control chart via bootstrap (MWSB chart) is summarized as follows, in a similar manner used in Phaladiganon et al. (2011).

- Compute the $T^2_{(S)i}$ statistic for $i = 1, \ldots, m$ with m observations from an in-control dataset.
- Let $T^{2(b)}_{(S)1}, T^{2(b)}_{(S)2}, \ldots, T^{2(b)}_{(S)m}$ be a set of m $T^2_{(S)i}$ values from the bth bootstrap sample ($b = 1, \ldots, B$) randomly drawn from the initial $T^2_{(S)i}$ statistics, with replacement. In general, B is a large number (e.g., $B > 5000$).
- In each of B bootstrap samples, determine the $100 \cdot (1 - \alpha)$th percentile value, given a user-specified value α, with a range between 0 and 1.
- Determine the control limit, by taking the average of B $100 \cdot (1 - \alpha)$th percentile value ($\bar{T}^2_{(S),100\cdot(1-\alpha)}$). Note that statistics other than the average can be used (e.g., the median).
- Use the established control limit to monitor a new observation. That is, if the monitoring statistic of a new observation exceeds ($\bar{T}^2_{(S),100\cdot(1-\alpha)}$), we declare that specific observation as out-of control.

Determination of the appropriate number of bootstrap samples to use is not obvious. However, with reasonably large numbers of bootstrap samples, the variation of results is insignificant. The computational time required has been perceived as one of the disadvantages of the bootstrap technique, but this is no longer a critical issue, because of the computing power currently available.

4 Illustrative Example: Bearing Condition Monitoring with Localized Defect

The bearing system in most rotating machinery needs a reliable measure for diagnosing its functionality, to avoid the risk of catastrophic accident, which may cause personal damage and economic loss. The most successful method for condition monitoring of a bearing system is using vibration measurements. The easiest way of measuring the vibration is mounting accelerometers on the surface of the outer frame of the bearing, and instrumenting them to the data acquisition system, for further signal processing. Instead of using experimental data, many studies exploit bearing simulation schemes, in which we can arbitrary control the severity and collision frequency of defects. Previously, some researchers proposed bearing simulation methods, by producing series of repeated periodic oscillation with proportional damping terms (McFadden and Smith 1984; Cong et al. 2010). Again, the frequency of impulse is related to the collision frequency between the defect and rotating elements. The limitation of the simulation approach is that it only represents a single frequency component. It is reasonable to believe that a bearing system has its own dynamics, and many modes of broad spectrum. Thus, we used a state-space model to represent the dynamics of the overall bearing system. Specifically, the state-space model consists of two-degree-of-freedom system, i.e., two masses connected by springs and dampers. Each component of spring, mass, and damper simulates the overall dynamics of inner-race, outer-race, rolling elements, and housing structure. The effect of bearing fault is simulated by inflicting periodic impulses to the state-space system.

It is reasonable to consider that the severity of fault in a bearing is proportional to the power of periodic impulse caused by collision impact between the defect and rolling elements. We quantified the level of fault severity, denoted SL (Severity Level), as the relative intensity of impact force on the state-space model, having a scale from 0 to 1. Here, 0 represents no fault or pristine condition, while 1.0 indicates the condition from the most severe defect. In this study, we used five levels of fault severity, SL = 0.05, 0.1, 0.15, 0.2, and 0.25. Thus, the lower the SL value, the more difficult it is to differentiate the incipient defect from measurement noise or background resonances. The time-history data simulated from the model of the bearing system for the healthy case is shown in Fig. 3a, and for the bearing having defect is depicted in Fig. 3b, respectively. It is worth noting that the severity of a bearing defect reflected in terms of signal power can be easily masked by environmental noise, making its detection extremely difficult in practice. Thus,

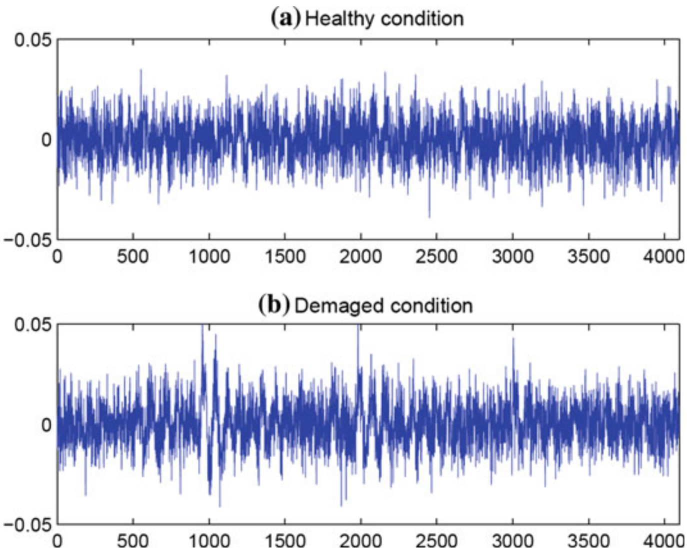


Fig. 3 Examples of time-history response from sensor for bearing; **a** healthy condition; **b** damaged condition with SL = 0.2

damage-sensitive features extracted by signal processing must be further interpreted, and evaluated for fault detection. Again, feature extraction is the process of identifying damage-sensitive properties derived from the measured vibration that allows one to differentiate the incipient fault from a healthy bearing. The number of data points has been fixed to 2^{12} = 4096 points, which is large enough to include at least two defect-driven impulses during the rotation of the bearing.

Once it is revealed that the extracted features are adequately sensitive to structural defects, the most direct way to deal with a vibration signal in SPC is treating features extracted from the vibration signal as a realization of a multivariate process. However, the use of standard multivariate SPC charts is ill-advised, because when the number of monitored features exceeds the number of collected samples, the covariance matrix is singular, and the common multivariate statistics cannot be calculated. Also, even though the number of monitored features is less than the number of collected samples, it is well known that the power of detecting process fault or structural damage will drop significantly, when the number of features becomes large (Fan 1996). This implies that the rate of false alarm will be significantly increased. Functional principal component analysis (FPCA; Ramsay and Silverman 1997) and related procedures (Hall et al. 2001) are useful as a dimension reduction tool, in modeling high-dimensional data. However, there are some drawbacks in applying the FPCA approach to SPC procedure. For instance, the FPCA method presents difficulties of interpreting the relationship between changes in a few principal components and the high-dimensional signal. To overcome the limitation posed by high dimensionality, Sect. 3 in this study proposed a wavelet-based multivariate control chart method, for fault detection of a bearing system.

5 MWSB Chart Performance Evaluation: ARL0 and ARL1

In order to use MWS features for bearing condition monitoring, we need to first decide which wavelet basis function type should be used. Several wavelet basis function types are available in the literature. Some of these are the Haar's, Daubechies', coiflets, symlets, and bi-orthogonal wavelets. For a detailed description on the mathematical properties of the wavelet basis function, refer to Goswami and Chan (1999). In this study, we used Daubechies' wavelet type (db5), with depth of decomposition 6 (i.e., $N = 4096$, $\log_2 N/2 = 6$). We generated two groups of 1000 sets of signals from healthy condition (in-control observations). The first group of 1000 in-control sets of signal was used to determine the control limits of the $T^2_{(s)}$ statistic, from the F-distribution and bootstrap approach. The second group of 1000 in-control sets of signal was monitored on the control chart, which were based on the control limits established by the first group of in-control sets of signal. The control chart that produces a similar value for both the actual FAR and the nominal FAR would be considered the better one. To check multivariate normality of the in-control observations, we conducted a Royston's H test (Royston 1983). The p-value, which measures the plausibility that the dataset follows a multivariate normal distribution, was almost zero (i.e., Royston's statistic: 188.067, degrees of freedom: 7.000048, p-value: 0.0000). This strongly indicates that this dataset does not come from a multivariate normal distribution.

As mentioned in earlier section, ARL0 is defined as the average number of observations required for the control chart to detect a change under the in-control process of Phase I. To compare the competing control charts, we need to determine the better one, corresponding to the higher value of ARL0. Given a type I error rate (α), which is assumed as equal to 0.01, 0.05, and 0.1, the two competing approaches (using F-distribution and the bootstrap approach) can be compared in terms of ARL0 (in-control ARL). In this study, the ARL0 was calculated using bootstrap techniques. Starting from a group of 1000 in-control sets of signal as a sample, we simulated 5000 bootstrap samples, where each sample is selected randomly, with replacement, and with the same size 1000 as the original sample. For each bootstrap sample, the run length (RL) to detect a change under the healthy condition was calculated. Figure 4a shows the nominal ARL0 values and the actual ARL0 values as obtained by F-distribution and the bootstrap approach, for different type I error rates α. This figure shows that across the different approaches, the actual ARL0 values are close to the nominal ARL0 values, when the type I error rate α is equal to 0.05 and 0.1. However, for the type I error rate $\alpha = 0.01$, the difference between the actual ARL0, as determined by the F-distribution and the nominal ARL0, increases. In contrast, the bootstrap approach generated similar actual and nominal ARL0 results. This suggests that choosing the bootstrap approach, instead of F-distribution, will reduce the false alarm rate.

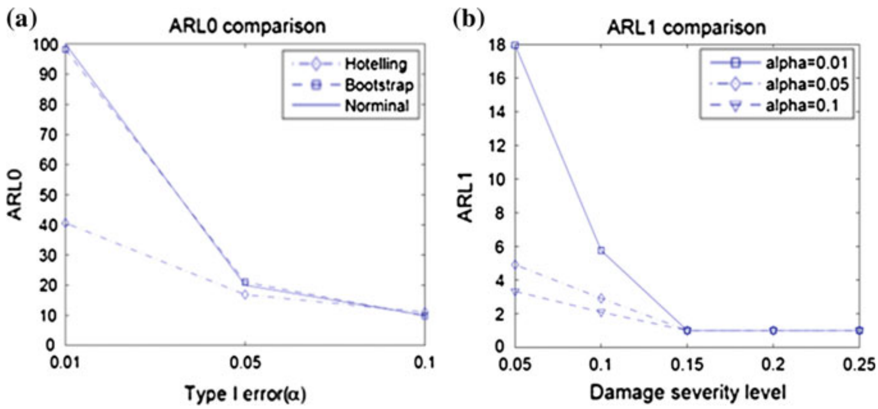


Fig. 4 MWSB chart performance evaluation; **a** ARL0 comparison between F-distribution and the bootstrap approach, for different type I error rates; **b** ARL1 comparison among several type I error rates, for different damage severity levels

The objective in Phase II of control charting is to quickly detect any change in the process from its in-control state. In this phase of control charting, the competing methods are compared in terms of ARL1, where the run length is defined as the number of samples taken, until an out-of-control signal is issued. In order to evaluate the performance of control charts in this operating phase, the occurrence of assignable causes (faults) was simulated by a total of five out-of-control conditions. Each condition is then characterized by a damage severity level parameter (SL), directly proportional to the damage severity of the out-of-control (i.e., $SL = 0.05, 0.1, 0.15, 0.2, 0.25$) introduced in the baseline model (i.e., SL = 0 for the in-control state). Figure 4b summarizes the simulation results under study. This figure reports the ARL1 performance estimated by computing a set of 5000 bootstrap run lengths for α = 0.01, 0.05, and 0.1, in a similar manner to the case of ARL0 calculation, given that the new sets of signal are simulated according to a specific out-of-control model (i.e., $SL = 0.05, 0.1, 0.15, 0.2, 0.25$). In order to select the better approach in each different study (where each case study is characterized by a specific out-of-control), we need to determine the better method, corresponding to the lower value of ARL1. Figure 4b shows that the actual ARL1 values with larger α are smaller, as one would expect. More importantly, as SL increases, the actual ARL1 decreases, until SL becomes 0.15. This means that as damage becomes more severe, it is more quickly detected. If the damage severity level is greater than or equal to 0.15, the damages are all detected in average at the first observation, across all different type I errors. This result indicates that the proposed control chart is very sensitive to the severity of incipient fault in a bearing system.

6 Conclusion and Future Work

This study introduced a new feature extraction and dimension reduction technique, to alleviate the high dimensionality problem of implementing multivariate SPC, when the quality characteristic is a vibration signal from a bearing system. A set of multiscale wavelet scalogram features was generated to reduce the dimensionality of data, and the bootstrapping technique applied as nonparametric density estimation, to set up an upper control limit of the control chart. Our example and numerical simulation of a bearing system demonstrated that the proposed method has satisfactory fault-discriminating ability, without any distributional assumption.

The proposed bearing condition monitoring approach has several desirable attributes. First, the fault detection is conducted in an unsupervised learning mode. That is, data from the damaged system are not needed. The ability to perform the fault detection in an unsupervised learning mode is very important, because data from damaged bearing are typically not available for most real-world practices. Second, the approach presented herein is very attractive for the development of an automated continuous monitoring system because of its simplicity, and because it requires minimal interaction with users. However, it should be pointed out that the procedure developed has only been verified on relatively simple numerical simulations, using a bearing model. To verify that the proposed method is truly robust, it will be necessary to examine many time records, corresponding to a wide range of operational and environmental cases, a wide range of damaged and pristine condition, as well as different damage scenarios.

In this research, we emphasized a damage severity-change problem, but this can be easily extended to damage classification problems. The other direction that this research could be extended to is to estimate the damage severity level from vibration signals using nonparametric regression techniques, such as support vector regression. Finally, there are several parameters still to be determined, such as the wavelet type, the level of decomposition, and the number of bootstrap samples for density estimation, before the proposed approach can be used effectively. Some theoretical guidelines are needed in future research, to select 'optimal' sets of parameters.

References

Bakir, S. (2006): *Distribution-free quality control charts based on signed rank-like statistics*. Communications in Statistics: Theory and Methods, 35, pp. 743–757.

Bajgier, S.M. (1992): *The use of bootstrapping to construct limits on control charts*. Proc. Decision Science Institute, San Diego, CA, pp. 1611–1613.

Chakraborti, S., Van der Laan, P. and Bakir, S. T. (2001): *Nonparametric control chart: an overview and some results*. Journal of Quality Technology, 33(3), pp. 304–315.

Chou, Y.-M., Mason, R. L., Young, J. C. (2001): "The control chart for individual observations from a multivariate non-normal distribution", *Communications in Statistics-Simulation and Computation*, 30(8–9), pp. 1937–1949.

Cong, F., Chen, J., and Dong, G. (2010): "Research on the order selection of the autoregressive modeling for rolling bearing diagnosis", Proceedings of the Institution of Mechanical Engineers, Part C: *Journal of Mechanical Engineering Science*, 224(10), pp. 2289–2297.

Dyer, D. and Stewart, R. M., (1978): "Detection of rolling element bearing damage by statistical vibration analysis", *Journal of Vibration and Acoustics*: Transactions of the ASME, 100(2), pp. 229–235.

Efron, B. (1979): *Bootstrap method: another look at jackknife*. Annals of Statistics, 7(1), pp. 1–26.

Fan, J. (1996): "Test of significance based on wavelet thresholding and Neyman's truncation", *Journal of American Statistical Association*, 91, pp. 674–688.

Ganesan, R., Das, T. K., and Venkataraman, V. (2004): "Wavelet-based multiscale statistical process monitoring: A literature review", *IIE Transactions*, 36, pp. 787–806.

Goswami, J. C. and Chan, A. K. (1999): *Fundamentals of Wavelets: Theory, Algorithms, and Applications*. Wiley, New York, NY.

Hall, P., Poskitt, D. S., and Presnell, D. (2001): "Functional data-analytic approach to signal discrimination", *Technometrics*, 43(1), pp. 1–9.

Hotelling, H. (1947): *Multivariate quality control, in Techniques of Statistical Analysis*. Eisenhart, C., Hastay, M. W. and Wills, W. A. (eds), McGraw-Hill, New York, NY, pp. 111–184.

Jardine, A. K. S., Lin D., Banjevic, D. (2006): "A review on machinery diagnostics and prognostics implementing condition-based maintenance", *Mechanical Systems and Signal Processing*, 20(7), pp. 1483–1510.

Jeong, M. K., Chen, D., Lu, J. C. (2003): "Thresholoded scalogram and its applications in process fault detection", *Applied Stochastic Models in Business and Industry*, 19, pp. 231–244.

Jones, L. A. and Woodall, W. H. (1998): “The performance of bootstrap control charts”, *Journal of Quality Technology*, 30(4), pp. 362–375.

Jung, U., Jeong, M. K. and Lu, J. C. (2006): “A vertical-energy-thresholding procedure for data reduction with multiple complex curves”, *IEEE Transactions on Systems, Man, and Cybernetics-Part B*, 36(5), pp. 1128–1138.

Jung, U. and Koh, B. H. (2009): “Structural damage localization using wavelet-based silhouette statistics”, *Journal of Sound and Vibration*, 321, pp. 590–604.

Kim, S. H., Alexopoulos, C., Tsui, K. L. and Wilson, J. R. (2007): “A distribution-free tabular CUSUM chart for autocorrelated data”, *IIE Transactions*, 39(3), pp. 317–330.

Koh, B. H., Nagarajaiah, S., Phan, M. Q. (2008): “Reconstructing structural changes in a dynamic system from experimentally identified state-space models”, *Journal of Mechanical Science and Technology*, 22(1), pp. 103–112.

Law, S. S., Li, X. Y., Zhu, X. Q., and Chan, S. L. (2005): “Structural damage detection from wavelet packet sensitivity”, *Engineering Structures*, 27, pp. 1339–1348.

Lei, Y., He, Z., and Zi, Y. (2011): “EEMD method and WNN for fault diagnosis of locomotive roller bearings”, *Expert Systems with Applications*, 38(6), pp. 7334–7341.

Li, H., Deng, X., and Dai, H. (2007): “Structural damage detection using the combination method of EMD and wavelet analysis”, *Mechanical Systems and Signal Processing*, 21, pp. 298–306.

Li, Z., Xia, S., Wang, J. and Su, X. (2006): “Damage detection of cracked beams based on wavelet transform”, *International Journal of Impact Engineering*, 32, pp. 1190–1200.

Lin, J. and Zhang, A. (2005): “Fault feature separation using wavelet-ICA filter”, *NDT&E International*, 38(6), pp. 421–427.

Lio, Y. L. and Park, C. (2008): “A bootstrap control chart for Birnbaum-Saunders percentiles”, *Quality and Reliability Engineering International*, 24, pp. 585–600.

Liu, R. Y. and Tang, J. (1996): “Control charts for dependent and independent measurements based on bootstrap methods”, *Journal of the American Statistical Association*, 91, pp. 1694–1700.

Liu, R. Y., Singh, K. and Teng, J. H. (2004): “DDMA-charts: nonparametric multivariate moving average control charts based on data depth”, *Allgemeines Statistisches Archiv*, 88(2), pp. 235–258.

Mallat, S. G. (1989): *A Wavelet Tour of Signal Processing*. Academic Press, San Diego.

McFadden, P.D., and Smith, J. D. (1984): “Model for the vibration produced by a single point defect in a rolling element bearing”, *Journal of Sound and Vibration*, 96(1), pp. 69–82.

Mason, R. L., Young, J. C. (2002): *Multivariate Statistical Process Control with Industrial Applications*. ASA/SIAM: Philadelphia, PA.

Palacz, M., and Krawczuk, M. (2002), Vibration parameters for damage detection in structures, Journal of Sound and Vibration, Vol.249, No.5, 999–1010.

Peng, Z., Chu, F., and He, Y. (2002): “Vibration signal analysis and feature extraction based on reassigned wavelet scalogram”, *Journal of Sound and Vibration*, 253(5), pp. 1087–1100.

Phaladiganon, P., Kim, S. B., Chen, V. C. P., Baek, J. G. and Park, S. K. (2011): “Bootstrap-based T^2 multivariate control charts”, *Communications in Statistics-Simulation and Computation*, 40, pp. 645–662.

Polansky, A. M. (2005): “A general framework for constructing control charts”, Quality and Reliability Engineering International, 21, pp. 633–653.

Prabhakar, S., Mohanty, A. R., Sekhar, A. S. (2002): “Application of discrete wavelet transform for detection of ball bearing race faults”, *Tribology International*, 35(12), pp. 793–800.

Qiu, P. (2008): “Distribution-free multivariate process control based on log-linear modeling”, *IIE Transactions*, 40(7), pp. 664–677.

Ramsay, J. O., and Silverman, B. W. (1997): “Functional linear models for scalar responses.” Functional Data Analysis. Springer New York, pp. 157–177.

Randall, R. B. and Antoni, J. (2011): “Rolling element bearing diagnostics–A tutorial”, *Mechanical Systems and Signal Processing*, 25(2), pp. 485–520.

Rioul, O. and Vetterli, M. (1991): “Wavelets and signal processing”, *IEEE Signal Processing Magazine*, pp. 14–38.

Royston, J. P. (1983): “Sone techniques for assessing multivariate normality based on the Shapiro-Wilk W”, *Applied Statistics*, 32(2), pp. 121–133.

Scargle, J. D. (1997): “Wavelet methods in astronomical time series analysis”, Application of Time Series Analysis in Astronomy and Meteorology, Chapman & Hall, New York, pp. 226–248.

Seppala, T., Moskowitz, H., Plante, R., and Tang, J. (1995): “Statistical process control via the subgroup bootstrap”, *Journal of Quality Technology*, 27, pp. 139–153.

Sohn, H., Farrar, C.R., Hemez, F.M., Shunk, D.D., Stinemates, D.W., and Nadler, B.R. (2003), A review of structural health monitoring literature: 1996 ~ 2001, Technical Reports LA-13976-MS, Los Alamos National Laboratory.

Sukchotrat, T., Kim, S. B., and Tsung, F. (2010): “One-class classification-based control chart for multivariate process monitoring”, *IIE Transactions*, 42, pp. 107–120.

Swamidas, A.S.J. and Chen, Y. (1995), Monitoring crack growth through change of modal parameters, Journal of Sound and Vibration, Vol.186, No.2, 325–343.

Yu, Y., YuDejie, Junsheng, C. (2006): “A roller bearing fault diagnosis method based on EMD energy entropy and ANN”, *Journal of Sound and Vibration*, 294(1–2), pp. 269–277.

Vidakovic, B. (1999): *Statistical Modeling by Wavelets*. John Wiley & Sons.

Wang W. and Wong A. K., (2002): “Autoregressive model-based gear fault diagnosis”, *Journal of Vibration and Acoustics: Transactions of the ASME*, 124, pp. 172–179.

Woodall, W. H. (2000): “Controversies and contradictions in statistical process control”, *Journal of Quality Technology*, 32(4), pp. 341–350.

Woodall, W. H. and Montgomery, D. C. (1999): “Research issues and ideas in statistical process control”, *Journal of Quality Technology*, 31(4), pp. 376–386.

Zhang, H. and Albin, S. (2007): “Determining the number of operational modes in baseline multivariate SPC data.” IIE Transactions 39(12) pp. 1103–1110.

Part II
Six Sigma Practices

Lip Hong Teo, Council Member of Singapore Quality Institute

In an ever-changing world, every company aspires to stay competitive and generate greater profits and revenue. Many companies have Operational Excellence with initiatives such as Lean, Six Sigma, Lean Six Sigma, etc., to improve their quality, delivery times, and prices which correspond to their reduction in defects, cycle times, and cost of operation.

Many companies assure that their companies are efficient as they have fewer headcounts in operation. What does a process actually mean? In layman terms, a process refers to “the way you manufacture a product” or “the way one delivers a service.” What, then, is the definition of an effective and efficient process?

Effectiveness (Lean) means “Doing the right things (customers’ values) right,” to eliminate non-valued activities (wastes). On the other hand, Efficiency (Six Sigma) means “Doing the right things (customers’ values) right at all times,” to reduce variations—which are defects.

To establish Effectiveness and Efficiency in companies, many implemented Lean, Six Sigma, and Lean Six Sigma. Most of them claim that Six Sigma is more prominent in manufacturing companies, whereas Lean is more applicable to service industries. I would say that there is no “one-size fits all” structure; Lean and Six Sigma complement each other. It is of utmost importance to understand the principles, concepts and applications of the appropriate tools and techniques in the processes.

Lean Enterprise embraces Lean principles and Lean thinking that matches the Deming cycle of Plan, Do, Check and Act (PDCA) to carry out Kaizen improvement projects. Lean aims to eliminate wastes, specifying “Customers’ Values,” defining “Values Stream,” establishing “Continuous Flow” to prevent unnecessary idling and stoppage. It also looks into overproduction and inventory through the lenses of the “Pull” system. Ultimately, “Pursuing Perfection” is to have continuous improvement over the horizon.

Lean speeds up the lead time in many service lines like the on-the-dot delivery of pizza, office stationeries, etc., to customers, market in Information Technology operations for business and end-user satisfaction. The eliminated waste in manufacturing like food industrials improves the cycle time of the process. Furthermore, Lean not only shortens the lead time, but also gives a velocity (direction) of how it is heading toward the values perceived by the customers. Lean principles encompass Lean tools like Visual management to make problems visible, Kanban and Just-in-Time to help reduce overproduction and inventory, Total productive Maintenance to enhance the Overall equipment effectiveness for a more productive equipment, etc. All these enable the Lean Principles and concept to move forward.

Six Sigma's methodology includes Define, Measure, Analyse, Improve and Control (DMAIC) for systematic executions of project. Six Sigma—3.4 Defect Per Million Opportunities (DPMO) reduces variation (defects) in the process and establishes the Sigma level of a company. Six Sigma stresses the importance of statistical hypotheses to analyse data. Eventually, these decisions translate into more practical implementations in your respective companies.

Companies use Six Sigma to minimise the usage of raw material, thus scaling down the operating costs. Universities use DMAIC methodology to improve the quality standard through the reduction the variations.

Six Sigma tools like Gage Repeatability and Reproducibility, Process Capability Study, Design of Experiment, Statistical Process Control, etc., are excellent in reducing variations.

Integrating both the Effectiveness (eliminate wastes) and Efficiency (reduce variation) approaches—we have Lean Six Sigma. It is not a simple addition; it is the complementation and organic combination of both. With such integration, Lean Six Sigma demonstrates the values specified by the customers through the reduction of the complexity in the process and improves the lead time, delivery time and time to market, whereas Six Sigma improves the quality.

This is well demonstrated in the Anderson leadership team—a new approach leading to continuous improvements in 2014. I believe that with Lean Six Sigma, the integrated result, companies with these initiatives would have a more competitive edge in many aspects like quality, delivery, and price.

About the author

Mr. Teo Lip Hong graduated with a Master of Science in Strategic Quality Management from the University of Portsmouth in United Kingdom, and his research enriches him with practical understanding of Six Sigma goal, philosophy and breakthrough strategy for Six Sigma.

He is a certified Master Black Belt and ASQ Lean Six Sigma trainer. Having more than 10 years' experience as a Lean Six Sigma Practitioner, he has been conducting Six Sigma and Lean Thinking training programmes in the service and manufacturing sectors. He is also a qualified lecturer in Singapore Quality Institute (SQI) training certified courses in Six Sigma and Lean Six Sigma. He also mentors employees and SQI's students regarding the implementation of Six Sigma DMAIC methodology and Lean principles in their projects.

He welcomes experience sharing with an open heart and is contactable at +65-91174028 and/or teoliphong@gmail.com.

From Belts to Baldrige

Charles Aubrey

Abstract The Anderson leadership team "discovered" a new formal approach to continuous improvement in 2004—Lean Six Sigma (LSS). Senior management integrated the two approaches into Lean Six Sigma—attacking both the product and process variation and waste. Our processes are supported with over 100 engineers, mechanics, and technicians. A significant training effort allowed all of them to be Lean Certified. They have saved over $10 M annually in improving equipment and setups alone. Anderson has developed 18 of 200 Green Belts (our best) into Black Belts and six of them into Master Black Belts. They saved almost over $7 M annually. The Steering Committee has committed everyone in every department to participate. To that end our projects are focused in every area of the company: Production 44%, Quality Control/Assurance 26%, Maintenance 8%, Warehouse 8%, Sales/Administrative/Finance 8%, Engineering 4%, and Human Resources 2%. The results are cost reduction, improved customer service and measurable improvement in quality, employee and customer satisfaction resulting in over $15 M annual savings plus cost avoidance, and increased capacity. In addition, over 25% of our projects improve our regulation and compliance. As a result of these efforts in 2011 Anderson won the IQPC's "Best Manufacturing Project Award." In 2012 Anderson won the IQPC's "Best Manufacturing Continuous Improvement Program Award." In 2013 they won the "State of Illinois Malcolm Baldrige Award."

C. Aubrey (✉)
Asia Pacific Quality Organization, 10186 Evergreen Dr.,
Grand Junction, MI 49056, USA
e-mail: AubreyPartners@gmail.com

C. Aubrey
International Academy for Quality, 10186 Evergreen Dr.,
Grand Junction, MI 49056, USA

C. Aubrey
Aubrey Partners, 10186 Evergreen Dr.,
Grand Junction, MI 49056, USA

C. Aubrey
Anderson Packaging, 10186 Evergreen Dr.,
Grand Junction, MI 49056, USA

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_5

Keywords Performance · Baldrige · Teamwork · Improvement · Excellence

1 Introduction

Anderson Packaging has always been committed to quality by meeting and exceeding customer requirements and continuous improvement. Founded in 1967, the Anderson Brothers manufactured packaging machines. They continued to refine and develop the equipment for their customers. They were, however, unhappy with how their customers operated the equipment. The quality of the packaging was not up to the standard they thought it should be for the ultimate customer. Therefore, in 1968, they began operating the packaging machines themselves in order to get the most effective packaging output for the final customer—the consumers. There has always been this desire to be "the best." Continuous improvement efforts and programs have always been the direction and way the Anderson leadership team has focused and led the organization. Today the vision of the company is

> To be 'Best in Class' in Pharmaceutical/Healthcare contract manufacturing and packing services. Our culture will be customer focused, embracing regulatory compliance, and continuous improvement. By utilizing operational excellence, state of the art technology, innovation, and training, we will ensure strong growth and profitability.

The vision is driven to realization by our daily mantra and mission:

> To provide:
>
> Our customers with an on-going competitive advantage.
>
> Our employees with a challenging and rewarding work experience.
>
> And our shareholders with an attractive Return on Investment.

We will accomplish this mission by providing innovative manufacturing and packaging services focused on the pharmaceutical and medical device markets. These services will encompass state-of-the-art manufacturing/packaging equipment and facilities, highly trained cGMP employees, unrivaled customer service levels, and a constant, driving continuous improvement effort (Fig. 1).

The Anderson leadership team "discovered" a new formal approach to continuous improvement in 2004. They began a drive to learn the approach of Lean Six Sigma. In that year they hired a Master Black Belt with extensive experience of implementing this approach in numerous Fortune 500 companies and the pharmaceutical and medical device industries. Senior management started their Lean Six Sigma journey by committing themselves to a formal intensive multi-day training program. They started with Lean Manufacturing and then Six Sigma Improvement. This resulted in integrating the two approaches into Lean Six Sigma (LSS)—attacking both the product and process variation and waste.

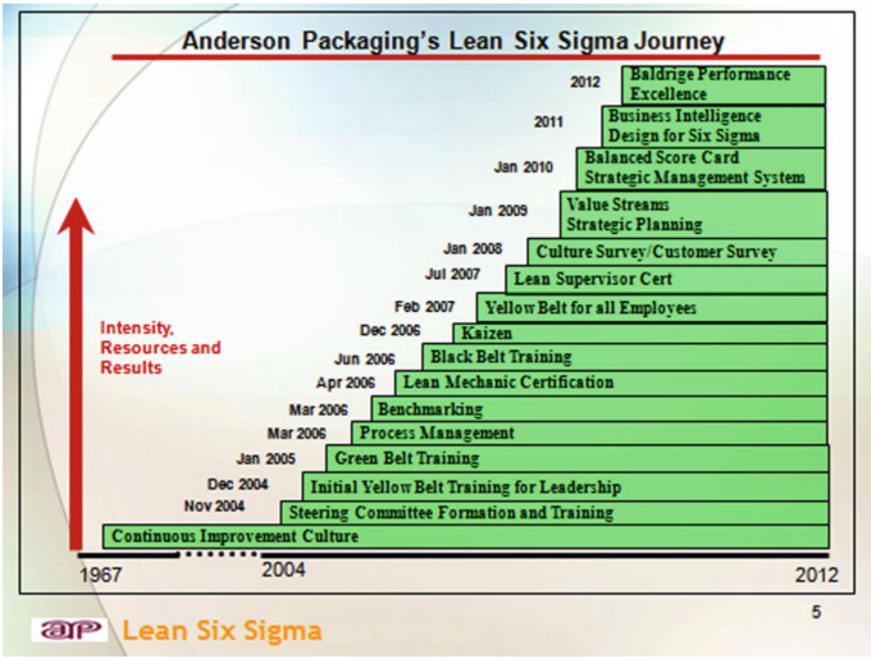


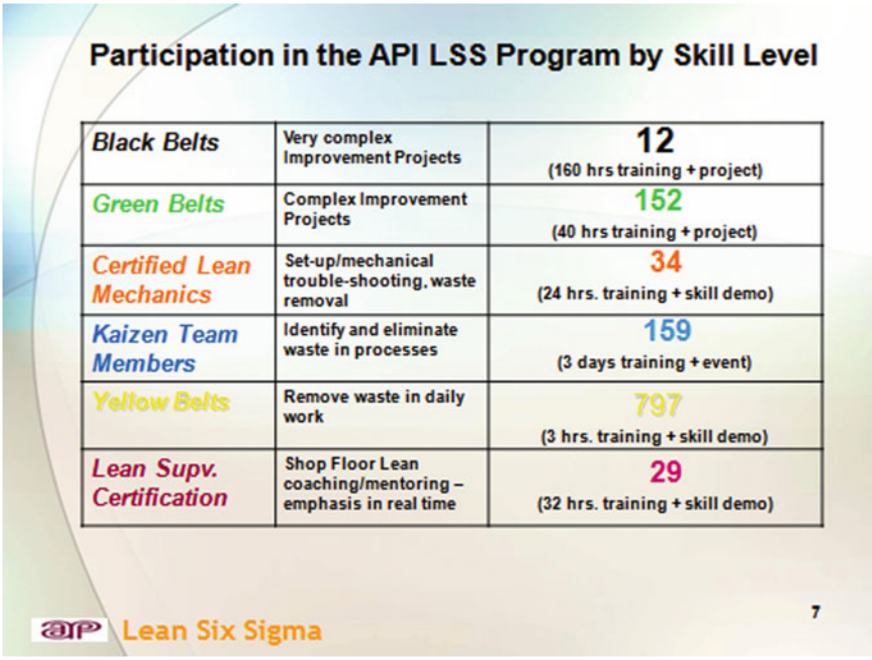
Fig. 1 Lean Six Sigma journey

2 Case Study

The strategic plan was developed integrating LSS into both the long-term and an annual business activities and goals with a serious financial and leadership commitment. Middle management all went through a multi-day Yellow Belt Training. Subsequently, the first round of projects and Green Belts were chosen with their five-day training delivered just-in-time as the projects unfolded. Subsequently, process management (PM) training was developed and a PM teams began documenting, simplifying, and eliminating waste in business processes.

A drive to improve on the part of our quality control group led us to initiate Benchmarking, with the desire to have the most efficient and effective quality control processes in the industry. Benchmarking project teams were trained and they are interacting with best-in-class Fortune 500 companies inside and outside our industry.

Our processes are supported with over 100 mechanics and engineers. A significant effort was put in for all of our mechanics and engineers to be Lean Certified. They have multiple days of class room training with tests and demonstration projects to assure skill acquisition and be certified. They have saved $500 K in improving machine and equipment setups alone. The have also achieved 100s of thousands of dollars with continuous improvement projects to make operations more efficient and effective (Fig. 2).



Participation in the API LSS Program by Skill Level

Black Belts	Very complex Improvement Projects	12 (160 hrs training + project)
Green Belts	Complex Improvement Projects	152 (40 hrs training + project)
Certified Lean Mechanics	Set-up/mechanical trouble-shooting, waste removal	34 (24 hrs. training + skill demo)
Kaizen Team Members	Identify and eliminate waste in processes	159 (3 days training + event)
Yellow Belts	Remove waste in daily work	797 (3 hrs. training + skill demo)
Lean Supv. Certification	Shop Floor Lean coaching/mentoring – emphasis in real time	29 (32 hrs. training + skill demo)

Fig. 2 Lean Six Sigma skill levels

Anderson has developed six Green Belts into Black Belts by having them acquire additional training and skills. They lead project teams to solve more difficult and sophisticated problems and implementation actions. In addition they assist in coaching Green Belt teams and deliver LSS training. Kaizen events are led by Lean Certified experts and Black Belts. These waste reduction efforts have saved almost $500 K in less than a year of operation.

The steering committee has committed to every employee who does not already have a "Belt" to be Yellow Belt trained and certified. We are well underway to completing this continuous goal. The expectation is that everyone in the organization is either identifying waste and improvement opportunities or eliminating waste and making improvements or both!

To that end our projects are focused in every area of the company: 44% production, 26% quality control/assurance, 8% maintenance, 8% warehouse, 8% sales/administrative/finance, 3% engineering, and 1% human resources. There are waste and improvement opportunities in every function or the organization (Fig. 3).

The resulting areas of impact likewise are numerous: Cost/capacity, customer service, regulatory compliance, employee satisfaction, and community (Fig. 4).

Listed below is a sample of the projects which Green Belts and Black Belts pursue (Fig. 5).

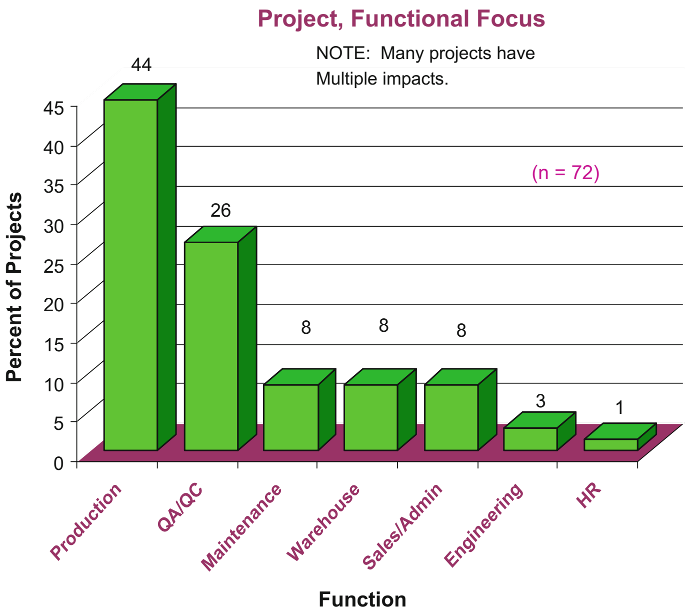


Fig. 3 Functional project areas

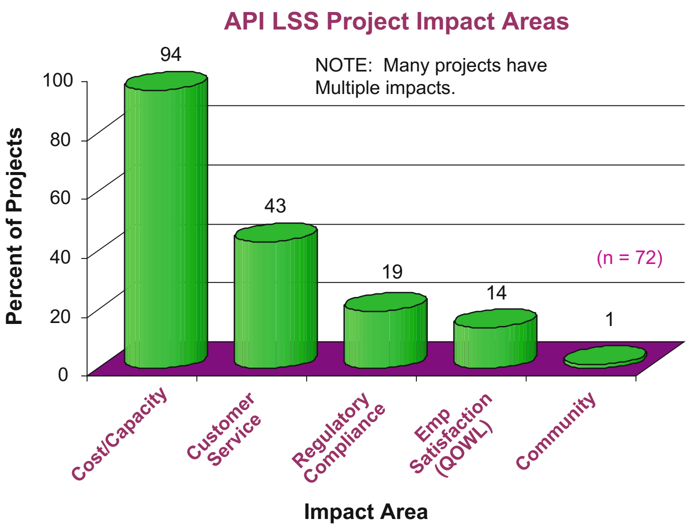


Fig. 4 Project impact areas

Anderson Packaging's Lean Six Sigma Projects

Production Projects
- Scrap Reduction
- Downtime Reduction
- Productivity, Cycle Time
- Changeover Reduction

QA/QC Projects
- Reduce Inspection Costs
- Improve Validation Cycle Time
- Reduce Documentation Errors
- Reduce Lab Costs

Maintenance Projects
- SMED
- Reduce Welding Burns
- Reduce Nominal OT

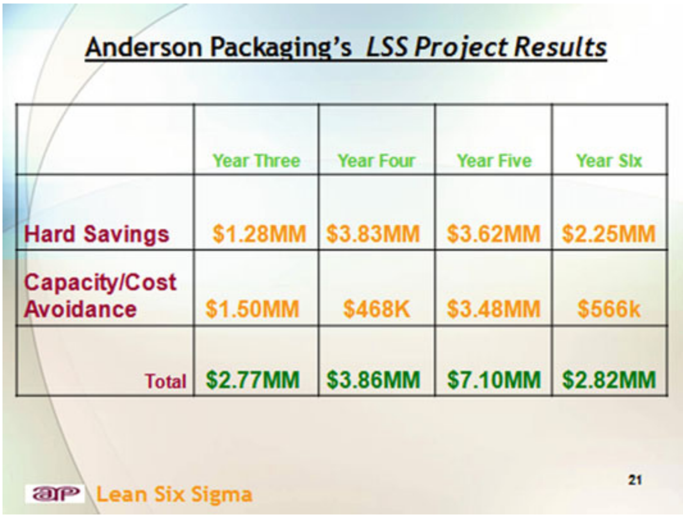
Warehouse Projects
- Improve Material Issue Cycle
- Reduce Inventory Variance
- Reduce Receiving Cycle

Sales/Admin/Finance Projects
- Improve Supply and MRO Ordering Processes
- Reduce Investigation Cycle Time
- Reduce Non-Conforming Material
- Revise Non-Routine Billing

Engineering Projects
- Reduce Tooling Costs
- Improve Tooling Cycle Time, Delivered Performance

HR/Community Projects
- Reduce Turnover
- Improve the Customer Project Management Process

Fig. 5 Sample projects



Anderson Packaging's *LSS Project Results*

	Year Three	Year Four	Year Five	Year Six
Hard Savings	$1.28MM	$3.83MM	$3.62MM	$2.25MM
Capacity/Cost Avoidance	$1.50MM	$468K	$3.48MM	$566k
Total	$2.77MM	$3.86MM	$7.10MM	$2.82MM

Fig. 6 Project results

3 Conclusion

The financial results of these LSS project efforts benefit both Anderson and our customers. Anderson has had hard savings between $2MM and $4MM annually. Soft savings have also been achieved each year (capacity increase and cost avoidance) of between $500K and $4MM annually. Every year we pass a portion of our saving to our customers. This has ranged from $500K to over $1.5MM over the years (Fig. 6).

Cost savings, capacity increases, and cost avoidance help Anderson be more competitive and hold down prices. At the same time this allows us to significantly reduce the risk of defects and assure on time customer deliveries. Customer savings typically are directly transferred to customers. Some examples are reduced bulk scrap or reduction of cost that customers are typically billed for such as obsolete material disposal or reduced tooling costs.

Continuous improvement and waste reduction are added value benefits of trusting Anderson to make and package the product. This is a free service that makes both Anderson and our customers more competitive with less risk. In addition, over 25% of our projects improve our regulatory profile and directly contribute to stronger compliance with Continuous Good Manufacturing Practices (cGMP), Drug Enforcement Agency (DEA), the Federal Drug Administration (FDA) regulations, and others.

Strategic planning, including annual self-assessments, linked to company-wide balanced score cards and executed through work teams in every area utilizing Six Sigma project and Lean Kaizen projects has been the key. Adding value stream teams, product line teams with responsibility for executing everything for that product line from "the incoming door to outgoing door" midway through our journey propelled us to now performance heights. One of our product lines actually improved profit margins by 50%.

Following the Malcolm Baldrige performance excellence model, dogmatic leadership to our mission vision and plans, engaging, training, empowering and holding everyone in the organization responsible, engaging and meeting our customers and supplier needs and finally rewarding our leadership, our employees, our suppliers and our customers created a "perfect storm" leading to a company worthy of winning the Baldrige.

Author Biography

Charles Aubrey is currently an Executive Coach and Consultant at Aubrey Partners. He is a Lean Six Sigma Master Black Belt. He recently retired as Vice President, Quality and Performance Excellence at Fortune 500 companies AmerisourceBergen Corporation which won the Baldrige Award in 2012. He held similar positions at four other Fortune 500 companies. Charles worked as Vice President Consulting for 10 years with Dr. Juran at the Juran Institute. He helped draft the original Baldrige Criteria and served as a Senior Examiner for the Baldrige Program for 4 years. He specializes in Leadership and Strategic Planning, Customer and Employee Measurement, Lean and Six Sigma, Process Management and Benchmarking and Performance Excellence.

He has written two books, *Quality Management in Services* and *Teamwork-Involving Employees in Quality and Productivity*. He has also written chapters in seven other books, most notably in *Dr. Juran's Quality Management Handbook*. His books have been published in Spanish, Italian, Portuguese, Chinese, and Indonesian. He has published over 150 articles and papers.

Part III
TQM and Quality Audit

Mr. Ellson Boey, Council Member of Singapore Quality Institute.

Total Quality Management (TQM) is a set of concepts and methods, which include all aspects of work within an enterprise with objective to achieve development and continuous improvement in quality levels for all products impacted processes.

Top organizations in today world deploy Total Quality Management (TQM) as their Business Model. The application of TQMs has been integrated in a wide diversity of business setup and operation models and presents itself in different forms and modules. Each organization terms TQM with a slight different terminology in their Business Model setup, however, basic concept of TQM applies.

Audit is a core element within TQM which provides assurance through provision of objective evidences that business processes impacting product quality are complying with underlying requirements and guidelines. The papers entitled "Risk Management in Internal Audit Planning" reinforces the concept that auditing is a critical building block for a TQM organization.

About the Author

Ellson Boey holds Honors Degree of Engineering in Electrical and Electronic Engineering from Nanyang Technological University (NTU). He has more than 13 years of working experience in Research and Development, Quality Management, Continuous Improvement and Training Management, which he plays a key role as project manager, functional manager, and facilitator. Ellson Boey also holds executive council position in Singapore Quality Institute.

In terms of professional certifications, Ellson Boey is a Certified Project Management Professional (PMP), Certified Six Sigma Black Belts (CSSBB), Certified Quality Engineer (CQE), Certified Quality Manager (ASQ CQM), Certified Quality Auditor (CQA), and ISO Lead Auditor Trained (IRCA). He also holds Advanced Certificate in Training and Assessment (ACTA) awarded by Singapore Workforce Development Agency (WDA).

Risk Management in Internal Audit Planning

Sharon Tay

Abstract Internal audit planning is a critical step in achieving effective and efficient internal audits. This study looks at how risk management approach is used to plan internal audit. Purpose is to reduce the subjectivity and time spent in audit planning. A two-step risk management approach is used. In Step 1, the audit frequency is determined based on the department's Workplace Environment, Health and Safety (EHS) risk. In Step 2, the level of readiness of a department is determined by looking at the department's level of documentation and level of measuring and monitoring of processes. From the risk assessment, a 3-year cycle plan is developed to ensure all departments are audited at least once in every 3 years. Higher risk departments would be audited more frequently. A review of risk factors is conducted every 3 years or as and when necessary. The subjectivity in selecting a department for internal audit is reduced. Another significant benefit achieved is the reduction in time spent to do audit planning. With the risk-based audit planning, new staff who is not familiar with audit planning can complete the audit planning process efficiently and effectively. Through the risk management approach, the internal audit planning is more structured and effective.

Keywords Risk management · Audit planning · Effective audit · Efficient audit · Internal audit

1 Introduction

Getting an acute tertiary 1600-bedded hospital ready for the various accreditations is a challenge. The effectiveness and efficiency of the internal audit is a top management concern as ill-coordinated audits can be disruptive to normal operations. The hospital has adopted a risk management approach to establish an effective and

S. Tay (✉)
Sengkang Health, Singapore, Singapore
e-mail: leekat96@gmail.com

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_6

efficient audit plan to meet the requirements of different management system standards.

2 Background of the Problem

The department responsible for establishing the audit plan for the entire year starts the planning cycle starts in November. This is where the audit plan is established for the integrated internal audit to be conducted in the following year.

Planning an effective audit plan for hospital is very challenging in a Hospital. Problems encountered during audit planning are as follows:

- How departments are selected to be audited is subjective
- Diversity in the processes, eg. Patient care, administrative (purchasing, call center), operations support (maintenance, housekeeping), diagnostic (laboratory testing, X-ray) which pose different risk levels in terms of environment, health, and safety
- Limited number of internal auditors; need to mix and match competency required to audit certain departments.
- Inexperience staff who is not familiar with the department functions, took a longer time to plan an audit

Depending on the experience of the staff, the selection process can be haphazard. This may result in critical department not being selected, rationale for selection is vague, or the competency of selected auditor is lacking in audit team. Therefore, the aim of this project is to develop an effective audit plan in an efficient manner.

3 Current Process

Audit planning takes about 30–40 man/days for an inexperienced staff. Staff in charge will look at the past year audited department, new services/equipment, and other factors like past audit findings to select departments to be audited. Next step is to select the internal auditors with relevant competency skills to form the audit team. Proposed audit plan can be rejected by the Management Representative (MR), mainly due to critical department not selected or rationale for selecting the department is not clear.

The current audit plan (see Fig. 1) only provides information on the processes in the department.

		Information about Department's Processes			
S/N	Name of Department	Patient Care	Diagnostics	Admin / Operations Support	Hazardous Material/Gases
1	Kitchen			Yes	Yes
2	X-Ray	Yes	Yes		Yes
3	Call Centre			Yes	
4	QA dept			Yes	

Fig. 1 Information available before implementation of solution (extracted from an audit plan)

3.1 Problem Identification

In year 2011, a brainstorming session was conducted to list the problems in audit planning. The possible causes identified are as follows:

- Lack of selection tool to prioritize departments that need to be audited
- Special set of competency skills required to audit a particular department is not available
- Selection process relies on staff experience. The tacit knowledge in audit planning is difficult to document into standard operating procedures

3.2 Proposed Solution: Risk Management in Audit Planning

At that time, the hospital has implemented a risk assessment framework to identify workplace hazards. The proposed solution was to use the risk management approach which departments are familiar with to plan the integrated internal audit. The risk assessment matrix used in the Code of Practice on WSH Risk Management (WSH Council 2011) was adapted as a tool to aid in audit planning.

A two-step risk matrix (see Figs. 2 and 3) was developed to identify high risk departments in terms of

- Environment, health and safety (EHS) risks
- Level for readiness in terms of documentation and monitoring of process parameters

Based on the two-step risk matrix, the frequency of audit is established. For departments that must be audited every year as specified by management system standards, the audit frequency is set as every 1 year.

EHS Risk Level* (High, Medium, Low)	Audit Frequency
High	Every 1 Year
Medium	Every 2 Years
Low	Every 3 Years

Fig. 2 Step #1 EHS risk level. *Note* The EHS risk level is determined based on the workplace hazards: *high*—use a lot of hazardous materials; registered factory. *Medium*—use some hazardous materials. *Low*—office/admin environment

	Documentation			
Monitor and measure of process	1 - Not Met	2 - Minimal	3 - Met	4 - Exceed Standard
1 - Not Met	Every 1 Year	Every 1 Year		
2 - Minimal	Every 1 Year	Every 2 Years	Every 2 Years	
3 - Met		Every 2 Years	Every 3 Years with option to extend one year	Every 3 Years with option to extend one year
4 - Exceed Standard			Every 3 Years with option to extend one year	Every 3 Years with option to extend one year

Fig. 3 Step #2 level of readiness

		EHS Risk Level		Level of Readiness					3 Year Audit Plan		
S/N	**Name of Department**	**EHS Risk Level (High, Medium, Low)** **Low -** Office **Medium -** Use some hazardous materials **High -** Use a lot of hazardous materials; registered factory	**Preliminary Audit Frequency (year)**	**Documentation & Records** **1 - not met** **2 - minimal** **3 - met** **4 - exceed std**	**Monitor & Measure** **1 - not met** **2 - minimal** **3 - met** **4 - exceed std**	**Preliminary Audit Frequency (year)**	**Other Factors**	**Recommended Audit Frequency** **1 — Every year** **2 — Every 2 years** **3 — Once every 3 yrs**	**Year 2015**	**Year 2016**	**Year 2017**
1	Kitchen	High	1	3	3	3	Registered factory	1	Yes	Yes	Yes
2	X-Ray	Medium	2	3	3	3	New equipment in 2017	2	Yes		Yes
3	Call Centre	Low	3	1	2	1	-	1	Yes	Yes	Yes
4	QA dept	Low	3	3	3	3	Mandatory by Certification Body	1	Yes	Yes	Yes

Fig. 4 Extracted 3-year audit plan

A 3-year audit plan is then developed (see Fig. 4) which acts as a guide in selecting the departments for internal audit. In the final audit plan, current business/work environment is taken into consideration, together with other factors like last external audit date, regulatory requirement, change in process, new equipment, or major incidents.

3.3 Results

The use of risk matrix does provide a degree of objectivity in selecting departments to be audited. Though there is still a certain element in subjectivity, the entire audit planning process is more transparent.

The time taken to plan an internal audit is reduced by about 50%, i.e., 15–20 man/days, with a reduction in the rejection rate by the MR. Staff is also less stressed during audit planning.

4 Conclusion

Risk management in audit planning removes the subjectivity element in audit planning. The structured approach allows staff to develop a 3-year audit plan, with the flexibility to update as and when there are major changes in processes, equipment, and current business/work environment.

References

Workplace Safety and Health Council (2011): *Code of Practice on Workplace Safety and Health (WSH) Risk Management*. WSH Council, Singapore.

Author Biography

Sharon Tay is an ASQ certified manager of quality/organization excellence and a MOM Registered Workplace Safety and Health Officer. She has been actively involved in various quality functions in both the manufacturing and healthcare industries for more than 20 years, focusing on implementing and integrating quality management system standards. She has been frequently invited to share her experiences on the implementation process of integrated management system at both local and overseas conferences. To further enhance and keep pace with changes, she is currently pursuing a part-time Degree of Science in Human Factors in Safety at UniSIM.

Part IV
Quality in Education

Prof. Cher Ming Tan, Chang Gung University, Taiwan

Quality tools and Total Quality management (TQM) have been successfully in improving the product quality for many manufactures. Over the recent years, there is also an increasing trend of using these tools and management methodology in service industry to address customer feedback. These tools and management methods are now also applied to Education sectors in Asia.

In the paper entitled "Analysis of Students' feedback to improve course programme", a real case example in Singapore on applying quality tools to address the feedback is demonstrated. A modification in assessing students feedback is also introduced in a paper from Taiwan, entitled "The evaluation of service quality for high education in Taiwan using Importance-satisfaction model", stemming from the fact that the standard customer satisfaction survey may not obtain the significant results in improving service quality as the identified improvement items may not of concern to the students. The importance-satisfaction model is used, and the areas for improvement which is of high important but low satisfaction are identified.

The use of TQM in education sector is also shown by a paper from India, entitled "Total Quality Management in Education Institutions", and the barriers for its implementation is also discussed. Chulalongkorn University in Thailand is also developed a quality assurance information system based on the following five concepts: harmonization, detailed focus of common data set; data managements on facts; business process concern; and data utilization for both internal and external quality assessments and management of the university and university ranking.

In fact, the awareness of academic quality of universities in ASEAN started as early as 1998 with the creation of an organization called AUN-QA. The AUN-QA assessment was inaugurated in Dec 2007. The paper entitled "The impact of ASEAN University Network-Quality Assurance (AUN-QA) Assessment on the Quality of Educational Programmes" shows the establishment of this organization and its impact in improving the educational quality of universities in ASEAN over the years since 2007.

Analysis of Students' Feedback to Improve Course Programme ANQ Congress 2014 Singapore

Francis Chua and Sam Lam

Abstract Over the years, there is an increasing trend of using quality tools and techniques in service sectors to address customer feedback. Deployment of such quality tools in the education sector is relatively new. Coleman College in Singapore faced feedback from the student community that the Advanced Diploma in Accounting course is too focused on academic learning. This issue was brought up to the Academic Board for further discussion. There were a couple of suggestions and the team embarked on a root cause analysis to analyse the feedback. One root cause identified is the lack of skill-based learning in the current course. The team then brainstorm to incorporate MYOB, the accounting software, into the course programme as a new module. In addition to adding more practical aspects to the course, this implementation also benefits the students by improving their employability. A post-implementation feedback was gathered again to ascertain the effectiveness of the learning outcome. The feedback was positive and the implementation is successful.

Keywords Root cause analysis · Feedback · Survey · Learning outcome · Brainstorm

1 Introduction

The vision of the school is to be the choice provider of quality and affordable education services in Singapore.

F. Chua (✉)
Master in Education, Coleman College, 167 Jalan Bukit Merah #02-15 Tower 4,
Connection One, Singapore 150167, Singapore
e-mail: francis@colemancollege.edu.sg

S. Lam
Master in Business Administration, Singapore Quality Institute, 17 West Coast Walk,
West Coast Recreation Centre #02-14/17/17A, Singapore 127157, Singapore
e-mail: samlam@live.com.sg

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_7

The school has attained EduTrust certification from the Council of Private Education (CPE). This certification is only awarded to schools which have met a set of stringent requirements stated by CPE.

1.1 EduTrust Certification

EduTrust is established as part of the Private Education Act of Singapore. The Council for Private Education (CPE) is a Singapore statutory board with the legislative power to regulate the private education sector. In addition to its role as the sectoral regulator of private education institutions (PEIs), the CPE facilitates capability development efforts to uplift standards in the local private education industry.

2 Background of the Problem

The school has a set of service targets established by the management. This is supported by the collection, analysis, reporting and action on student feedback. The school adopts a process to collect data about the student's learning experience in the school through regular feedback and surveys. Analysis of this data provides a holistic understanding of trends in the student experience, and the performance of the faculty in relation to Learning and Teaching Performance Key Performance Indicators.

2.1 Service Targets

To track the school's key performance indicators, the school has set service targets to monitor the entire process of student enrolment up to student graduation or withdrawal. See Table 1 for the list of service targets.

2.2 Processes Adopted in the School

There are generally the following key processes:

- Curriculum Design and Development Process. This process involves planning and design of the curriculum, development of the curriculum materials, and finally the delivery processes to the students.

Table 1 Service targets

Student requests	Course transfer approvals and withdrawals—within 4 weeks from application (subject to ICA's approval)
Resolution of customer complaints	Within 21 working days from the date of feedback
Student feedback	Feedback is gathered on a regular basis for continual improvements
	Higher learning (certificate level onwards)—at end of every module
Exam results release	Within 4 weeks from date of exam
Exam result appeal release	Within 4 weeks for Coleman College's courses and within 8 weeks for collaborated courses with external partners after student writes in within 5 days upon receipt of exam result release
Refund policy	Notified within 7 working days from date of request
Teacher:student ratio	1:40 for language and preparatory courses
	1:100 for lectures/practical/revision/tutorial sessions for higher education courses

- Student Admission and Development Process. This process involves admission of students, and subsequent monitoring of the student development process. The assessment process includes a mix of formative and summative assessment, and the students' progress is being tracked by the school management
- Selection of Academic Lecturers and Non-academic Staff. This process improves the chances that the lecturers and staff can make a positive contribution to the students. After selection, management needs to train and develop the staff.
- Regular Process Review. After establishing the processes it is essential to regularly review the processes to assess their effectiveness.

2.3 Six Sigma Methodology

The school adopts the Six Sigma methodology as part of its continuous improvement programme (Fig. 1).

3 Define Phase

3.1 Goal Statement

To increase student satisfaction score from 60 to 70% by June, for the Accountancy course cohort.

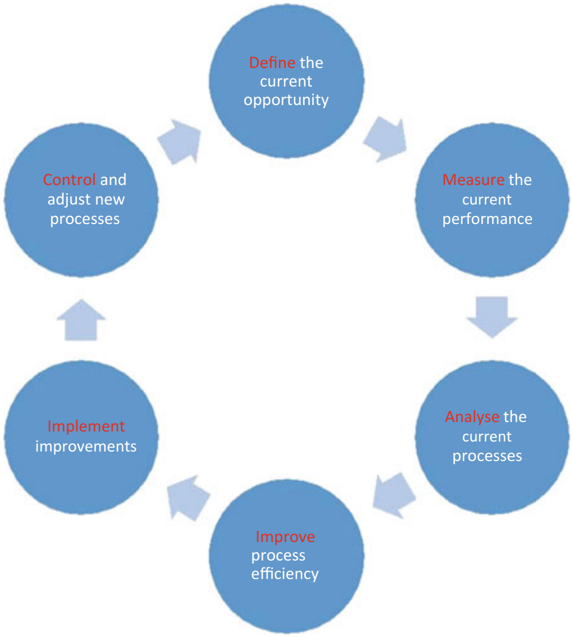


Fig. 1 Six Sigma methodology

3.2 SIPOC (Supplier—Input—Process—Output—Customer)

Figure 2 shows a brief SIPOC model with reference to the project. The supplier in this case is the accountancy course which includes the curriculum, course materials, lesson plans, and the lecturers delivering the lessons. The input will be the students' survey results in the current year. The data collected will be analysed for possible

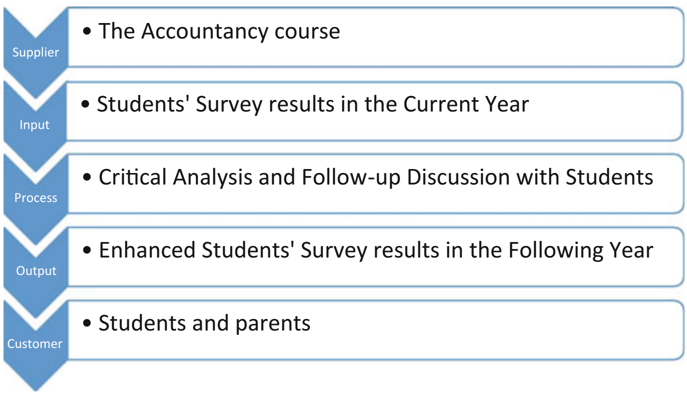


Fig. 2 SIPOC

improvement in the survey results in the following year. The school values the relationship with its customers whom are the parents and the students.

4 Measure Phase

4.1 Data Collection

The survey is conducted for the 100 students currently enroled in the Accountancy course. The students are required to indicate their level of agreement or disagreement (along a scale from 'definitely agree', scoring 5, to 'definitely disagree', scoring 1) with each statement as a description of their programme of study.

Abstract of the survey questions is shown in Table 2.

4.2 Fish Bone (Cause and Effect Diagram)

Fish bone diagram was used to brainstorm all the causes of the low survey score, as shown in Fig. 3.

Table 2 Abstract of survey questionaire

Category	Survey question
Course objectives	The course objectives are clear to the students
Course workload	The course workload is appropriate to achieve the course objectives
Course assignments	The course assignments support the course objectives
Course relevancy	The course is relevant for students' employability when they graduate
Course content	The course materials provided are sufficient
Lecturer knowledge about the course	Lecturer is knowledgeable about the subject
Lecturer effectiveness	Lecturer is effective in the lesson delivery
Facilities	The facilities including the learning environment is conducive
School management support	The school management regularly communicates with the students

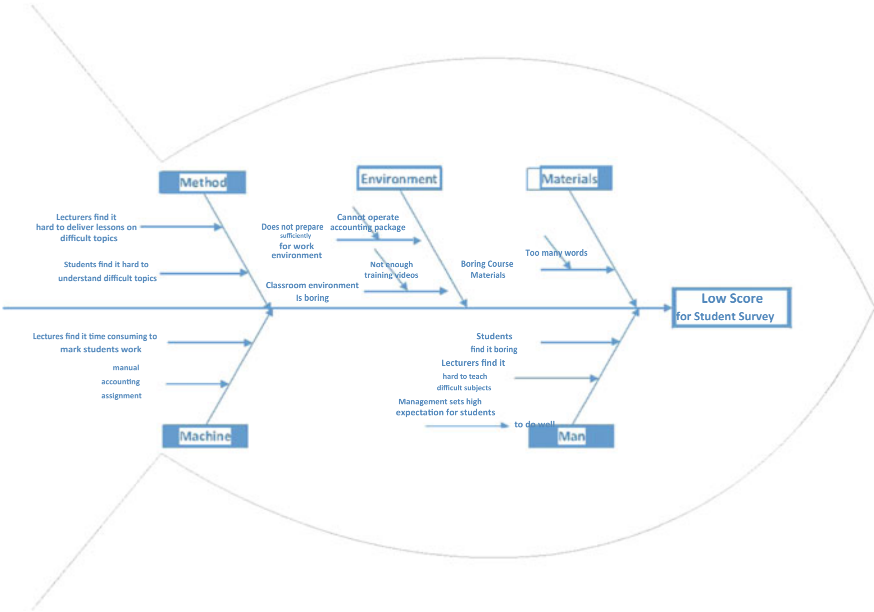


Fig. 3 Fish bone diagram

4.3 *Simple Attribute Gauge Repeatability and Reproducibility (GR&R) Studies*

To ascertain whether the survey results can be attributed to particular lecturers, a simple GR&R analysis was done. Also a group of senior teachers sits in each lecturer's class and write an observation report.

The results show that average scoring for each lecturer in Fig. 4. The average score is quite uniform across all the lecturers and the observation reports also confirmed that there is no issue with individual lecturer.

Lecturer	1	2	3	4	5
Average Score	70	75	72	80	78
Standard deviation	2	3.5	4	6	5.5

Fig. 4 Survey score for each lecturer

5 Analysis Phase

5.1 *Pareto Analysis*

Figure 5 shows the Pareto chart for the survey score of each survey category. It is clear that the course relevance and course content attribute to 85% of the overall low score.

5.2 *Root Cause Analysis*

Based on the Pareto chart, and the earlier fish bone diagram, a root cause analysis was created to confirm the key process input variables (KPIV) as shown in Fig. 6. The finding was to include training of accounting systems into the curriculum.

6 Improve Phase

6.1 *Focus Group Discussion*

To further confirm the necessity to include training on accounting systems, focus group discussions with students, lecturers, and curriculum planners were arranged. Students were very keen to be trained on accounting systems but there is concern that the curriculum topic coverage may be affected with this additional lesson hours. In the end, it was decided that because certain manual workload can be reduced with automation by the accounting system, the curriculum topic coverage can be left intact.

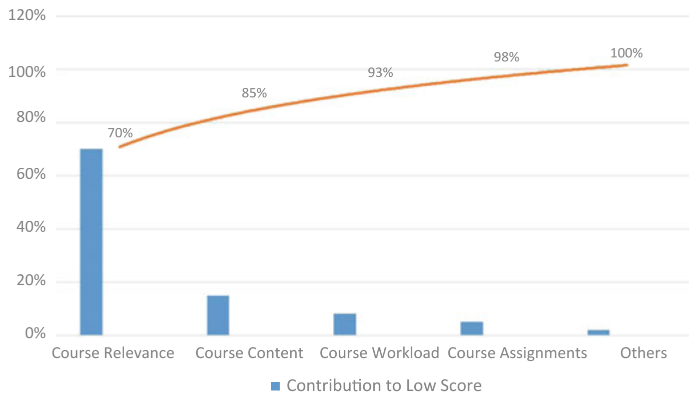


Fig. 5 Pareto chart for survey results

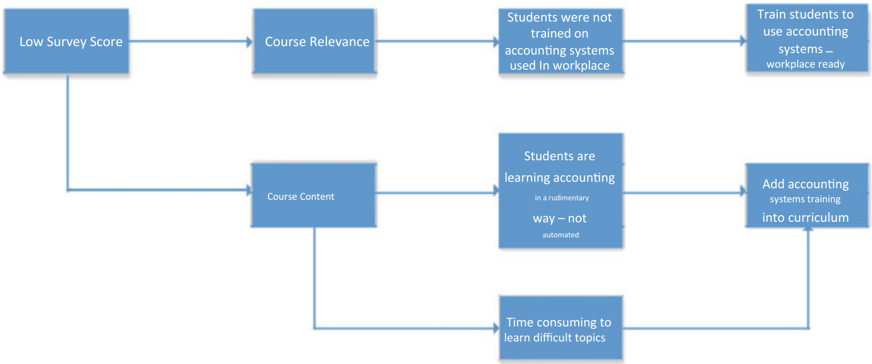


Fig. 6 Root cause analysis for survey results

6.2 *Trial Run for Selected Students*

With management support, the curriculum planning team then started integrating the additional training hours of accounting system into the curriculum, and at the same time review the curriculum to reduce the manual workload.

The school management also gets the students in the focus group discussion to experience the new modified curriculum for 3 months. During the 3 months, the students in the trial run programme provided valuable input and their progress was compared to other students who did not participate. After the 1 month, a summative assessment test was also administered to all the students. 90% of the students in the trial run programme scored higher than the 75 percentile of the cohort.

6.3 *Modified Curriculum Offered to All Students*

The next step is to offer the new modified accountancy curriculum to all the cohort. For those students who are still uncertain about the new curriculum, they can seek additional guidance from the lecturers.

7 Control Phase

7.1 *Results Confirmation*

Survey results were very promising. Two important evaluation factors, course content and course relevancy, were rated the highest among all criteria. Based on the student feedback, the survey results improve from 60 to 70%. In addition, accounting programme enrolment increased by more than 30% in the following year.

7.2 *Close Monitoring*

Formative and summative assessment is conducted regularly and the results are compared to past years results. In addition, the training on accounting systems was certified by the system vendor.

8 Conclusion

By taking an integrated and data-driven approach, the school has been able to implement a comprehensive student tracking process taking into account student feedback. Six Sigma methodology has provided a quality process for school management to act decisively with confidence.

This creates a 'virtuous cycle' where the good student outcomes improved the school's reputation by being ready for the workplace. The employers then view the graduates as well trained by the school. The school is then more willing to continue to invest in the students. This attracts better students.

References

Coleman. About Us. (WWW page). URL : http://www.colemancollege.edu.sg

Council for Private Education. Guide for Choosing a PEI. (WWW page). URL : http://www.cpe.gov.sg/legislation/private-education-act.html

Authors Biography

Francis Chua is a Lecturer in Coleman College. His main interest is Accounting.

Sam Lam is a member of Singapore Quality Institute. He holds a Financial Risk Manager title from Global Association of Risk Professionals. His main interest is Six Sigma Black Belt, Financial Risk and Operational Risk.

The Impact of ASEAN University Network-Quality Assurance (AUN-QA) Assessment on the Quality of Educational Programmes

Ong Chee Bin Johnson

Abstract The creation of the AUN-QA was initiated in 1998 as a network of Chief Quality Officers (CQOs) appointed by the AUN member universities as the focal point for coordinating activities to realise the mission of harmonising educational standards and to seek continuous improvement of academic quality of universities in ASEAN. Since then, AUN-QA has been actively promoting, developing, and implementing quality assurance practices based on an empirical approach where quality assurance practices are shared, tested, evaluated and improved. AUN-QA assessment was inaugurated in December 2007. From 2007 to 2012, 16 AUN-QA assessments covering four countries, 7 universities and 37 programmes were carried out. These 37 programmes were assessed by AUN-QA assessors based on the 15 criteria of the AUN-QA Model at the programme level [1]. To assess the impact of AUN-QA assessments on the quality of educational programmes, an analysis of the results was made using medians and categories of criterion rated based on the AUN-QA Rating Scale. Based on the medians of the AUN-QA criterion rated at the programme level, the number of criteria rated "Better than Adequate" has increased from 1 in 2007 to 13 in 2012. This demonstrates that the quality of the assessed educational programmes at the participating universities has improved over the last 6 years. Recommendations to further improve the quality of the educational programme in the 15 criteria are made.

Keywords Educational quality · ASEAN University Network · AUN-QA · Quality assessment · AUN-QA models

O.C.B. Johnson (✉)
AUN-QA Expert, Education Quality International,
Singapore 544265, Singapore
e-mail: johnsonongcb@gmail.com

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_8

1 Introduction

This paper documents the findings and analysis of the AUN-QA assessments from 2007 to 2012 with the objective of assessing the impact of AUN-QA assessments on the quality of educational programmes of the participating ASEAN Countries namely; Indonesia, Philippines and Vietnam.

2 Background of AUN-QA Network

2.1 Establishment of AUN-QA Network

The creation of the AUN-QA was initiated in 1998 as a network of Chief Quality Officers (CQOs) appointed by the AUN member universities as the focal point for coordinating activities to realise the mission of harmonising educational standards and to seek continuous improvement of academic quality of universities in ASEAN. The AUN-QA activities are carried out by the CQOs in accordance to the Bangkok Accord adopted in 2000, which provides a series of guidelines to promote the development of a quality assurance system as instruments for maintaining, improving and enhancing teaching, research and the overall academic standards of AUN member universities. Since then, AUN-QA has been actively promoting, developing, and implementing quality assurance practices based on an empirical approach where quality assurance practices are shared, tested, evaluated, and improved. The AUN-QA timeline is depicted in Fig. 1.

Fig. 1 AUN-QA timeline

2.2 Objectives of AUN-QA

ASEAN University Network-Quality Assurance (AUN-QA) is the de facto regional quality assurance network for higher education in ASEAN. Its main objectives are:

- Develop, promote, implement and enhance the AUN-QA Guidelines and criteria;
- Facilitate and conduct AUN-QA quality assessments;
- Serve as the authority for AUN-QA quality certification;
- Develop and train quality assurance professionals and practitioners;
- Provide advisory and consulting services on best quality assurance practices; and
- Collaborate with other quality agencies inside and outside ASEAN in relation to harmonisation of quality assurance frameworks and development of professionals in quality assurance in higher education.

3 AUN-QA Assessments

3.1 List of AUN-QA Assessments

ASEAN University Network-Quality Assurance (AUN-QA) assessment was inaugurated in December 2007 after a pilot assessment of the AUN-QA models at Burapha University, Thailand. Since the inaugural assessment, 16 AUN-QA assessments covering 4 countries, 7 universities and 37 programmes were carried out from 2007 to 2012 as listed in Fig. 2.

3.2 AUN-QA Assessment by Country

A summary of the AUN-QA assessments by country is tabulated in Fig. 3.

4 AUN-QA Assessment Results

4.1 AUN-QA Rating Scale

The 37 programmes were assessed based on the AUN-QA Rating Scale as documented in Fig. 4. To fulfil the requirements of the AUN-QA criterion, a minimum score of 4 (i.e. Adequate as Expected) is required.

No.	Period	University	Programme
1st	Dec 2007	Universiti Malaya	• Biomedical Engineering • Computer-Aided Design/Manufacture Engineering
2nd	Aug 2008	De La Salle University	• Chemical Engineering • Applied Economics
3rd	Dec 2008	Universitas Indonesia	• Civil Engineering • Mechanical Engineering
4th	Dec 2008	Institut Teknologi Bandung	• Physics • Pharmacy
5th	Oct 2009	Universitas Gadjah Mada	• Pharmaceutical Sciences • Chemistry • Medical Education
6th	Dec 2009	Vietnam National University – Hanoi	• Information Technology
7th	Dec 2009	Vietnam National University – Ho Chi Minh City	• Information Technology • Computer Science & Engineering • Electronics & Telecommunications
8th	Oct 2010	Universitas Indonesia	• Architecture • Electrical Engineering • Chemistry • Metallurgy & Material Engineering
9th	Nov 2010	De La Salle University	• Chemistry • Psychology • Literature
10th	Dec 2010	Vietnam National University - Hanoi	• Economics
11th	Jun 2011	Universitas Gadjah Mada	• Biology • Geology Engineering • Civil & Environmental Engineering
12th	Nov 2011	De La Salle University	• Physics • Applied corporate Management • Software Technology
13th	Dec 2011	Vietnam National University – Ho Chi Minh City	• Biotechnology • Manufacturing Engineering • Vietnamese Studies
14th	May 2012	Vietnam National University – Hanoi	• English Language • Chemistry
15th	Oct 2012	Universitas Indonesia	• Chemistry • Management
16th	Dec 2012	Vietnam National University – Ho Chi Minh City	• Business Administration

Fig. 2 List of AUN-QA assessment

Country	University	No. of Programmes
Indonesia	Institut Teknologi Bandung Universitas Indonesia Universitas Gadjah Mada	2 8 6
Malaysia	Universiti Malaya	2
Philippines	De La Salle University	8
Vietnam	Vietnam National University – Hanoi Vietnam National University – Ho Chi Minh City	4 7
4 countries	7 universities	37

Fig. 3 AUN-QA assessment by country

Score	Interpretation
1	Absolutely inadequate; immediate improvements must be made
2	Inadequate, improvements necessary
3	Inadequate, but minor improvements will make it adequate
4	Adequate as expected (meeting the AUN-QA guidelines and criteria)
5	Better than adequate (exceeding the AUN-QA guidelines and criteria)
6	Example of best practices
7	Excellent (world-class or leading practices)

Fig. 4 AUN-QA rating scale

4.2 Overall AUN-QA Assessment Results

The overall AUN-QA assessment results (i.e. Mean, Minimum and Maximum Score) of the 15 AUN-QA criteria for the 37 programmes are depicted in the radar chart in Fig. 5.

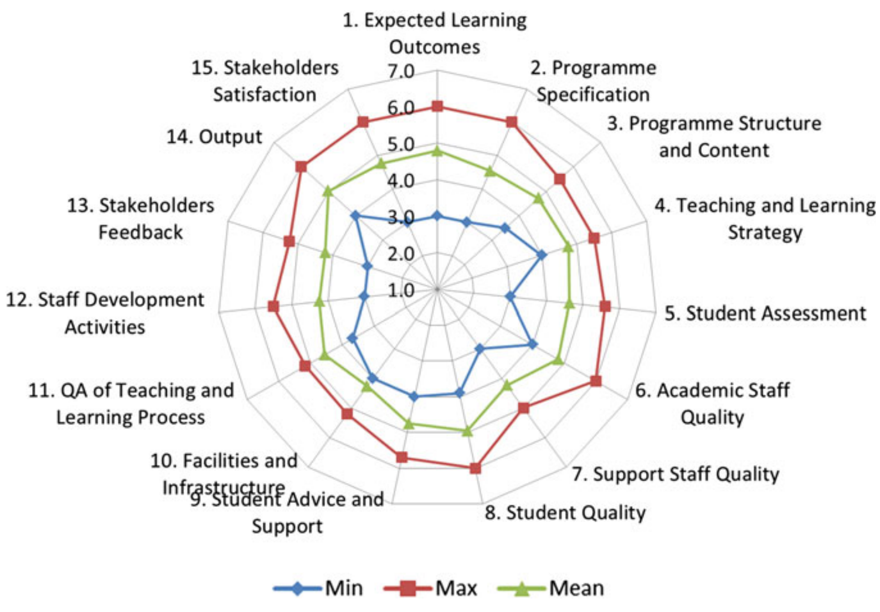


Fig. 5 Overall AUN-QA assessment by criterion

AUN-QA Criteria	Overall	2007	2008	2009	2010	2011	2012
1. Expected Learning Outcomes	4.8	4.8	5.1	4.7	4.7	4.8	4.7
2. Programme Specification	4.5	4.9	4.7	4.3	4.6	4.5	4.6
3. Programme Structure and Content	4.7	4.9	5.1	4.5	4.5	4.7	4.9
4. Teaching and Learning Strategy	4.8	4.1	4.7	4.8	5.0	4.8	4.8
5. Student Assessment	4.6	4.8	4.7	4.5	4.7	4.7	4.6
6. Academic Staff Quality	4.8	4.8	4.7	4.7	4.9	5.0	4.8
7. Support Staff Quality	4.2	4.0	4.3	4.2	4.1	4.4	4.2
8. Student Quality	5.0	4.5	5.1	4.8	5.1	4.8	5.2
9. Student Advice and Support	4.7	4.3	5.0	4.8	4.8	4.6	4.8
10. Facilities and Infrastructure	4.3	4.8	4.5	4.2	4.1	4.2	4.1
11. QA of Teaching and Learning Process	4.6	4.5	4.6	4.7	4.6	4.3	4.7
12. Staff Development Activities	4.2	4.3	4.0	4.7	4.2	4.1	4.2
13. Stakeholders Feedback	4.2	4.0	3.6	4.5	4.2	4.3	4.6
14. Output	5.0	5.0	5.4	4.9	5.0	4.8	5.1
15. Stakeholders Satisfaction	4.8	4.5	4.1	4.9	4.6	5.0	5.2
Overall	**4.7**	**4.5**	**4.6**	**4.6**	**4.7**	**4.6**	**4.8**
No. of Programmes	37	2	6	7	8	9	5

Fig. 6 AUN-QA assessment results by year

The AUN-QA assessment results by year (i.e. Mean, Minimum and Maximum Score) of the 15 AUN-QA criteria for the 37 programmes are tabulated in Fig. 6.

5 Impact on the Quality of Educational Programmes

5.1 Analysis of AUN-QA Assessment Results

To assess the impact of AUN-QA assessments on the quality of educational programmes, an analysis of the results was made using medians and categories of criterion rated based on the AUN-QA Rating Scale. Based on the medians of the AUN-QA criterion rated at the programme level, the number of criteria rated "Better than Adequate" has increased from 1 in 2007 to 13 in 2012 as illustrated in Fig. 7. This demonstrates that the quality of the assessed educational programmes at the participating universities has improved over the last 6 years.

AUN-QA Criteria	Overall	2007	2008	2009	2010	2011	2012
	Medians						
1. Expected Learning Outcomes	5.0	4.8	5.0	4.7	5.0	5.0	5.0
2. Programme Specification	4.8	4.9	5.0	4.2	4.9	4.7	5.0
3. Programme Structure and Content	4.9	4.9	5.0	4.5	4.5	4.8	5.0
4. Teaching and Learning Strategy	5.0	4.1	4.8	5.0	5.0	4.8	5.0
5. Student Assessment	4.9	4.8	4.7	4.6	4.9	4.9	5.0
6. Academic Staff Quality	5.0	4.8	4.9	4.8	5.0	5.0	5.0
7. Support Staff Quality	4.0	4.0	4.4	4.0	4.0	4.3	4.0
8. Student Quality	5.0	4.5	5.0	5.0	5.0	4.8	5.0
9. Student Advice and Support	5.0	4.3	5.0	5.0	5.0	4.9	5.0
10. Facilities and Infrastructure	4.0	4.8	4.4	4.0	4.0	4.0	4.0
11. QA of Teaching and Learning Process	4.7	4.5	4.5	4.7	4.7	4.2	5.0
12. Staff Development Activities	4.0	4.3	4.1	4.5	4.0	4.0	4.0
13. Stakeholders Feedback	4.0	4.0	3.3	4.5	4.0	4.0	5.0
14. Output	5.0	5.0	5.1	5.0	5.0	4.8	5.0
15. Stakeholders Satisfaction	5.0	4.5	4.4	5.0	5.0	5.0	5.0
Overall	4.7	4.5	4.6	4.6	4.7	4.6	5.0
No. of 5s (Better than Adequate)	7	1	6	5	7	3	13

Fig. 7 Analysis of AUN-QA assessment results

Using the categories of criterion rated based on the AUN-QA Rating Scale, there is a divergent trend towards an increasing percentage of higher ratings for “Best Practice” and “Better than Adequate”; and a reduction of percentage in ratings for “Inadequate needs Minor Improvement” from 2007 to 2012. The percentage of ratings in “Best Practice” and “Better than Adequate” has increased drastically from 33% in 2007 to 56% in 2012; while the percentage in ratings for “Inadequate needs Minor Improvement” has significantly reduced from 7% in 2007 to only 1% in 2012. This clearly demonstrates that the quality of the educational programmes of the participating universities has improved after each round of AUN-QA assessment as illustrated in Fig. 8.

The above improvement trend on the quality of educational programmes of the participating universities cut across most countries, especially in Indonesia and Vietnam as tabulated in Fig. 9.

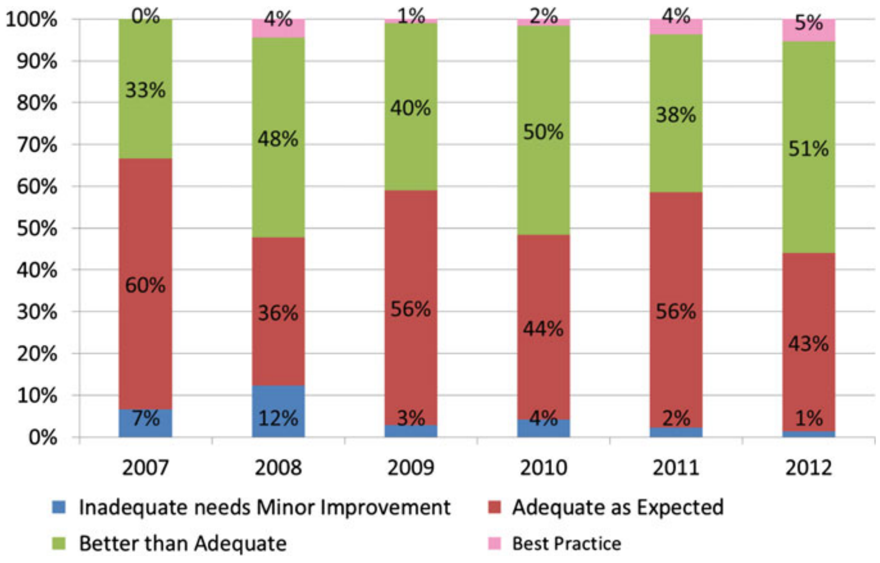


Fig. 8 Improvement of AUN-QA assessment ratings

AUN-QA Rating	Indonesia		Vietnam	
	2008	2012	2009	2012
Inadequate need Minor Improvement	15%	0%	5%	2%
Adequate as Expected	48%	40%	50%	44%
Better than Adequate	35%	53%	45%	49%
Best Practice	2%	7%	0%	4%

Fig. 9 Improved ratings of educational programmes

6 Recommendations for Quality Improvement

6.1 Areas for Improvement

For the purpose of suggesting improvement for the universities in the participating countries, a comparison of the AUN-QA assessment results based on the country's means is used to identify the areas for improvement as illustrated in Fig. 10.

Participating universities can use Fig. 10 as a gauge to seek improvement on the quality of their educational programmes for the criterion where the country's mean of the AUN-QA criterion falls below the overall's AUN-QA criterion mean. The areas for improvement for the respective country are denoted by a "tick" against the AUN-QA criterion as shown in Fig. 11.

AUN-QA Criteria	Overall Ratings	Indonesia	Philippines	Vietnam
1. Expected Learning Outcomes	4.8	5.0	4.9	4.4
2. Programme Specification	4.5	4.7	4.8	4.1
3. Programme Structure and Content	4.7	4.7	4.8	4.6
4. Teaching and Learning Strategy	4.8	4.8	4.9	4.8
5. Student Assessment	4.6	4.4	4.9	4.7
6. Academic Staff Quality	4.8	4.8	5.0	4.7
7. Support Staff Quality	4.2	4.1	4.6	4.2
8. Student Quality	5.0	4.9	5.1	5.0
9. Student Advice and Support	4.7	4.8	5.0	4.5
10. Facilities and Infrastructure	4.3	4.1	4.5	4.2
11. QA of Teaching and Learning Process	4.6	4.5	4.7	4.6
12. Staff Development Activities	4.2	4.1	4.3	4.3
13. Stakeholders Feedback	4.2	4.1	4.1	4.5
14. Output	5.0	5.1	5.3	4.7
15. Stakeholders Satisfaction	4.8	4.5	4.8	5.1

Fig. 10 Identifying areas for improvement

AUN-QA Criteria	Indonesia	Philippines	Vietnam
1. Expected Learning Outcomes			√
2. Programme Specification			√
3. Programme Structure and Content			√
4. Teaching and Learning Strategy			
5. Student Assessment	√		
6. Academic Staff Quality			√
7. Support Staff Quality	√		
8. Student Quality	√		
9. Student Advice and Support			√
10. Facilities and Infrastructure	√		√
11. QA of Teaching and Learning Process	√		
12. Staff Development Activities	√		
13. Stakeholders Feedback	√	√	
14. Output			√
15. Stakeholders Satisfaction	√		

Fig. 11 Areas for improvement

6.2 *Key Improvement Initiatives*

The key initiatives to improve the quality assurance of each AUN-QA criterion are recommended in Fig. 12.

AUN-QA Criteria	Areas for Improvement
1. Expected Learning Outcomes	• Adapt appropriate educational taxonomy for the formulation of learning outcomes. • Implement a systematic way to map learning outcomes to the requirements of various stakeholders. • Enhance life-long learning skills (employability, and professional & academic pathways). • Demonstrate the constructive alignment of learning outcomes, teaching & learning strategy and student assessment. • Lead industry in identifying strategic knowledge & skills to support economic growth and employability.
2. Programme Specification	• Align programme specification to AUN requirements. • Standardise programme and module specifications.

3. Programme Structure and Content	• Offer multiple academic pathways:
	  • Concurrent degrees
	  • Double/joint degrees
	  • Double majors
	• Promote multi-disciplinary and research approach.
	• Provide regional and international student exchanges.
	• Promote industry and community collaboration in learning, projects, research and shared facilities.

4. Teaching and Learning Strategy	• Develop quantitative measures to monitor student-centred learning. • Leverage on information & communication technology to enhance teaching & learning strategy and spaces. • Establish pedagogy research and continuing professional development of academic staff.
5. Student Assessment	• Emphasize formative assessment across the programme. • Demonstrate the constructive alignment of learning outcomes, teaching & learning strategy and student assessment. • Standardise examination appeal policy and process.
6. Academic Staff Quality	• Standardise the calculation of staff-to-student ratio. • Benchmark staff-to-student ratio. • Align employment policy of academic staff to university's mission and vision for internationalisation and research. • Provide continuing professional development of academic staff. • Enhance incentive and reward schemes to improve teaching & research
7. Support Staff Quality	• Establish quantitative measures of service quality. • Continuing technical & professional development of support staff.
8. Student Quality	• Carry out analysis of student quality to academic performance (GPA) • Establish guidelines on student's study load and individual academic ability, and to support multiple academic pathways.
9. Student Advice and Support	• Enhance physical, social and psychological spaces inside & outside the classroom to support student-centred learning and the use of information & communication technology.
10. Facilities and Infrastructure	• Provide continuing development and upgrading of facilities, infrastructure and equipment. • Enhance education & implementation of health, safety and environment standards.
11. QA of Teaching and Learning Process	• Embrace AUN-QA model or adapt appropriate model as a holistic programme design & development process rather than using it as a certification model.

Fig. 12 Key improvement initiatives

AUN-QA Criteria	Areas for Improvement
	• Shorten curriculum design & development cycle, and improve responsiveness to stakeholders' needs. • Promote continuing development and benchmarking of best QA practices and systems within and outside the university.
12. Staff Development Activities	• Adapt a holistic approach to human resource management and development of academic & support staff • Provide continuing training & professional development of academic and support staff. • Establish quantitative measures of staff training & development activities.
13. Stakeholders Feedback	• Establish systemic processes to gather purposeful cyclical feedback from staff, students, alumni, employers, etc. to measure stakeholders' requirements & satisfaction.
14. Output	• Establish benchmarking of output measures. • Enhance undergraduate research activities (vertical and horizontal integration) • Encourage greater research activities of academic staff. • Promote industry collaboration & consultation and applied research.
15. Stakeholders Satisfaction	• Establish quantitative measures of stakeholders' satisfaction. • Establish benchmarking of stakeholders satisfaction.

Fig. 12 (continued)

Reference

1. Guide to AUN Actual Quality Assessment at Programme Level. ASEAN University Network, Version 2.0, 2011.

Author Biography

Ong Chee Bin Johnson is an AUN-QA Expert and a key member of the ASEAN University Network-Quality Assurance (AUN-QA) Council. He is actively involved in the AUN quality assessment and training of Chief Quality Officers (CQOs). He is instrumental in writing the "Guidelines for AUN Quality Assessment and Assessors" and the editing of the "Guide to AUN Actual Quality Assessment at Programme Level". He is appointed as the facilitator for AUN-QA Assessor Workshops, CLMV Countries Workshops, and the AUN-QA Training Course for Accomplishing Programme Assessment Workshops. He is an IQA expert in the ASEAN-QA project and an appointed trainer at the East African Quality Assurance Forum organised by HRK and DAAD.

The Evaluation of Service Quality for Higher Education in Taiwan by Using Importance-Satisfaction Model

Kuo-Chun Tsai, PoTsang B. Huang and Ching-Chow Yang

Abstract Due to the government's education policy significantly change in higher education, number of new universities has established dramatically in the past decades in Taiwan. As a result, the universities are facing the strongly competitive pressures. Most of the universities adjust their strategy in improving service quality to enhance their competition among the universities. Usually they used the customer satisfaction survey to ensure the actual service quality from student's perspectives, and then find out the service items or quality attributes with low satisfaction levels as the items to be improved. Then they will adjust and response to those items which need to be enhanced or improved furthermore to improve their overall service quality and attributes. Several studies assert that only using the satisfaction survey and the related improvement actions may not be obtained the significant results in improving service quality. The main cause is that the improvement items may not concerned by student. Usually the student will concern the service context and quality attributes with high importance. Therefore, we consider the famous I-S model and the related questionnaire to evaluate the service quality for university. We have conducted this model and the related questionnaire survey in a private university in Taiwan. According to I-S model analysis, the results of this research survey show the main pole to be improved (high importance and low satisfaction), including equipment, facilities need to be replaced or upgraded. The survey results will be improved subsequently as further reference for improvement.

Keywords Importance-satisfaction model · Customers' satisfaction · Higher education

K.-C. Tsai (✉) · P.B. Huang · C.-C. Yang
Department of Industrial and Systems Engineering,
Chung Yuan Christian University (Taiwan),
No. 200, Chung Pei Road, Chung Li 32023, Taiwan
e-mail: rogertsaisir@gmail.com

P.B. Huang
e-mail: pthuang@cycu.edu.tw

C.-C. Yang
e-mail: chinchow@cycu.edu.tw

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_9

1 Introduction

Higher education institutions are increased significantly, it has been recognized that higher education is a service industry, and are placing greater emphasis on meeting the expectations and needs of their participating customers, in other word—students. This becomes even more importantly in those universities budgets as a tuition-based model. The rapid expansion of colleges and universities, and significant increases in college education costs combined with demographic shifts in the population, which force colleges to think differently about the role of student satisfaction for their survival (Kotler and Fox 1995). Furthermore, intense competition in today's competitive educational market causes universities to adopt a market orientation strategy to differentiate their offerings from their competitors. Similarly, higher educational institutions in Taiwan are operating in a competitive marketplace. Thus, they need to understand their target markets (i.e. students, external stakeholders of different types), evaluate the target audience's needs, revised their offerings to meet those needs, moreover enhance customer satisfaction by delivering superior quality services (Keegan and Davidson 2004).

From the beginning of the 1980s, quality awareness and customer consciousness have been growing steadily. As a result, enterprises have not only been seeking business growth and an improved competitive position, but have also been devoting themselves to upgrading "quality" (Anderson and Zeithamal 1984). Along with this, service "quality" has also received significant attention.

Service quality aims to confirm the requirements of customers, to meet their expectations and to satisfy them (Deming 1982).

Customer surveys have been used by most enterprises to measure service equality (Babbar 1992). Surveys can be conducted by brief questionnaire, telephone, or mail.

Improving customer satisfaction not only raises company profits, but also facilitates company development (Dubrovski 2001). In higher education, most studies focus on students as "customers", and evaluate their level of satisfaction/dissatisfaction with their programs of study (Comm and Mathaisel 2000). The objectives of higher education are to provide in-depth knowledge, seek academic development, educate students, and coordinate national development demands.

Customer satisfaction and customer loyalty are the key concepts in planning marketing strategies. The importance of customer satisfaction has already been recognized by private industry. Successful and innovative companies, such as Procter and Gamble and American Express, are increasing their efforts to monitor and improve satisfaction, because they realize the value of these activities in enhancing their images, understanding their markets, and increasing their profits (Hennig-Thurau et al. 2002). Same as education business, student satisfaction can bring important impact on future behavior.

The concept of a student satisfaction is derived from the concept of customer satisfaction. Customer satisfaction not only plays an essential role in marketing, but also a critical factor in persuading consumers to make repurchases (Churchill and Surprenant 1982).

Colleges and universities use student satisfaction data to better understand, improve and change campus environments, thereby creating settings more conducive for student development. In this sense, student satisfaction is an indicator of the university's responsiveness to students' needs and a measure of institutional effectiveness, success, and vitality (Hallenbeck 1978; Low 2000; Nichols 1985; Upcraft and Schuh 1996). To make university administration management efficiency, we survey a university's student to know which student care about, what task school should be ranked in priority order.

The purposes of student satisfaction surveys are not only to discover student satisfaction levels, but also to determine necessary improvements by the results of student satisfaction surveys. Student satisfaction surveys commonly apply questionnaire and complaint analyses. However, complaint analysis is a passive method, which cannot fully determine student satisfaction. Recently, firms have increasingly started using questionnaire surveys. This study applies the I-S model conducted by Yang (2003) to analyses student satisfaction.

1.1 Importance-Satisfaction Model (I-S Model)

Low-quality attributes should not be the only consideration when designing improvement plans. Usually, the customer (student) measures the quality of goods or services based on several important attributes or elements (Berry et al. 1990; Deming 1986). Customer (student) evaluates product or service quality by considering several important quality attributes. Therefore, firms must take actions to improve the attributes that are important to the customer but which have low satisfaction levels. Figure 1 shows concept of I-S model conducted by Yang (2003). The results for each quality attribute are placed in the model and then improvement strategies considered based on the position of each item will be established.

2 Methodology

This study employed the Questionnaire Survey methodology. Hence, research was conducted through the distribution of survey questionnaires. The methods and procedures are shown in the following:

1. Make a comprehensive study of all related documents and literatures to determine research topic and structure.

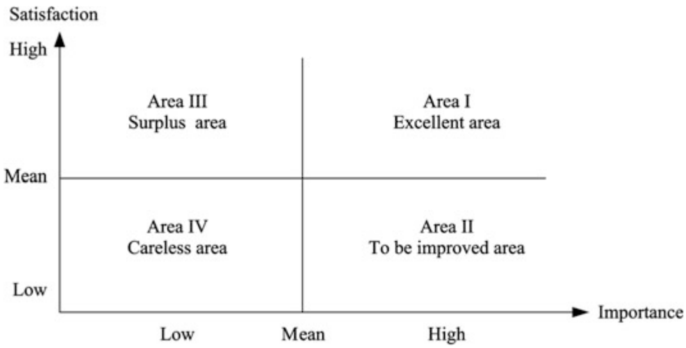


Fig. 1 Importance-satisfaction model [*Source* Yang (2003a)]

2. Compile and analyze related documents to produce self-developed questionnaires.
3. Conduct formal questionnaire survey and perform statistical analysis.
4. Draw a research conclusion and present implications and recommendations based on the findings.

This study issued a questionnaire to a private university located in north Taiwan. Questionnaire respondents using the online way, a total of 495 valid questionnaires were recovered. Questionnaire has a total of 53 question items based on the service level provided by the schools daily contact. This questionnaire was divided into "Importance Survey" and "Satisfaction Survey", each of the scale ranged from 55 to 95 (with 55 representing extremely low importance and 95 representing extremely high importance).

Reliability is generally measured by Cronbach's (a; the Cronbach's of students measurement for higher education was calculated using the statistical software, SPSS. 17.0). The reliability can be confirmed when Cronbach's alpha coefficient for all the latent variables was found to be well above the accepted threshold value of 0.75. The Cronbach's of "Importance" was 0.981, and the Cronbach's of "Satisfaction" was 0.978. The differences between these two figures are small, among 0.97, 0.98, which indicates that the questionnaires administered in this study are highly reliable.

The assessment of validity can be confirmed by factor loading the corresponding endogenous and exogenous latent constructs. In this research, the values of factor loading of the latent constructs for "Importance" are ranged from 0.544 to 0.923, and "Satisfaction" are ranged from 0.485 to 0.891, which are greater than the threshold level of 0.4 proposed by Nunnally (1978). This study also conducted interviews with HR directors. The feedback indicated that the questionnaire had extremely high reliability and validity.

2.1 I-S Model Results

Those quality attributes were placed in the I-S model, as shown in Fig. 2:

(1) **Excellent area**

1. The attitude of school administrative staff overall satisfaction
2. The School General administrative efficiency overall satisfaction
3. The School development and performance overall satisfaction
25. Health care service
27. Campus security facilities and equipment
31. Campus clean
32. Professional competence of service personnel, can actually reply to questions from readers
33. Library catalog easy to search
34. Information exchange and acquisition service
35. Audio-visual equipment, computer equipment, photocopying equipment and reading seats
36. Collections and library resources
45. Department of academic research outcome
46. Departmental activities announcement
47. Departmental office service
48. Departmental resources management

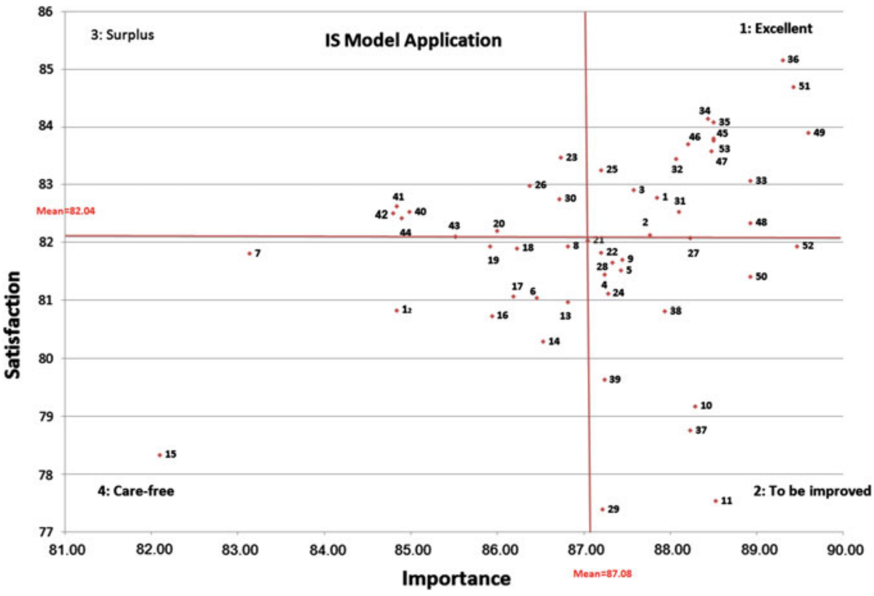


Fig. 2 I-S model application

49. Teaching quality
51. The faculty teaching seriously
53. Overall satisfaction for department.

(2) **To be improved area**

4 Crisis Handling
5 School Announcement for Important Information
9 Schools provide students with learning and sharing environments (such as classrooms, laboratories, etc.)
10 Classroom overall facilities
11 i-learning network teaching platform
22 Employment, internship support, and counseling plan,
24 Scholarship and assistant application business
28 Security and access control
29 Student automobile and motorbike application and car park management
37 Campus network maintenance and service
38 I-touch school administration system maintenance and integration
39 Campus web page planning
50 Classroom and laboratory equipment and maintenance
52 Department of Curriculum Planning.

(3) **Surplus area**

21. Provide work–study opportunities and training
23. Students loan business
26. Student counseling and guidance
30. Campus beautification
40. International students apply for cultural events
41. Exchange students related services promotion
42. International student admissions service promotion
43. Travel learning and study abroad scholarship announcement
44. Overall satisfaction for the school.

(4) **Careless area**

6 School web page maintenance
7 Faculty sport activities organized
8 Sports and Leisure Facilities Management Sports and Leisure Facilities Management
12 Learning alert system
13 Performance management
14 Convenience of equipment loan
15 Semester naming system
16 Teaching evaluation

17 Student dormitory management and counseling
18 Student out of campus housing assistance
19 Students general counseling and management.

Figure 2 shows the average student satisfaction value of 82.04, demonstrating that the school has acceptable operation performance. The "excellent area", in which students are completely satisfied, includes 18 quality attributes. "To be improved area" in which students are dissatisfied with the quality attributes, indicate student hope school can actively improved these items contains 14 attributes. "Surplus area", indicating that schools have acceptable performances in these items, contains 10 attributes. Furthermore, "Care-free area" contains 11 attributes. If school resources are limited, these quality attributes have a low priority. Furthermore, if school resources are abundant, these items should also be improved.

3 Conclusions

Using higher education students as examples, this study proposed the improvement priority based on the perspectives of importance and satisfaction by I-S model theory. Based on school's resources, the result can determine the improvement strategies and priorities to satisfy actual student requirements.

This study found the following should be the highest priority for improvement projects focused on school facilities and the environment (for example: the campus parking problem, updating and maintenance of teaching equipment, network environment), shows that private universities have resources indeed far less public universities, Government should consider that the quality of public and private institutes should be balanced, and provide reasonable subsidy to the private institutes with high reputation.

Of course, if the school has sufficient resources should also give priority to these "To be improved(high importance and low satisfaction)" projects try to improve, however, if the school does not have enough resources, but also without consuming too many resources from the comparison of "Excellent area (high importance and low high satisfaction)" in the project to continue their efforts, for example through continuing education and training to improve service quality administrative service personnel, so that students can continue to enhance the school to increase overall satisfaction.

According to the previous researches (Fornell 1992; Anderson and Sullivan 1993), customer satisfaction is the antecedent of relational outcomes, such as customer loyalty and positive word-of-mouth advising. Testimonials from successful former students might be one of particularly effective strategies to market the institutes.

References

Adee Athiyama, (1997): *Linking student satisfaction and service quality perceptions*: the case o university education, European Journal of Marketing, Vol. 31, No. 7.

Anderson, C. and Zeithaml, C.P. (1984), "*Stage of the product life cycle, business strategy, and business performance*", Academy of Management Journal, Vol. 27, March.

Anderson, E. W. & Sullivan M. W. (1993). *The Antecedents and Consequences of Customer Satisfaction for Firms.* Marketing Science, 12.

Babbar, S. (1992), "*A dynamic model for continuous improvement in the management of service quality*", International Journal of Operations & Production Management, Vol. 12 No.2.

Berry, L.L., Zeithaml, V.A. and Parasuraman, A. (1990), "*Five imperatives for improving service quality*", Sloan Management Review, Summer.

Besterfield, D.H. (2004): *Quality Control*. Pearson Education International, New Jersey.

Ching-Chow Yang (2003a), *Improvement actions based on the customers satisfaction survey, TQM & Excellence,* Vol. 14, No. 8.

Ching-Chow Yang (2003b), *Establishment and applications of the integrated model of service quality measurement, Managing Service Quality,* Vol. 13.

Ching-Chow Yang (2005), *The refined Kano's model and its application,* Total Quality Management, Vol. 16, No. 10.

Comm, C.L. and Mathaisel, D.F.X. (2000), "*Assessing employee satisfaction in service firms: an example in high education*", The Journal of Business and Economic Studies, Fairfield, Spring.

Churchill, G. & Surprenant, C. (1982). *An Investigation into the determinants of consumer satisfaction.* Journal of Marketing Research, 19.

Deming, W.E. (1982), *Quality Productivity and Competitive Position,* MIT Press, Cambridge, MA.

Deming, J. (1986), *Out of Crisis,* MIT Press, Cambridge, MA

Dubrovski, D. (2001), "*The role of customer satisfaction in achieving business excellence*", Total Quality Management, Vol. 12 Nos 7/8.

Fornell, C. (1992). *A National Customer Satisfaction Barometer: The Swedish Experience.* Journal of Marketing, 56.

Hallenbeck, T. R. (1978). *College student satisfaction: An indication of institutional vitality. N.A. S.P.A.* Journal, 16(2).

Hennig-Thurau, T., Gwinner, K. P. & Gremler D. D. (2002). *Understanding Relationship Marketing Outcomes.* Journal of Service Research, 4(3).

Keegan, W.J. and Davidson, H. (2004), *Offensive Marketing: Gaining Competitive Advantage,* Elsevier, Amsterdam.

Kotler, P. and Fox, K.F.A. (1995), *Strategic Marketing for Educational Institutions*, 2nd ed., Prentice-Hall, Englewood Cliffs, NJ.

Low, L. (2000). *Are college students satisfied? A notional analysis of changing expectations.* The USA Group Foundation New Agenda Series, 2(1). Indianapolis, IN: USA Group Foundation.

Nichols, H. J. (1985). *Black college student satisfaction as a measure of effectiveness of higher education*. Unpublished doctoral dissertation, Kent State University, Kent, Ohio.

Nunnally, J. C. (1978). *Psychometric Theory*, 2nd ed., McGraw-Hill, New York.

Oscar W. Deshields Jr, Ali Kara, Erdener Kaynak, *Determinants of business student satisfaction and retention higher education: applying Herzberg's two-factor theory,* International Journal of Educational Management, Vol 19.

Upcraft, M. L., & Schuh, J. H. (1996). *Assessment in student affairs: A guide for practitioners.* San Francisco: Jossey-Bass.

Yu-Fen Chen, Chin-Hui Hsiao, & Wen-Ching Lee, *How Does Student Satisfaction Influence Student Loyalty –From the Relationship Marketing Perspective.*

Authors Biography

Kuo-Chun Tsai Kuo-Chun Tsai *is a Ph.D. student, Department of Industrial and Systems Engineering, Chung Yuan Christian University, Taiwan, R.O.C.*

Experience:
Senior Lecturer, Top 500 Lecturers in Taiwan
Part-Time Lecturer, Vanung University and Yaun Ze University
Chairman, Elite Association of Republic of China
Senior Tutor of Team momentum

PoTsang B. Huang PoTsang B. Huang *is a Assistant Professor, Department of Industrial and Systems Engineering, Chung Yuan Christian University*

Education: Ph.D., Iowa State University Industrial Technology Education and the U.S.
Experience:
Assistant Professor Department of Mechanical and Electrical Engineering Technology, Georgia Southern University
Assistant Professor, Department of Industrial Engineering, Chung Yuan Christian University
"Taiwan Ministry of Education to study Elite Office" Director

Ching-Chow Yang Ching-Chow Yang *is a Professor, Department of Industrial and Systems Engineering Chung Yuan Christian University*

Education: Ph.D., Management Science National Chiao Tung University University R.O.C.
Experience:
Head of Department of Industrial Engineering, Chung Yuan Christian University
Chung Yuan Christian University alumni director-cum-employment counseling

Quality Assurance Information System of Chulalongkorn University

Thongchai Pitimanuskul, Grittapat Akkaphuriwong, Wibool Khongdee and Vimolvan Pimpan

Abstract Data management is one of the necessary processes in quality management. Chulalongkorn University has developed the tool for this process called Quality Assurance Information System (QAIS). This system was originally developed based on "Lotus Notes" software. Currently, its system consists of the processes of data input, reference input, data checking and correction, data evaluation, and report preparation. By using this system, the data collection process changes from "Offline" to "Online via Intranet." It has been used to support internal and external quality assessments according to national quality standards of Thailand since this system can connect and transfer data to the national database called CHE QA Online (Commission on Higher Education Quality Assessment online system). Chulalongkorn University was nationally awarded by the Office of the Public Sector Development Commission in the category of Servicing Innovation for implementing this system. Therefore, Chulalongkorn University would like to share and exchange the concept and the operation of this system for further improvement.

Keywords QAIS · Quality assurance · Quality management · Data · Chulalongkorn University

---

T. Pitimanuskul (✉) · G. Akkaphuriwong · W. Khongdee
Office of Strategy Management and Budgeting, Chulalongkorn University, Phayathai Road, Pathumwan, Bangkok 10330, Thailand
e-mail: Thongchai.Pi@chula.ac.th

G. Akkaphuriwong
e-mail: Grittapat.A@chula.ac.th

W. Khongdee
e-mail: Wibool.K@chula.ac.th

V. Pimpan
Chulalongkorn University, Pathumwan, Thailand
e-mail: vimolvan.p@chula.ac.th

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_10

1 Introduction

Quality Assurance Information System (QAIS) of Chulalongkorn University is developed based on five concepts as follows:

(a) *Harmonization*
Since there are many operational units in Chulalongkorn University including 19 faculties, 3 colleges, 9 research institutes, and more than 60 centers and administrative offices and these units which have different missions, roles and functions, there are many types of data needed to handle. Furthermore, there are many IT systems or programs used in these units. Therefore, to avoid the effect on the operation of each unit, each unit can still use its own system to collect the data but the definition and the criteria of the data as same as in QAIS are applied. Then QAIS gathers the data from each unit via standard template using excel program as shown in Fig. 1.

(b) *Detailed Focus of Common Data Set (CDS)*
Because the data supplied and used in QAIS is the data gathered from the operation and management of all units in the university, some data is the combination of several sub-data. In addition, the definition and the criteria of the data used for different purposes may be different. For example, the definition of “Academic Staff” used for internal quality assessment based on national quality standards of Thailand is different from that used for university rankings by QS and Times Higher Education. Moreover, some data is collected based on the fiscal year while the others are collected based on the academic year. Therefore, QAIS is designed to gather in the level of the sub-data as much as possible.

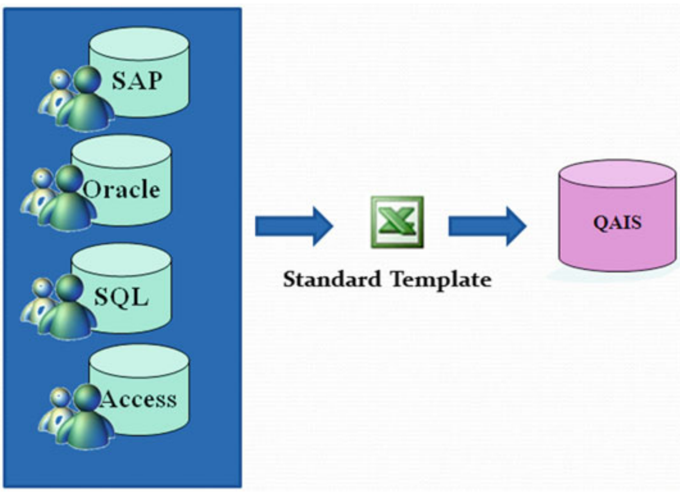


Fig. 1 Data collecting process of QAIS

(c) *Data Managements on Facts*
Files of the references or the evidences supporting the data are important in QAIS. These files are useful in the checking, reviewing, and assessment processes. Therefore, the system requires the attachment of these files at the first time the data is submitted and every time the data is changed or corrected.

(d) *Business Process Concern*
Since the main objective in the development of QAIS is to find the most suitable and effective process in data management of Chulalongkorn University, "Lotus Notes" is used as a tool for the development. This is due to its ease in using and changing the requirements of the program that can support the needs of each unit of the university.

(e) *Data Utilization*
Data supplied in QAIS is not only used for internal and external quality assessments but also for the management of the university and other purposes such as university ranking. Therefore, the system is designed to support the diversity of data and the variety of the format of the reports.

2 Management and Databases of QAIS

The control and management of QAIS can be divided into three levels depending on the units responsible for the data and sub-data. The Office of Strategy Management and Budgeting (OSM) has the responsibility for overall control and management of QAIS. Other main administration offices of the university besides OSM are responsible for the data and sub-data based on their roles and functions. For example, the Office of Research Affairs is in charge of all research data and sub-data while the Office of Student Affairs is in charge of all data and sub-data related to student activities and awards. These offices are called "Data Controllers." All other units of the university excluding OSM and Data Controllers are called "Data Owners." They are responsible for the data and sub-data resulting from their operation.

The overall working processes can be summarized as follows:

1. OSM sets the parameters and prepares the CDS Survey Form.
2. Data Controllers submits the original data and files of references and evidences.
3. Data Owners checks the original data and files submitted by Data Controllers. If these data and files are the same as what they have, they are required to confirm the information. On the other hands, if the data and those files are different from what they have, they can submit the request for changing the data and must attach the files of the evidences to support the change request.

4. Data Controllers review the change request and if they agree with the new information, they change the data and the files in the system. Then Data Owners and OSM acknowledge the change.
5. Data Controllers confirms all the data.
6. OSM closes the data submission system.
7. The next step is depended on the purpose of data utilization. For example, in the case of internal quality assessment, Data Owners prepare the KPI (Key Performance Index) Survey Form, analyze the KPIs based on data related to the KPIs, and submit the evaluation of the KPIs to the system. The Self Assessment Report (SAR) of each Data Owner is then printed.
8. In the same case at the university level, OSM prepare the KPI (Key Performance Index) Survey Form. Then Data Controllers analyze the KPIs they are responsible for based on data related to the KPIs and submit the evaluation of the KPIs to the system. OSM concludes the final evaluation and prints the SAR.

QAIS involves the following databases as shown in Fig. 2.

In addition, QAIS is designed to link to CHE QA Online (Commission on Higher Education Quality Assessment online system) which is national database system. Therefore, the data of Chulalongkorn University can be easily transferred to CHE QA Online as shown in Fig. 3. Some data can be posted on the OSM website for anybody to access or printed as reports for administration.

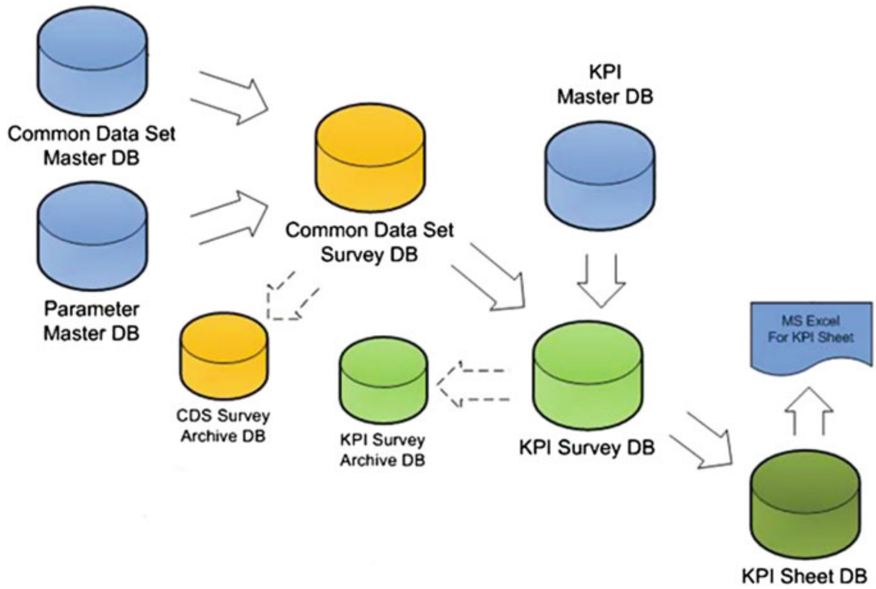


Fig. 2 Databases of QAIS

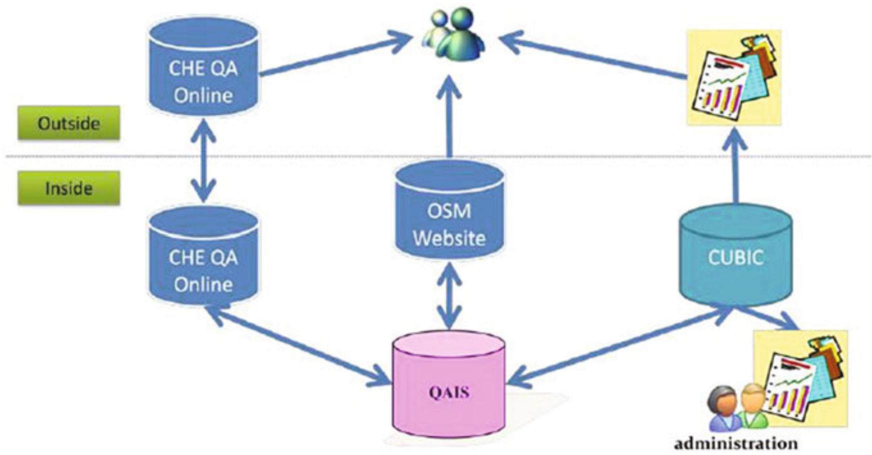


Fig. 3 Data utilization of QAIS

3 Conclusion

Since QAIS is implemented, the data used at both university level and other units in the university are the same. This reduces the problem of the inconsistency and the difference of the data. The data supplied in QAIS is based on "Fact." Therefore, not only the numerical data is supplied, the details of each data are also included in the system. Moreover, all users at each level can communicate to each other, check and review the data at all times. The important outcome is that Chulalongkorn University has the databases that effectively support internal and external quality assessments, strategic management, university ranking, and other purposes.

Acknowledgments The authors would like to acknowledge Assistant Professor Prasert Akkharaprathomphong, Mr. Worachok Chaiwong, Mrs. Sirinart Tupsang and Chulalongkorn University for their supports.

References

www.mua.go.th
www.topuniversities.com
www.timeshighereducation.co.uk

Authors Biography

Mr. Thongchai Pitimanuskul They are the staff of the Office of Strategy Management and Budgeting. They are in Quality Management Division. Their responsibilities are in internal and external quality assessment processes of the university according to national quality standards of Thailand by the Office of the Higher Education Commission and the Office for National Education Standards and Quality Assessment, respectively. They are also in charge of the control and management of QAIS.

Mr. Grittapat Akkaphuriwong They are the staff of the Office of Strategy Management and Budgeting. They are in Quality Management Division. Their responsibilities are in internal and external quality assessment processes of the university according to national quality standards of Thailand by the Office of the Higher Education Commission and the Office for National Education Standards and Quality Assessment, respectively. They are also in charge of the control and management of QAIS.

Mr. Wibool Khongdee They are the staff of the Office of Strategy Management and Budgeting. They are in Quality Management Division. Their responsibilities are in internal and external quality assessment processes of the university according to national quality standards of Thailand by the Office of the Higher Education Commission and the Office for National Education Standards and Quality Assessment, respectively. They are also in charge of the control and management of QAIS.

Dr. Vimolvan Pimpan Graduated from Chulalongkorn University in 1994 with her B.Sc. degree in Materials Science (first honor). Then she continued her study for Ph.D. degree in Polymer Science and graduated from The University of Akron in 1999. She started working as a lecturer at the Department of Materials Science, Faculty of Science, Chulalongkorn University in 1994. She is now the Assistant to the President of Chulalongkorn University for quality management and administrative affairs.

Total Quality Management in Educational Institutions

A.K.M.F. Hoque, M.S. Hossain and M.A. Zaman

Abstract The concept of Total Quality Management (TQM) has been discussed in the present work stressing its application in the field of education. Side by side definition, concept, principles of TQM, and philosophy are discussed. In this present work also discussed synergistic relationship, continuous improvement, self-evaluation, a system of ongoing process leadership, etc. TQM process in educational institutes generally includes (a) Mission and Passion, (b) Administrator as a role model, (c) Environment factors, (d) Accountability, (e) Humans relations and (f) feedback. Implementation principles and processes of the concept in education have been narrated after identifying the barriers for implementing total quality in education. TQM will help achieve excellence, which only can guarantee the survival of institutions now in a highly competitive world, with ever decreasing subsidy in the education sector.

Keywords Principles of TQM · Philosophy · Mission and passion · Administrator as a role model · Environment factors · Accountability · Humans relations · Feedback · SQCC · 5-S · TWIT · Management commitment · Employee empowerment · Customer focus · Continuous improvement

1 Introduction

The concept of Total Quality Management (TQM) was developed by an American, after World War II for improving the production quality of goods and services. The concept was not taken seriously by Americans until the Japanese, who adopted it in

A.K.M.F. Hoque (✉)
Physics Department, Tejgaon College, Farmgate, Dhaka, Bangladesh
e-mail: dfhoque@gmail.com

M.S. Hossain
Physics Department, National University, Gazipur, Bangladesh

M.A. Zaman
Physics Department, Jahangirnagar University, Savar, Dhaka, Bangladesh

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_11

1950 to resurrect their postwar business and industry, used it to dominate world markets by 1980. The TQM is quite popular in management circles today. TQM refers to a set of philosophies that reorient management systems to focus on the efficient achievement of organizational objectives in order to exceed customer expectations and maximize stakeholder value. In the last two decades, TQM have enabled Japanese companies increase their competitive advantage in costs and quality, and to consequently gain significant global market share in a number of variety of industries. TQM, in many ways, is practically a formalized common sense.

While TQM was pioneered in the manufacturing sector, its benefits have been brought to service and public sector organizations. Many educational institutions in the US and Canada are applying TQM techniques to improve their organizational effectiveness and the quality of their programs.

A total quality approach in running our schools is necessary for equipping learners to function to their fullest potential in an environment of depleting resources. The schools have to be dynamic and flexible to meet the demand. The expectations of students, parents, and the public in general vis à vis educational priorities, costs, accessibility, programs, and relevancy, make it imperative for schools to undergo continual assessment and improvement. Funding resources for education are diminishing at a rapid rate. Schools have to find innovative ways of cutting costs without cutting quality.

2 Introducing TQM

2.1 TQM Defined

TQM is a management philosophy that seeks to integrate all organizational functions to focus on meeting customer needs and organizational objectives. It views an organization as a collection of processes and maintains that organizations must strive to continuously improve these processes by incorporating the knowledge and experiences of workers. Although originally applied to manufacturing operations, and for a number of years only used in that area, TQM is now becoming recognized as a generic management tool, just as applicable in service and public sector organizations.

2.2 TQM Concept

Quality of education is a multi-dimensional concept, with varying conceptualizations. It includes the quality of inputs in the form of students, faculty, support staff and the infrastructure: the quality of processes in the form of learning and teaching activity: and the quality of outputs in the form of the enlightened students who move out of the system. Quality control is an effective system of ensuring quality, ensuring continuing excellence. TQM is a modern term wider in scope than the total quality control (TQC). TQC considers the role of employees in improving the productivity. But it remains silent about the quality of work life, employee satisfaction, and organizational development. TQM takes into its fold not only ensuring productivity and efficiency but also ensuring individual satisfaction and institutional building and human well being.

2.3 TQM Philosophy

TQM approach keeps the organizational goals at the supreme but there is a fundamental shift in philosophy from work centered to employee centered. TQM believes in the following:

(a) There are no workers and no managers, all employees of an educational institution have important roles to play. The role of each one is important. Therefore, we must realise that all are facilitators and team members, the head of them being the leader of the team. The team can never succeed unless everyone puts into his/her best.
(b) Involvement is the key word. It means participation plus commitment and pride.
(c) Everyone is made to identify oneself with the institution. Employees voluntarily come forward as the relationships should be family oriented where everyone gives his/her best even without asking.
(d) TQM requires a new set of values. There should be openness, transparency, trust, patience, respect, and discipline.

3 TQM Concept for Education

Components of TQM in education
Awareness and Commitment for Everyone, A Clear Mission, A Systems Planning Approach, Teaming Replacing Hierarchy, Enabling and Empowerment Replacing Fear.

Focus on Mastery Learning
The sequences are: (1) Plan, (2) Teach, (3) Test

The TQM alternative is: (1) Plan, (2) Teach (DO), (3) Check, (4) Revised Teaching (ACT), (5) Test

PDCA Cycle Activities

PDCA Cycle will be repeated again and again for "**Continuous Improvement**"

Focus of practicing TQM in educational institutions must embrace at least the following

(1) Knowledge: 100% understanding of subjects (2) Wisdom: ability to apply knowledge and experiences to make the right decisions, the correct choices and the right judgment (3) Eloquent Speech: Clear, fluent, and flawless presentation of ideas and living a participated life (4) Spiritual Perception: Ability to differentiate between right and wrong.

TQM concept must be applied to all activities of the institution such as

Morning assembly, Lesson plan, Games, Teachers training, Admission, Recruitment, Examination/Assessment, Discipline, SQCC, 5-S, TWIT, etc.

5S's stand for five Japanese words

1. **SEIRI** (Sorting): Deciding what is/is not necessary, and disposing the latter.
2. **SEITON** (Setting in Order): Storing the necessary things in order, so that the right items can be picked up efficiently (without any delay/waste) at the right time.
3. **SEISOU** (Shining): Creating a clean workplace without dust, dirt, and garbage.
4. **SEIKETSU** (Standardization): Keeping the good state of SEIRI, SEITON, SEIKETSU.
5. **SHITSUKE** (Self-Discipline): Training officers/employees for making the four steps as a regular rule at the workplace.

The TQM is applicable to academics. Many educators believe that the Deming's (1900) concept of TQM provides guiding principles for needed educational reform. In his article, "The Quality Revolution in Education," John Jay Bonstingl (1985) outlines the TQM principles he believes are most salient to education reform. He calls them the “Four Pillars/Principles of Total Quality Management.” These are described below.

(a) **Synergistic Relationships**

The very application of the first pillar of TQM to education emphasizes the synergistic relationship between the “suppliers” and “customers”. The concept of synergy suggests that performance and production is enhanced by pooling the talent and experience of individuals. In a classroom, teacher–student teams are the equivalent of industry’s front-line workers. The product of their successful work together is the development of the student’s capabilities, interests, and character.

(b) **Continuous Improvement and Self-Evaluation**

The second pillar of TQM applied to education is the total dedication to continuous improvement, personally, and collectively. TQM emphasizes self-evaluation as part of a continuous improvement process.

Management by Measurement in TQM

The two basic purposes of TQM in education

(a) Improved learning (b) Improved cost effectiveness

Form a TQM steering committee that.

(a) Develops a plan for supporting the staff in TQM implementation (b) Builds a positive connection between that committee and the traditional supervisors (c) Use advice from consultants and/or from schools that have succeeded at TQM transformation

Implementing TQM

1. Aligning the Institution
   (a) People (b) Process
2. Ensuring Mind-set/Attitude to Understand the following
   (a) Quality: A new Philosophy, (b) KIAZEN, (c) Internal Customer, (d) Working with Facts, (e) Respect for Humanity
3. Through Participation and Involvement Utilize Following Tools
   (a) Policy Management, (b) Students Quality Control Circle, (c) Teachers Work Improvement Teams, (d) 5-S Techniques, (e) Quality Assurance (and Process Orientation)

(c) **A System of Ongoing Process**

The third pillar is the recognition of the organization as a system and the work done within the organization must be seen as an ongoing process.

(d) **Leadership**

The fourth TQM principle is that the success of TQM is the responsibility of top management. The school teachers must establish the context in which students can best achieve their potential through the continuous improvement that results from teachers and students working together.

4 Principles of TQM

The key principles of TQM are given below (Martin 1993):

Management commitment	**Fact based decision making**
Plan (drive, direct)	Statistical process control
Do (deploy, support, participate)	DOE, FMEA
Check (review)	The seven statistical tools
Act (recognize, communicate, revise)	Ford 8D—team-oriented problem solving
Employee empowerment	**Continuous improvement**
Training	Systematic measurement and focus on CONQ
Suggestion scheme	Excellence teams
Measurement and recognition	Cross-functional process management
Excellence teams	Attain, maintain, improve standards
Customer focus	
Supplier partnership	
Service relationship with internal customers	
Never compromise quality	
Customer driven standards	

5 TQM Process

In an educational institution, the TQM process brings with it the commitment to quality, along with commitment to the employees, and the organization. All those who contribute to the system should be involved, with a clear understanding of the purpose. It is an approach to improve the effectiveness and flexibility of the organization as a whole. The improved performance is directed towards satisfying cross-functional goals as quality, cost, manpower development, quality of work life, etc. These activities ultimately lead to increased students and employee satisfaction. The process to introduce TQM in colleges should generally have the following steps:

(a) Mission and Passion: The determination and announcement of mission statement is the first and foremost task on which the whole TQM will depend. There should be no or minimal conflict. Everyone should share a passion to move continuously close to the ideal vision.

(b) Administrator as a Role Model: It must also be clear from the side of administrator/Principal of the institution that he/she is committed to total quality. The commitment should be communicated in meetings with employees and students and must be practiced, i.e. it must be by word of mouth and by action visibly demonstrated. The process of pursuing this agenda should be continuous and never lost sight of.

(c) Environment Factors: The next step is identification of the factors of internal and external environment, which have a bearing on the institution building. These include factors affecting the work environment in the institution (proper cleanliness, lighting, teaching aids, projectors, computer labs, lab materials, canteen, sports, gardening, water, etc.) and factors helpful in image-building of the institution (industry–institution interaction, debates, conferences, seminars, public relation including media management, etc.). Outsiders should be involved for mutual benefit.

    Liberalization and globalization have set new trends in domestic and global competitive environment. This has led to a great disparity between what is taught and what is needed at the work place. Due to the socioeconomic, cultural, and technological transformation which has taken place during past decade, newer demands are being placed upon educational institutions. Educational system can effectively react to these internal and external challenges only when it emphasizes on total quality.

(d) Accountability: We have to develop the system in which every group (student, teachers, researcher, manager) is accountable to all other groups and members of each group are accountable to one another. Students should be accountable to teachers. We should develop a system in which teachers are accountable to students through instruction surveys and are accountable to management through self assessment and assessment of teacher by outside organizations and in which researchers are prepared to be assessed by outside agencies and funding agencies for their work. We have to develop a system in which managements are accountable for their work through assessment by accreditation process. Moreover all accountability at all levels has to be in terms of criteria laid down sufficiently in advance.

(e) Human Relations: There is a need to enhance quality in the whole setup, including the relationship. All individuals, small or big must be viewed as important human beings with physiological, psychological, social, and ego needs.

(f) Feedback: TQM is a continuous process. There is a need for continuous performance appraisal of all the subsystems as well as the system as a whole.

6 Implementation Principles and Processes

There are several barriers for implementing total quality in education:

(a) There is often a conflict between administration and academic functions. The two groups often form parallel worlds without a *shared* vision or mission for the school. A total quality approach requires that the two groups should work together to meet customer expectations; (b) Within the academic group, there is often too much division. Identity with the entire school must take precedence over subgroup identity; (c) Most schools have entrenched cultural practices and beliefs that may create resistance to change. (d) The concept of a customer may be difficult to adopt in an academic environment; and (e) The need for control, measurement, and feedback systems for the purposes of standardization is somewhat foreign to academic environments. There may be the fear of stifling creativity.

To successfully implement a total quality approach to education we must do the following:

(a) Obtain COMMITMENT to total quality from the relevant authority. (b) Recognize a school as a system with interacting subsystems, namely, a social/cultural subsystem dealing with human interactions and motivation. (c) Identify all the customers and stakeholders; (d) Develop and communicate throughout the school a shared vision and mission that reenforce the needs of its customers; (e) Analyze the behavior of the school; (f) Develop goals and objectives consistent with the vision and mission; (g) Study the impact of each major process on the ability to meet the school's goals and objectives; (h) Develop measurement and feedback systems for each major process (such as curriculum development, student intake, teaching, etc.); (i) Form cross-functional teams to improve major processes. Ensure that all the customers of each process are directly involved in the improvement effort; (j) Train all teams in techniques consistent with the nature of their activity; (k) Implement systems to hold the gains that are made; (l) Document all improvement exercises; (l) Repeat Steps narrated before.

A preliminary step in TQM implementation is to assess the organization's current reality. Relevant preconditions have to do with the organization's history, its current needs, precipitating events leading to TQM, and the existing employee quality of working life. If the current reality does not include important preconditions, TQM implementation should be delayed until the organization is in a state in which TQM is likely to succeed.

7 Conclusion

Quality education is all about systems that lead to good academic culture, excellent academic result, progressive and adaptive management, clear and transparent administration, prominent profile of outgoing students, and, above all, review and modification of inputs. All stakeholders have a prominent role to play. TQM in educational institutions is the need of the hour. It must be tried in colleges/ Universities for maximum performance of the students and the employees. It must also be implemented for institutional image-building. It has become all the more necessary with the entry of the private sector in education in a big way. TQM will help achieve excellence, which only can guarantee the survival of institutions in a highly competitive world, with ever decreasing subsidy in the education sector.

References

William Edwards Deming (20 October '1900–20 December '1993) was an American Statistician, Professor, author, lecturer and consultant. He is perhaps the best known for his work in Japan.

John Jay Bonstingl (1985) International consultant, author, and speaker on Quality Learning and Improvement Columbia,

Martin, L. (1993). "Total Quality Management in the Public Sector," National Productivity Review, 10, 195–213.

Part V
Quality in Healthcare

Sharon Tay, Singapore

The quality of healthcare processes is affected by patient care process, hospital operation, and the people involved in designing and implementing these systems. Thus, human factors play a major role in ensuring quality and safety in an inherently complex healthcare system.

Competencies of healthcare professionals (doctors, nurses) are crucial to the delivery of quality care. Evaluation of the effectiveness of medical education is being researched in the following paper.

Patient care process can be complicated and involves multiple healthcare professionals. In the case of home care, more patients are opting for better continuity of care following their discharge from hospitals. Success of home care involves assigning the right mix of healthcare professionals with appropriate skills to render treatment while visiting patient's home. Evaluation of patient care and an effective and efficient method to plan resources for home visits are the subject of research in Asia healthcare industry as in the papers of this Chapter.

The two common incidents in the hospital setting are patient falls and medication error. These incidents may result in extended stay in hospital and additional costs to the health system. On the prevention side, Fall Risk Checklist is developed and its effectiveness is evaluated as shown in the paper of this chapter. When the incident happens, an effective investigation flow is needed, and one way is to do a task analysis to identify specific problem of that process method as presented in another paper of this Chapter.

About the Author

Ms. Sharon Tay has been actively involved in various quality functions in both the manufacturing and healthcare industries for more than 20 years, focusing on implementing and maintaining quality standard systems since 1993. She has work experience in managing accreditation and licensing related work in the healthcare industry for the past 16 years, pioneering the implementation and maintenance of the integrated management system in the largest tertiary hospital in Singapore. She is an ASQ Certified Manager of Quality/Organisation Excellence and a MOM registered Workplace Safety and Health Officer. As part of her lifelong learning journey, Sharon has just completed her part-time Degree of Science in Human Factors in Safety.

Contact details:
Name: Sharon Tay
Designation: Manager, Quality Governance
Organisation: Sengkang Health
Email: Sharon.tay.l.k@skh.com.sg

A Study on the Falls Prevention System in an Acute Hospital

Masataka Sano and Masato Yoshida

Abstract Preventing falls is one of the clinical problem and causes patients' significant problem, since it may cause serious injuries such as an abrasion and fractures. An assessment score sheet for inpatients is one of the falls prevention system. In this paper, we focus on sheet for inpatients' fall risk. In this paper, we validate the appropriateness of the falls estimate in patients hospitalising. We reassessed the score of 84 cases which consist of 42 sets of the actual patient data for patients who have fallen and 42 sets of the data for those who have not from seven wards from October 2013. The results showed four items as greatly influence the falls. With these items, we calculated fall probabilities of each patients, and compared them of patients who have fallen and those who have not. Then, we could not recognise the difference between two groups since the number of people whose probability of fall is more than 60% are 34 cases in patients who fallen and 24 cases who did not fall. From the result of the analysis, we discuss it is not enough that nurses estimate the patient's fall risk only using the assessment in patients at the time of hospitalising. Instead, the importance of daily assessment is shown.

Keywords Preventing falls · Assessment score sheet · Inpatients' fall risk · Probability of fall · Importance of daily assessment

1 Introduction

1.1 Background

Much recent focus has been given to the system of medical safety, as it is currently being constructed by medical institutions. Falls can result in severe injuries for patients, such as fractures, and it is the second most common form of medical accident.

M. Sano (✉) · M. Yoshida
Tokyo University of Science, Tokyo, Japan
e-mail: sano.masataka@p.chibakoudai.jp

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_12

It has been suggested that assessment sheets be used as tools for determining patient fall risk in order to improve prevention and reduction efforts. When a patient enters a hospital, the nurse uses the sheet constantly. Although many hospitals use assessment sheets, satisfactory results have not yet been achieved; thus, nurses often doubt their effectiveness and do not tend to use them precisely.

1.2 Previous Study

In a previous study, Kato et al. (2013) collected 150,883 sets of assessment data and 2085 sets of accident data from between 2009 and 2011 for patients at Iizuka Hospital. They excluded improper data, and 109,474 sets of data for analysis, consisting of 1429 events, remained. Thus, they developed an assessment system that enabled them to assign a total risk grade and to set priorities among candidate countermeasures for each patient. The nurses in Iizuka Hospital used the sheet constantly, so they were able to collect large quantities of data for utilisation in analysis.

1.3 Purpose

Many Japanese hospitals use a sheet created by the Japanese Nursing Association, Morita et al. (2010), Nakagawa (2010) and Ogi and Iijima (2013) but we do not know how applicable this sheet is to other hospitals. However, the sheet cannot be revised, since each hospital would revise the sheet differently and would not consult the other hospitals after doing so. In this study, we revised this risk assessment sheet and confirmed its appropriateness for estimating the fall risks of patients at “A” hospital.

2 Grasping the Present Situation

2.1 Summary of the Sheet

A hospital is an acute hospital with about 500 beds. The sheet used as a reference was created in 2007, and has not been revised since then. It contains 10 assessment criteria and 19 items, but each item has multiple contents so the 19 items consist of 42 components, each of which has a score. The maximum total risk score is 41, and the minimum is 0 (Table 1).

2.2 Problems with Its Use

We interviewed the head nurses regarding the problems with the risk sheet. The interview focused on the demands of the nurse, the wording of the items, the usage

Table 1 A hospital's current fall risk assessment sheet (a portion)

Assessment	Options	Score	Total risk score		
			Hospitalisation		
Age	Over 65 years, under 9 years	2			
Previous history	Previous history of falling, previous history of fainting	2			
Sensory obstacle	Equilibrium, visual impairment, hearing impairment	2			
Physical disability	Leg or waist weakness, muscle weakness	3			
	Paralysis, paresthesia, dysarthrosis, contracture	1			

of the sheet, and how well it identified the risk factors for falls. As a result, the nurses wanted to reduce the number of items and how the items were worded, such as whether the contents of the item were connected with "and" or "or." They reported it as difficult to easily judge vague expressions; for example, "the distance to the restroom is long." There were 12 items that they could not easily assess when the patient first enters the hospital. When patients are first admitted, nurses are supposed to use the risk sheet constantly; however, the nurses in the present study reported not using it after the first point of admission. They did note, however, that walking without assistance and without having to call for a nurse qualified as risk factors during hospitalisation.

2.3 *Validity of the Sheet*

We collected the scores of patients who had fallen during a specified period in order to assess the risk sheet's usefulness. We analysed scores of 75 patients from April to July 2013, although the scores of eight patients were not recorded (Fig. 1). As a

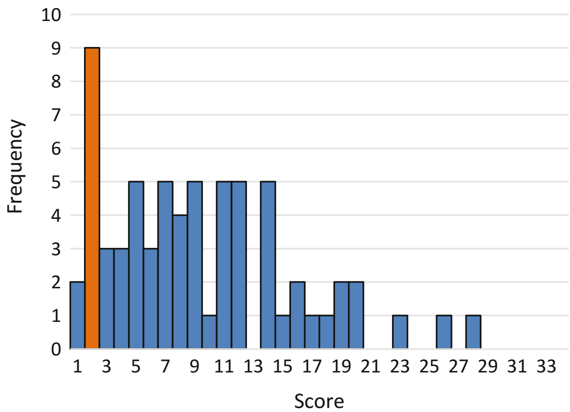


Fig. 1 The distribution of risk scores for 75 patients

result, 56 patients (84%) were classified as Risk I (having scores of under 15), which was the lowest fall risk level. When we checked this in detail, patients who had scores of 2 made up the majority. Thus, we reassessed 10 patients to validate the accuracy of our assessment (number of patients with scores under 5:7; number of patients with scores over 13:2; number of patients who had not fallen: 1). The scores of 4 patients who had scored below five rose over time, so we judged it necessary to use these scores to reassess and revise the risk sheet.

3 Methods and Reassessment Results

3.1 *The Reassessment Format*

We used 42 responses, those which concern falls on detail. We changed the expressions of two items and added two more that had influenced fall risk in a previous study (Warashina 2003). The timings of the reassessment were at hospitalisation, discharge from the hospital, and when a fall occurred (Table 2).

3.2 *The Decisions About Wards and the Number of Reassessments*

We limited number of objects with the data from April 2011 to June 2013 because reassessment took too much time. Specifically, we limited this to seven words. We reassessed every word by questioning 6 patients who had fallen and who had not (82 patients in total). We requested head nurses each word for reassessment with medical records and stories of nurses.

Table 2 A portion of the reassessment

[For patients who have fallen]

Day						Ward
Hospitalisation	7/ 19	Fall	7/ 24	Discharged	7/ 26	C

Reassessment sheet		Timing				
		Hospitalisation		Fall		Discharge
		Before reassessment	After reassessment	Before reassessment	After reassessment	—
Age	①Over 65 years / under 9 years					
Previous history	②Previous history of falling	✔	✔	✔	✔	✔
	③Previous history of fainting			✔	✔	✔
Sensory obstacle	④Equilibrium					
	⑤Visual impairment					
	⑥Hearing impairment					
Physical disability	⑦Leg or waist weakness		✔			
	⑧Muscle weakness		✔			
	⑨Paralysis					
	⑩Parasthesia					
	⑪Dysarthrosis, contracture		✔			

3.3 The Result of the Reassessment

The deflection of the score's distribution decreased before and after the hospitalisation reassessment, and the centre of the distribution skewed to the right. However, there were no patients classified as Risk III (i.e., a score of over 31). We confirmed some items with a few applicable patients.

4 Selection of Items

4.1 Methods

In this study, we analysed the timing as samples because this made it easy to detect significant differences and which timing we should use during the analysis was not clear. There were some missing data, so we marked those as "0." We confirmed correlation between many items with Statistical hypothesis testing. We did not combine the activity items, since the contents did not agree.

First, we considered whether we were able to reduce number of the items on the risk sheet. In the same fields, we investigated the differences between the 44 item components with chi-squared or Fisher's exact tests (with $p < 0.05$), and then integrated the significantly different items. We deleted options that did not involve patients. The remaining 32 item components were then analysed as explanatory variables of falling using logistic regression analysis Tango et al. (1997) (with significance at $p < 0.10$). Thus, we investigated the items that most influenced falls in 21 ways with timings, choosing at least one point in the assessment time as patients who have fallen compared to those who have not.

4.2 Results

As a result of having analysed the 21 items, we confirmed nine factors that affected fall with the data for discharge from the hospital, and when a fall occurred for patients who had fallen, and for hospitalisation and discharge from the hospital for those who had not fallen (Model 1) (Table 3). We detected the most items with those timing. We applied the assessment of the fallen patients to Model 1 and calculated fall probabilities for each patient. This resulted in the distribution shown in Fig. 2. As the probability of fall rises, so does the frequency of falls, when we use Model 1.

Table 3 Analysis result

Items	Estimate	
	Model 1	Model 2
Over 65 years/under 9 years	−2.102	−1.669
Leg or waist weakness (or) muscle weakness	1.375	1.505
Paresthesia, (or) dysarthrosis, contracture	−2.889	−2.809
Walking evaluation: independent with stagger	1.335	–
Using a sleep-inducing drug (or) a tranquilliser	1.041	–
Using a an antipyretic	−3.179	–
Going to the restroom frequently at night	−3.905	–
Tendency to act without pushing the nurse call (or) can't recognise or use the nurse call	3.331	–
Bewildered in unfamiliar environments	−3.338	–
Hearing impairment	–	1.722

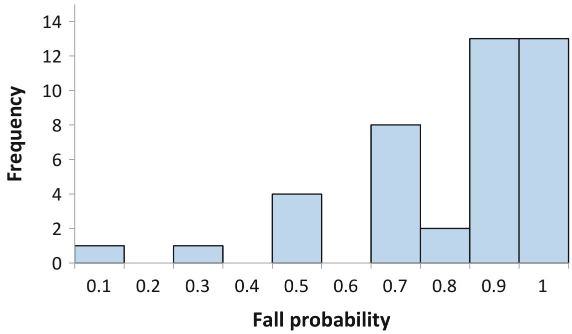


Fig. 2 The fall probability by reassessment data among the fallen (Model 1)

4.3 *Applying the Hospitalisation Reassessment Data to the Models*

To validate the effectiveness of assessment for the hospitalisation time period, we applied the reassessment data to Model 1. We found that, as the probability rose, the frequency decreased in patients who had not fallen. As such, we think that Model 1 is suitable for patients who have not fallen. However, the frequency did not show many significant differences from the probabilities for patients who have fallen.

Next, we investigated the factors that are more difficult for the nurses to judge when the patient is initially hospitalised. Using Model 2 (with data for hospitalisation, discharge from the hospital, and when a fall occurred for patients who had fallen, and for hospitalisation and discharge from the hospital for those who had not fallen), we applied it to the hospitalisation results and were able to confirm four factors (Table 3). With these, we calculated fall probabilities for each patient and compared these among patients who had fallen and those who had not. We found

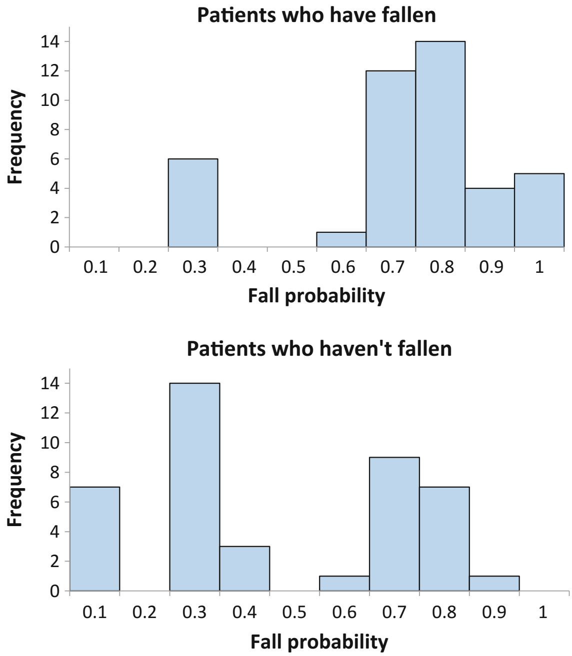


Fig. 3 Fall probability according to reassessment data among hospitalised patients (Model 2)

no significant differences between the two groups: only 34 patients who had fallen and 24 patients who had not had fall probabilities of over 60% (Fig. 3).

Thus, it is not sufficient for nurses to estimate a patient's fall risk only by using the patient assessment from the time of hospitalisation. Instead, the importance of daily assessment is shown.

5 Consideration

First, the nurses in A Hospital said that they wanted to promote use of the risk sheet and remove some of its content. Most patient risk scores at a certain time were low, even though they had fallen. This suggested that nurses had not followed the sheet exactly, and this was confirmed by a reassessment. To create a risk assessment sheet suitable for every hospital, we need to confirm the nurses' use of these sheets.

We found a number of factors that strongly influenced fall probability, but only among patients who had been hospitalised. Furthermore, we did not find a significant difference in fall risks between patients who had fallen and those who had not. Thus, it is not enough for nurses to estimate patients' fall risk only by using patient assessment sheets at the time of hospitalising. Instead, our results point to the need for daily assessment.

6 Future Issues

We evaluated the necessity of revising risk assessment sheets using previously collected data, and found that we need to re-evaluate with newer data. We need to make constant assessments and compare changes in fall risk over time to completely validate the benefits of periodic risk assessment.

Nurses can get a good idea of patients' fall risks using the risk sheet described herein. We hope that nurses will be able to predict the places and the timings of potential falls using the data obtained from the risk sheet. Furthermore, it is necessary to let the fall prevention system progress and evolve naturally, as medical professionals glean increasingly more information about their patients to reduce an accident from the origin.

References

Kato Shogo et al., (2013); "An Assessment System for Preventing Patient Falls based on Repeated Events Analysis", ANQ Congress 2013.

Warashina Erika (2003); "A Study on a fall accident in hospitals (in Japanese)", Journal of the Society of Medical Management, 4(1), 145.

Emiko Morita et al., (2010); "Revision of an Assessment Score Sheet for Fall Prediction and Examination on Inter-Rater Agreement among the Nurse Evaluators", The Journal of the Japan Academy of Nursing Administration and Policies 14(1), 51–58.

Youichi Nakagawa, (2010); "Incidence and Consequence of Falls among Stroke Rehabilitation Inpatients in Convalescent Rehabilitation Ward: Data Analysis of the Fall Situation in Multi-institutional Study", The Japanese Journal of Rehabilitation Medicine, 47(2), 111–119.

Yuko Ogi, Sachiko Iijima, (2013) "A Literature Review on the Effect and How to Combine the Fall Preventive Interventions with the Patients' Risk of Falling", The Journal of the Japan Academy of Nursing Administration and Policies, 17(2), 116–125.

Toshiro Tango et al, (1997); "Logistic Regression Analysis - Statistical Analysis with SAS – (in Japanese)", Asakura-Shoten.

Authors Biography

Masataka Sano is an associate professor at the Faculty of Social Systems Science, Chiba Institute of Technology. He earned a doctoral degree in Engineering in 2011. His research interests are in the areas of patient safety and healthcare quality.

Masato Yoshida is a master of science in Tokyo University of Science. He graduated from TUS in 2014. His research interests are in the areas of patient safety and healthcare quality.

A Study on the Measurement Method of Effect for HealthcareSafety Education

Takeyuki Goto, Masahiko Munechika and Chisato Kajihara

Abstract Medical technology has been consistently improving in recent years. With that, people are more concerned with medial safety and its quality. In medical service provided by hospitals, quality is greatly influenced by human knowledge and skills. Therefore, it is fundamental to improve these factors. Actually, at many hospitals, medical safety education is provided to employees. To conduct more efficient education, the effectiveness of the education must be evaluated, and the specific contents must be revised after its administration. It is said that effects of an education can be evaluated in terms of learning (knowledge and skill) and the degree of behavior modification and results. In evaluation, we should begin by evaluating learning, and then move to behavior and results. However, the methodology for measuring these has not been established. In this study, we analyzed the long-term findings in a hospital and implemented an evaluation model for medical safety education. Further, we designed the effectiveness measurement tool based on that model. We then propose a method for measuring the effects of health-care safety education. Using this method, medical workers can evaluate students' learning free of bias and extensive work after providing education. In addition, they can review the contents of education and provide sufficient feedback on the test results to students.

Keywords Educational evaluation · Health-care safety · Behavior modification · Skill · Knowledge

1 Study Purpose

Delivering safe medical service is an essential aspect of health care. The quality of medical services is influenced by human knowledge and skills. Therefore, many Japanese hospitals conduct educational programs on medical safety for their

T. Goto (✉) · M. Munechika · C. Kajihara
Waseda University, 51-14-12 3-4-1, Okubo, Shinjuku, Tokyo 169-8555, Japan
e-mail: noelgot326@ruri.waseda.jp

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_13

employees. Such programs help students acquire knowledge (i.e., gaining an understanding about a concept or a term) and skills (i.e., the ability to apply their knowledge and techniques). However, it is also important that the students are able to utilize the knowledge and skills that they learn in the workplace. Therefore, to examine the effectiveness of an educational program, it is necessary to assess the behavioral changes it brings about (changes in actions on the job). However, due to the lack of evaluation methods, it is not currently possible to review the contents of the program and provide sufficient feedback to the students. Therefore, in the present study, we propose a method to evaluate educational programs for healthcare safety.

2 Previous Research and Approach of This Study

2.1 Previous Research

Tsutsumi et al. (2007) assume that there is insufficient knowledge on technology for measuring the effects of education. Further, they state that it is difficult to measure effects of education in a given field, as deep, specialized knowledge, such as the knowledge of statistics used for ascertaining knowledge and reliability of a learning theme, is necessary. The situation is similar with medical safety education.

To grasp the present conditions of the measurement of education effects in medical safety education, we conducted an actuary investigation hearing regarding impeller medical care security of 10 hospitals, which participated in a 2013 meeting for a study on quality and security of first aid medical care. Ultimately, the hospitals indicated that, "We conducted medical safety education, but evaluation is a problem." Further, they said "We did the understanding degree investigation, but we wish we could evaluate the action transformation." From this result, at many hospitals, medical safety education is carried out without measurement of its effect, and thus, a review of the education and feedback from students attending the lecture were lacking.

Regarding a trend in measuring education effects among general companies, according to the Association of Training and Development (ATD, formerly ASTD) (Human Value 2004), the models that attracted attention from the 1990s through the 2000s were the four phases of Kirkpatrick's model (Tsutsumi et al. 2007), the return on investment (ROI) model of Philips (1999), and the success case method (Brinkerhoff 2009). Above all, Phillips' ROI model (cf. Table 1) is a widely used effect measurement model. However, it is an inflection example of these models, and the concrete methodology is not shown. It may be said that the company builds an evaluation method in accord with its training methods while receiving the support of the consulting company.

As a result of having investigated a study in conjunction with medical safety education, we determined that many studies have been conducted in accordance

Table 1 The five-level framework suggested by Phillips (1999)

Level	Impact of education	Details
1	Reaction and planned action	Reaction of the student to attending a lecture
2	Learning	Knowledge and skills
3	Job application	Changes in actions on the job
4	Business result	Performance and result of the work
5	Return on investment	Effect of the investment

with the medical safety education program, and there are few studies on effect measurement. Regarding the study of the medical safety education program, for example, Kajihara (2012) defined education as the fostering talented individuals who can practice medical safety management in terms of medical safety education. She derived the medical safety education items that should be explained by education programs with reference to a way of thinking and the activity of quality management systems (QMSs) in the industry, and she arranged these items in a table. In this way, each hospital was able to establish and review the medical safety education system.

As in other studies, there is a standard text development for the promotion of medical safety (Japanese Nursing Association 2013), a study on simulation education (Takahashi 2014), a study on policies toward security awareness in the health-care setting (Ebara 2014), and so on. It may be said that various studies have been conducted on enforcement methods of medical safety education.

Regarding the study of the effectiveness of medical safety education, we include a study on the measurement of the medical safety consciousness and a study of the attitude survey with regard to specific training. For example, Takemura (2011) indicated that the tools that evaluated medical safety awareness, such as the Safety Attitudes Questionnaire or the Hospital Survey on Patient Safety Culture, were developed mainly in the United States. Further, they utilized those tools and observed variability of medical safety consciousness and subsequently analyzed this variability. In addition, Inaga (2007) suggested a questionnaire survey method as a measure of the effectiveness of a medical safety self-prevention training program from a psychological viewpoint. In addition, Goto et al. (2013) discussed the methodology for effect measurement at level 2 ("Learning and knowledge" in Table 1).

The purpose of this study was to establish the method for measuring effectiveness of a medical safety education and to improve the educational program and the feedback provided to students attending the lecture. It may be said that the tool that can be used as an attitude survey of the medical safety culture can help to generalize to the degree of recognition of medical security, and to evaluate this as well. However, using the result of the generalization evaluation, it is difficult to improve the educational program and to perform the feedback in detail for the students attending lectures. In addition, most questionnaire surveys that measure effective training methods include an unclear development process. Therefore, it is difficult to determine whether a logical technique for developing questionnaire items was used. In addition, we cannot ensure the reliability or validity of the evaluation. Moreover, in Goto et al., only level 2 in Table 1 is considered to examine the measurement of the method's effectiveness. The purpose of our study is to introduce medical safety education and to ensure the security of patients, and thus it is not solely intended to acquire knowledge.

We should grasp all the effects of Table 1, levels 1–5 to achieve the final purpose, namely, ensuring patient security. To obtain such an effect, it is desirable to improve the education curriculum and contents. In addition, not only did we seek to improve education, but also we should provide appropriate feedback and manage motivation. To construct an environment in which education contents can be effectively used is essential for students attending the lectures.

Our aim then, based on the above-mentioned investigation, is as follows, it is useful for a hospital which conducts medical care safety education, there should be a method to improve feedback toward improving the program, and students attending a lecture of the educational program should conduct this feedback after performing the measurement of the effectiveness of the medical safety education.

2.2 Approach of This Study

In this study, we examined the measurement method based on two educational programs. The Process-Oriented Analysis Method (POAM) (Sano et al. 2013) and Kiken Yochi Training (KYT) (2011) were carried out in Hospital A and B, respectively. We first examined the effectiveness of "Reaction and Planned Action "and "Learning," and investigated the opinion of the students who attended a lecture. Specifically, we explored the students' opinions on their present knowledge and acquisition level, as affected by their participation in the POAM program. Based on those results, we then examined the aim of a program with reference to the action model suggested by Marzano and Kendall (2013) and make the evaluation criteria of the degree of action transformation. Subsequently, we developed a questionnaire based on the evaluation criteria and examined the effectiveness of "Job Application." Based on these findings, in the present paper, we propose a method to examine the effectiveness of medical safety education.

3 Examining the Effectiveness of the Proposed Method to Evaluate Medical Safety Education

3.1 The Measurement and Attitude Survey on Level 1 and 2

We examined the effectiveness of Level 1 and 2 using the method proposed by Goto et al. (2013). Thus, we evaluated the level of understanding of 30 student nurses who attended a lecture on POAM that was carried out with an X-disease ridge of Hospital A, in June 2013. In addition, we attempted to introduce POAM into the whole ward in the X-disease ridge, within one year, during 2013. To examine Level 1, we used a form to evaluate the degree of understanding of the course contents in students attending a lecture, their degree of interest, and their evaluations of how the lectures were conducted, in five phases each. As for the evaluation of the degree of understanding for Level 2, we addressed two aspects: knowledge and skills. Thus, we evaluated their degree of acquisition of skills on a five-point scale, and that of knowledge on a 14-point scale, where a total score of 19 points indicated a perfect score. Similar to the evaluation of Level 2 was conducted thrice, in June, September, and December, after the completion of the program. The findings from the evaluation of Level 1 are presented in Table 2. Figure 1 presents findings of the evaluation of Level 2 for six people, to illustrate two or three.

As shown in Table 2 the degree of intelligibility, usefulness, and degree of interest tended to be high, perhaps because the evaluation was conducted soon after the program. Further, as observed in Fig. 1, the evaluation scores improved in September, as compared to June, which indicates that most of the students attending the lectures implemented positively the POAM in the ward. On the other hand, the scores decreased in December. To explore this further, we obtained the opinion of the manager of the X-disease ridge to examine on-site activity. In addition, we

Table 2 Results from the evaluation of Level 1

	Intelligibility	Usefulness	Degree of interest	How the lectures were conducted
Student A	4	4	4	4
Student B	4	4	5	5
Student C	3	4	4	4
Student D	4	4	5	3
Student E	4	3	4	4
Student F	4	4	5	5

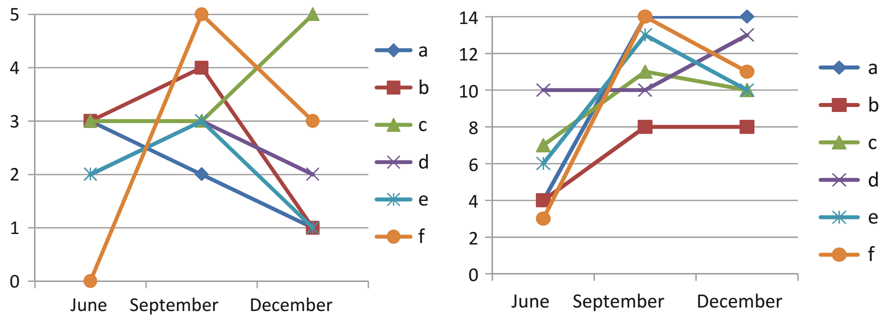


Fig. 1 Results from the evaluation of level 2 (*Left* Knowledge; *Right* Skill)

interviewed a student who attended the lectures and provided the highest score in the evaluation of Level 2. We found that their consciousness changed by education. It is thought that a change in the consciousness, such as in the degree of interest in the student attending a lecture, or the degree of usefulness, influences the degree of acquisition, and the degree to which learning is transformed into action.

3.2 Examining the Method to Evaluate Level 3

3.2.1 Examining the Use of the Action Model in Education

Based on the findings presented in Sect. 3.1, we hypothesized that a change in the consciousness of the student attending a lecture influenced an action by affecting the degree of acquisition. This suggests that, to evaluate a program that aims to change the consciousness of the student attending a lecture, it is necessary to measure the degree to which the learning is transformed into action, thus implying the need to develop a method to evaluate the same. Therefore, I utilized the action model suggested by Philips (1999) (see Fig. 2).

In this model, it is suggested that when a student who attends a lecture faces a new problem, the autonomous system closely evaluates the importance of the problem, the feelings, and the effectiveness of education. If it is judged that one should work on a problem, the metacognitive system sets an aim for the solution of a problem while the cognitive system manages the knowledge derived from the

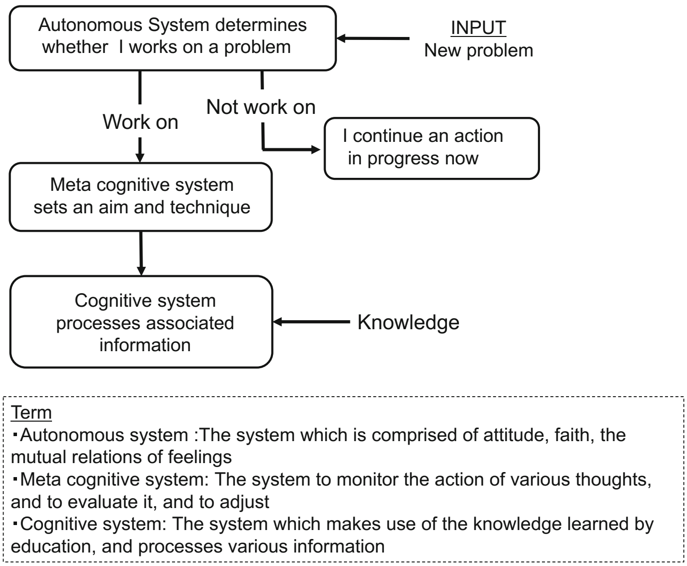


Fig. 2 The action model suggested by Masataka Sano et al. (2013)

contents of the educational program, which is saved in the student's memory. Section 3.1 revealed that the change in the consciousness occurs during the process of education, as well as during on-site activity after the completion of the program.

Therefore, it is desirable to evaluate Level 3 based on the autonomous system before and after the education, to measure the degree to which the learning is transformed into action and the time of this transformation. Thus, we used the change due to education in the consciousness of the student attending a lecture as the goal or end point of the action model, and evaluated the process of change by using the action model in the beginning and ending. The model that we used in the present study is presented in Fig. 3.

On attending an educational program, an autonomous system simultaneously undergoes changes and saves the knowledge to memory. When a student decides to work on a problem based on the autonomous system, she establishes an aim, and the cognitive system processes the problem. It is believed that this process leads to a change in the action. Thus, based on this model, for an action to be transformed, it is necessary for the following requirements to be met:

(a) The student's will to attend a lecture increases after education.
(b) An aim and a plan need to be established.
(c) The student can handle a new problem on-site.

Thus, if these three requirements are met, there may be a high level of action transformation. Therefore, in order to examine the level of transformation in an action, we studied the level of these requirements.

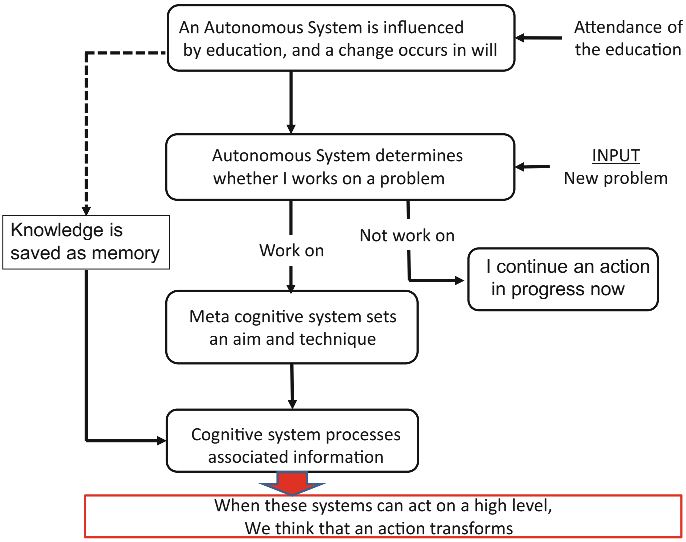


Fig. 3 The action model used in this study

3.2.2 Examining the Evaluation Criteria

Based on the action model explained in Sect. 3.2.1, we determined that bringing about a transformation in the action is the goal of POAM. Further, we specified the content of the knowledge to be provided through the educational program and the mental procedure involved, and explained the content of each system in the action model. These results are presented in Table 3.

For example, the importance, effectiveness, feelings, and will are examined in the autonomous system, and it is desirable that they are affirmative evaluations. Thus, I could derive four outcomes for each aim. This process can also be applied to other items.

3.2.3 Design of the Survey Questionnaire

We designed a survey questionnaire to evaluate the evaluation criteria we derived. This evaluation was conducted by the students who attended a lecture (Table 4).

With reference to the items of this questionnaire, based on the inputs of the medical safety manager of Hospital A, we created an item each, on will, goal setting, and the degree of understanding. We made each evaluation item to evaluate with five phases. We carry out an evaluation using this item.

3.2.4 Design of the Evaluation Standard List

We constructed an evaluation standard list to evaluate the on-site influence of education contents. It is a characteristic of medical safety education that penetration activities are carried out in hospitals to consolidate education contents to students attending lectures after education. Thus, the evaluator can collect activity results as an evaluation document, and it may be said that this leads to the evaluation of the on-site utilization to evaluate activity results. Therefore, we collected documents from the POAM analysis activity performed in A Hospital, and constructed an evaluation standard list to evaluate them.

First, we grasped contents of the POAM analysis activity through an investigation of an impeller activity. Next, we went into detail about the activity contents from the viewpoint of standard enforcement procedures and derived the arrival target of each enforcement procedure based on the argument with the POAM teacher as a check item. The results are shown in Table 5.

We set a marking standard of some check items to raise the objectivity of the evaluation. For example, for item (21), if measures were not written along 4W1H, zero points were given. Therefore, we could evaluate the utilization situation without depending on the evaluator.

Table 3 Goal of the POAM

Aim classification	Contents	Evaluation item
Autonomous system	Examining the importance	I realize that it is important to analyze an accident in terms of the process
	Examining the effectiveness	I believe that I can apply the POAM
	Examining one's feelings	I have affirmative feelings about the POAM
	Examining one's will	I am willing to use the POAM
Metacognition	Realizing the aim	I can determine an aim to utilize the POAM
	Process monitoring	When I utilize the POAM, I can understand for how long an aim is accomplished
	Monitoring the accuracy	I can articulate when I am confused about the POAM
	Accurate monitoring	I can demonstrate, with evidence, for how long I have known to use the POAM
Application of the knowledge	Decision making	I can make decision using the POAM
	Solution to the problem	I can solve the problem using the POAM
	Experiment	I can develop a hypothesis using the POAM and can inspect it
	Investigation	I can use the POAM for investigating a future event, based on the past
Analysis	Comparison	I can identify the differences between the POAM and other analysis techniques
	Classification	I can classify the POAM in the technique used to analyze the accident
	Error analysis	I can compare the POAM by a thinking that another person showed and can point out a mistake
	Generalization	Based on my learning, I can speak about the POAM
	Realization	I can generate about the POAM and, under a certain condition,I can provide evidence for the same
Understanding	Unification	I can compare the characteristic of the POAM by that when I do not use the POAM
	Symbolization	I can depict a main element of the POAM in a figure
Extraction	Recognition	I can check whether the words that it is process-oriented exactly or not
	Reproduction	I can explain the purpose of the POAM
	Practice	I can apply the analysis procedure suggested in the POAM

Table 4 List of items in the questionnaire on the POAM

	Contents
Q1	I realize that it is important that I practice a process point in accident analysis
Q2	I am sure that I can utilize the POAM well
Q3	I am aware of the weak points in the analysis technique of the POAM
Q4	I usually willingly use the POAM for analysis
Q5	I set an aim for myself, to utilize the POAM, and make a plan for the same
Q6	When I analyze using the POAM, I can understand how my aim was accomplished
Q7	With reference to the POAM, I can indicate the contents that I understand, and those I do not
Q8	I can make a decision by thinking things by a process
Q9	I can solve a problem by improving the way I fulfill everyday duties
Q10	I can indicate the errors in the analysis conducted by another person, and modify them
Q11	I can express my thoughts about the POAM
Q12	I can explain a purpose and a process point of the POAM to another person
Q13	I can carry out the POAM according to an analysis procedure

Table 5 Evaluation standard list of the POAM analysis activity (part A)

Process of the activity	Evaluation items	Check item	No.
Choice of the accident report	Whether they choose an example suitable for POAM analysis or not	Whether they choose the example of the process type instead of a non-process type	(1)
		Is it written in accordance with the standard in the stage of the instructions tray, preparation, and enforcement?	(2)
…	…	…	…
Drafting of measures	Whether they can perform measure drafting or not	As for the measures, is it written about a process, such as standardization and error proofing?	(20)
		Are the measures written in detail?	(21)
		Can the measures solve the problem that they gave in the factor analysis?	(22)
		Are the measures to be carried out decided under the agreement of all members?	(23)

4 A Suggested Method to Evaluate the Effectiveness of Medical Safety Education

Based on the findings of this study, we suggest a method to evaluate the effectiveness of medical safety education.

Step 1 **Grasp of the present state of the will of the student for the education contents**
We constructed evaluation items for the medical care safety education based on the guidance of the education target classification of Marzano. The first questionnaire survey was designed based on those evaluation items.
We then conducted the questionnaire survey including the item about the consciousness of the effectiveness, importance, and the state of feelings before education.

Step 2 **The measurement of the degree of understanding of student knowledge**
We designed the degree of understanding investigation using the construction method suggested by Goto and others, and carried it out after education. We conducted an investigation regularly in order to observe the degree of knowledge acquisition and the skill of the student attending a lecture in chronological order.

Step 3 **Measurement of the influence of the education contents on the spot**

3.1 We carried out the second questionnaire survey after education.
3.2 We constructed an evaluation standard list based on the activity process on spot.
3.3 We conducted an evaluation by the evaluation standard list and the questionnaire survey regularly to evaluate changes in the utilization situation.

This method allows for the evaluation of the degree of action transformation, as well as the degree of knowledge acquisition and skill in the student attending a lecture.

5 Examining the Effectiveness of the Goal

5.1 Examining the Effectiveness of Items in the Questionnaire

To examine the effectiveness of the investigation on the goal, as explained in Sect. 3.2.3, I conducted a survey on 15 members of the accident analysis team of Hospital A, using a questionnaire. The accident analysis team utilized the POAM to analyze a real accident example from the previous year and drafted rules for the ward. In addition, in this section, we also present the findings of the interview of the manager of the team and inspect the effectiveness of the contents of the evaluation.

Subjects of the survey: 14 link nurses of Hospital A and 30 nurses of the X-disease ridge. Investigation form: Questionnaire survey

The completed questionnaires were received from all the link nurses (100%) and 63% of the nurses of the X-disease ridge. They rated each item on a five-point scale, specifically, "not applicable," "applicable to some extent," "was not which," "largely applicable," and "totally applicable." Their responses were scored from one to five, respectively, and means were calculated. The related findings are presented in Table 5. It is also important to note that a link nurse and a nurse of the X-disease ridge worked together on the analysis of the POAM, for one year.

It can be observed from Table 6 that a low rating, specifically, that below three points, was predominant, with very few items depicting a score of four points or above. Specifically, both, link nurses as well as X-disease ridge nurses exhibited a higher score on Q1, which was related to the importance of the POAM. Thus, the acknowledgement of the importance of conducting an accident analysis seems to be apparent. On the other hand, with reference to Q3, on the awareness of the weak points of the POAM, the rating of the X-disease ridge nurses was particularly high.

Then, based on the responses of the link nurses and the nurses of the X-disease ridge, we performed a correlation analysis to examine the relationship between the items of the questionnaire. These results are presented in Table 7.

With reference to Table 7, the following point needs to be noted:

- The correlation between the items on the influence of education contents on the understanding (Q11–13) and the item on weak point awareness (Q3) (this section is highlighted in green in the table).

With reference to this point, it is thought that "the weak point awareness" influences the understanding at the processing level, specifically, in the autonomous system (the system that closely examines will). In other words, it is thought that if an individual has a strong awareness of her weakness for POAM, it will be difficult for her to understand the POAM, and to utilize it on-site.

When the manager of Hospital A was questioned about the cause of the weak point awareness, she reported that it may be influenced by the lack of power to think things by a process, which is characteristic of nursing education, and to solve

Table 6 Mean on each item of the questionnaire

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13
Link nurses	4.36	2.50	3.86	3.21	2.79	2.57	2.86	2.62	3.08	2.64	2.71	2.57	3.36
X-disease ridge nurses	4.26	2.53	4.32	3.33	2.17	2.28	2.28	2.72	3.22	2.26	2.32	2.21	3.05

Table 7 Correlation analysis on the items of the questionnaire

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13
Q1													
Q2	0.068												
Q3	−0.220	−0.304											
Q4	−0.052	0.028	−0.293										
Q5	−0.142	0.210	−0.122	0.251									
Q6	−0.128	0.302	−0.215	0.188	0.739								
Q7	−0.020	0.314	−0.393	0.190	0.653	0.442							
Q8	0.054	0.269	−0.250	0.364	0.439	0.446	0.352						
Q9	0.107	0.205	0.018	0.167	0.296	0.317	0.204	0.370					
Q10	0.328	0.189	−0.243	−0.157	0.293	0.207	0.346	0.265	0.335				
Q11	0.010	0.323	−0.429	0.003	0.391	0.475	0.362	0.327	0.188	0.449			
Q12	0.187	0.137	−0.559	0.035	0.152	0.286	0.282	0.235	0.092	0.517	0.591		
Q13	0.082	0.116	−0.506	0.276	0.037	0.191	0.217	−0.199	−0.094	0.063	0.151	0.330	

a problem. We knew that in the nursing school, there were few opportunities to understand why things are taught the way they were. With reference to the POAM, the students learnt the procedure of analysis even if they did not know its purpose. Owing to this, it is thought that thinking about things by a process is difficult for them such as POAM investigating a thing without the correct answer.

We agreed with the manager on this aspect.

5.2 Inspection of the General Purpose of the Suggested Method

We applied the suggested method to the training of KYT carried out in B Hospital. In the training, the purpose was for students to acquire the ability to practice KYT on site. The subjects of the evaluation using the suggestion method were 13 safety managers and some students attending KYT lectures who had learned KYT.

In the evaluation, we administered a questionnaire survey before and after the training, and conducted an evaluation of the influence of education contents using an evaluation standard list. We also conducted a degree of understanding investigation composed of 10 questions (2 points per question) immediately after the training. The questionnaire findings and the result of the degree of understanding investigation are shown in Figs. 4 and 5.

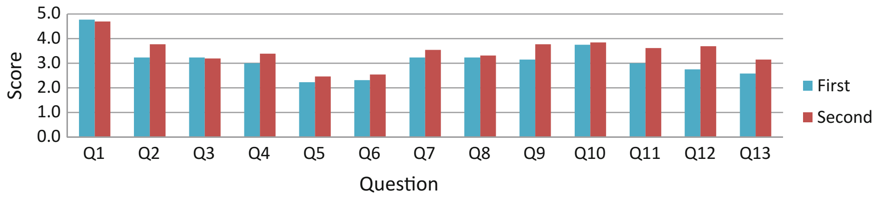


Fig. 4 Questionnaire findings before and after the training

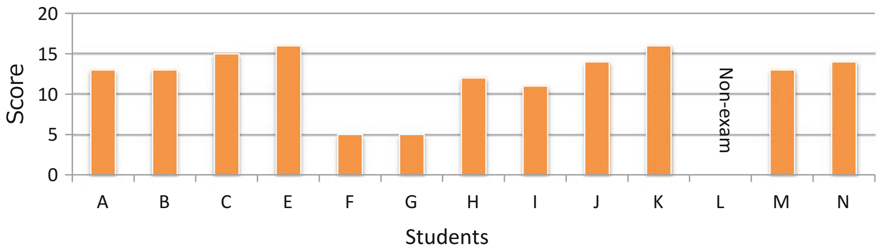


Fig. 5 Results of the degree of understanding investigation

Figure 4 demonstrates an improvement trend in the questionnaire results in most items after the training. Further, from the degree of understanding investigation, we were able to distinguish students with low scores from those with high marks. In addition, regarding the scores of the evaluation standard list, 19 points indicated a perfect score, and the average score increased from 10.5 points (standard deviation: 2.32) before the training to 13.3 points (standard deviation: 3.86) after the training.

Regarding these results, three students who had high scores in understanding the investigation, the evaluation standard list, the item about students' will for education may be able to best utilize the education contents.

Accordingly, we compared these with the questionnaire results of three students who enforced the knowledge in a medical safety promotion room of B Hospital after the training. These students reported changes in on-site action: "The critical eye for danger improved, and knowledge was deepened as a function of KYT, and I was able to teach it to a younger student" and "I divided it into teams and taught KYT and was able to publicize it." Therefore, it was suggested that, if the will of the student to attend lectures improves after the training, the degree of the influence of KYT rises. We were then able to apply a suggestion method to measure the effect of KYT training.

6 Contributions of the Study

Previous studies used conventional methods to evaluate the degree of acquisition of knowledge and skills. Additionally, there was no method to evaluate the degree of transformation of actions of the student attending a lecture after the completion of the program. In this study, we focused on the influence of education on the change of consciousness in the student attending a lecture and derived the goal of the transformation of action by using an action model. We also suggested a method to evaluate the degree of transformation of action by regularly measuring the same after the education. Previous studies on the measurement of the effectiveness of education did not demonstrate the process of deriving a goal of examining the degree of transformation of action. To identify the goal in this study, we utilized a framework of the education target taxonomy used in the action model; therefore, we could explain it in detail.

7 Conclusion and Future Research

In this study, we examined a method to evaluate the degree of transformation of action after education, in addition to a method to measure the knowledge of the student attending a lecture and the degree of acquisition of a skill, and suggested a method to evaluate the effectiveness of medical safety education. Future studies can explore the utility of the suggested method in evaluating other educational programs.

References

Tsutsumi. U, Aoyama. M, Kubota. S (2007): First Education Effect Measurement, Nikkagiren-Shuppansya

Human Value (2004): "ASTD2004 Report"

Jack J. Philips (1999): "Education training effect measurement handbook," Efficiency Management Center, Japan

Robert Brinkerhoff (2009): 「The Success Case Method: Find Out Quickly What's Working and What's Not」,Berrett-Koehler Publishers

Chisaato KAJIHARA et al. (2012): "Proposal of Educational Contents for Healthcare Safety", Quality, 42[3], 106–117

Japanese Nursing Association (2013): "Standard text for medical safe promotion", Nonprofit foundation Japan Nursing Association

Makiko TAKAHASHI (2014): "Improve the skill that the school teacher teaches with a learner by simulation education", NURSING business 8(12), 1122–1125

Futoshi EBARA (2014): "Raise awareness of security and prevent an error, by visualizing a characteristic by a safe action investigation and setting goal that fitted each person," NURSING business 8(12), 1134–1138

Toshihiko TAKEMURA (2011): "Study on Changes in Medical Safety Awareness-Comparison of time series changes in Japanese national university hospital-", Japan Society for Healthcare Administration

Kyoko inaga (2007): "Measurement methods of effect for education for Prevention of medical accident" Kitakyusyu municipal university, (21), 145–166

Takeyuki Goto et al. (2013): "A study on the measurement method of effect for healthcare safety education," The 11th ANQ Congress, pp. 210

Masataka Sano at el. (2013): "Analysis of Medication Incident for Improvement of Medication Process", Total Quality Management & Business Excellence

KAJIHARA Chisato et al. (2011): "A Proposal of Designing Method of Kiken Yochi Training Sheets in Medication Incidents", Quality 41(3), 361–370, JSQC

Robert J. Marzano, John S. Kendall (2013): "Designs educational objectives" The Kitaoji bookshop

Author Biography

Takeyuki Goto is a student who is the second year of the master's program of Waseda University and he study management design. In 2012, he graduated from Waseda University. He made the study on Education for Healthcare Safety. And he will study for evaluation for behavior of students after education in hospital. He is a member of the Japanese Society for Quality Control (JSQC).

Patient Assignment and Grouping Problem on Home Health Care: A Case in Hong Kong

M.Y. Lin, K.S. Chin and K.L. Tsui

Abstract Home healthcare (HHC) service systems have been attracting researchers' attentions in recent decades. Many resources are involved in delivering HHC service, consisting of different categories of human resources (nurses, physicians, physiotherapists, social assistants, and psychologists), support staff, and material resources. One of the main issues encountered when managing HHC services is the patient assignment problem, i.e., assigning operators to admitted patients given some sets of constraints (e.g., the continuity of care). This paper addresses the patient assignment and grouping problem for HHC system, Alice Ho Miu Ling Nethersole Hospital (AHNH) in Hong Kong, which focuses more on assigning new-admitted patients in a period to operators with achieving the aims of workload balance and minimum delay visits, and meanwhile grouping the assigned patients for each operator based on the patients' geographical location and next visit due time to reduce the travel time and improve the efficiency of operators. This hospital suffers from ineffective manual planning methods to manage their human resource when providing home health care. In order to improve the quality of service and efficiency of operators, we model the patient assignment and grouping problem and will use heuristic method such as discrete particle swarm optimization, Tabu Search, genetic algorithm to obtain the optimal solution.

Keywords Home health care · Patient assignment and grouping · Markov chain model · Assignment policy · Heuristic methods

M.Y. Lin (✉) · K.S. Chin · K.L. Tsui
City University of Hong Kong, Kowloon, Hong Kong
e-mail: meiyanlin1990@gmail.com

K.S. Chin
e-mail: mekschin@cityu.edu.hk

K.L. Tsui
e-mail: kltsui@cityu.edu.hk

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_14

1 Introduction

Home Health Care (HHC) as an alternative to conventional hospitalization has been developed to reduce overall costs of healthcare systems, and improve the life quality of patients in recent decades. It will increasingly become an indispensable and integral part of healthcare systems in the future due to many factors, such as the aging of population, dramatic changes in the needs of those with non-communicable diseases, the rising healthcare cost, the increased focus on user-centred services, and the availability of technology-based healthcare products that enable home-based disease management. Because of the increasing demand of HHC and limited resources available during the delivery process, HHC providers find it challenging to operate healthcare delivery systems in an efficient manner. These operators, including Home Health Aid, Certified Nursing Assistant, Nurse (Licensed Practical Nurse, Registered Nurse), Therapist (Occupational Therapist, Physical Therapist, Speech Therapist), Medical social worker, skilled homecare expert, etc., are one of the most important resources as they are all equipped with appropriate skills for providing healthcare services to admitted patients. Patients admitted to a HHC system are divided into classes, such as palliative and non-palliative care patients. Each patient receives different cares from different categories of operators according to his/her health situation, and has a care profile that describes his/her specific needs of treatment, preferences (i.e., time and operators), home address, etc. After admitting new patients, several planning decisions need to be made or revised, such as patient assignment, home visit sequences/time of each operator, etc. All these decisions need to be made by satisfying a number of operational constraints and meeting several objectives, such as balanced workload among operators, maximal home visit times each working day without overload, maximal satisfaction of patients, etc.

In this connection, we focus on the patient assignment problem in a local HHC system, and propose an effective method for solving this problem. This paper is organized as follows. In Sect. 2, we will provide a critical review of existing models and methods related to patient assignment in the HHC context. In Sect. 3, the details of patient assignment problem in the local HHC system will be explained with logical justifications, and model will be built to formula this problem. Section 4 will present heuristic methods to solve this patient assignment and grouping problem. Finally, conclusion and future work will be carried out in Sect. 5.

2 Literature Review

The assignment problem consists of assigning personnel to visits (patients) in fair way with considering different features, such as the continuity of care and the uncertainty of patients' demands. If continuity of care is not considered, the assignment problem turns out to be an assignment of personnel to visits rather than patients. Uncertainty of patient's demand affects personnel workloads. Focused on the problem of assigning a HC patient to an operator under the constraint of continuity of care, Lanzarone et al. (2010a) and Lanzarone and Matta (2012a) proposed a cost-assignment policy to make sure the workload balance among operators and lowest cost increment. They assumed a cost quadratic in the number of extra visits (each operator supplies in surplus to his/her capacity in accordance with working contract), and modeled the stochastic workload with a stationary triangular distribution and patient demand with uniform distribution. Dealing with uncertainty, such as patient clinical and social conditions, Lanzarone et al. (2010b) presented a Stochastic Markov Chain Model to present the patient's care pathway to estimate patient requirements based on the historical data. The model predicted several major variables, such as the number of patients in following period, the duration of care, and the amount of required visits. The outcomes can be used for a robust assignment of patients to the operators with preserving the continuity of care and balanced workload among the operators. Coupling the patient stochastic model proposed by (Lanzarone et al. 2010b) to an assignment policy presented by (Lanzarone and Matta 2012a, b) proposed a methodology that provided useful instruments and yielded good results when applied to a representative HHC provider. This methodology allowed the development of a tool for automatically assigning patients to operators.

Mathematical models such as linear programming, mixed integer programming, and cardinality-constrained approach have been developed for patient assignment problem in the literature. Borsani et al. (2006) developed a mathematical support tool for HHC providers, which was composed of two models of linear programming solved by LINGO, first one regarded the assignment of each patient to a reference operator, the second one was a scheduling model that deals with the weekly plan for each operator. They considered the continuity of care, patient satisfaction, outsourced visits, operators' efficiency, workload balance, service quality and efficiency (coherence between operators' skills and patient needs, coherence between operator and patient geographic areas, and the impact of patient on the burn out level of the corresponding operator) in their models to minimize the number of visits outsource, the number of visits carried out by non-reference operator and the number of visits carried out during non-preferential days. Considering the problem of assigning patients to nurses for home care service according to a territorial approach, whose aim is to balance the workload (depends on three components: visit load, case load, and travel load) of the nurses while avoiding long travels to visit the patients, Hertz and Lahrichi (2009) proposed a mixed integer programming model with some nonlinear constraints and a nonlinear objective which were solved by a Tabu Search algorithm. They classified patients into five categories and assigned different weights for each category when built the model. The nurses were also distinguished into three types, especially, consisted of the "surplus" nurse team, which were asked to deliver specific nursing care treatments by the professionals responsible for the patients. Under the assumption that patients' demands are either deterministic (demand for care in each period has been determined before assignment) or stochastic, Lanzarone et al. (2012) proposed a set of mathematical programming models to balance the workloads (measured by maximizing the minimum utilization rate in each period) of the operators within specific category. The models run using OPL 5.1 consider several peculiarities of HC services, such as the continuity of care (hard, partial, or completely neglected), operators' characteristics (skills, geographical area he/she operates, capacity in each period), independent or integrated districts (coherence between operator and patient geographic areas). The numerical experiments demonstrated that the model produces assignments with perfect continuity of care and better workload balancing than those actually implemented by the provider. Yalcindag et al. (2012) assumed that the assignment is held within a single category of operators who are divided into groups (known before the assignment process) based on their main skills and geographical areas to serve. In each group, all operators are homogeneous with same professional capabilities. Newly admitted patients are assigned to one of the operators in the beginning of the following week under the consideration of balanced workloads of operators and satisfaction of a set of specific constraints, such as the continuity of care. The demand of newly admitted patient is expressed as the total amount of time requested for the visits including travel time. They analyzed three approaches for the assignment problem, including two structure policies (assign one patient at a time after having ranked newly admitted patients with some criteria) and one mixed integer programming model (assign all of the new patients of a week together). When modelling the problem of assigning a set of HHC patients to a set of nurses over a time horizon under the consideration of different continuity of care requirements (hard continuity of care, partial continuity of care, none continuity of care), Carello and Lanzarone (2014) apply the robust cardinality-constrained approach proposed by (Bertsimas and Sim 2004). The problem aims to minimize the overall overtime costs and the cost associated to reassignments. The approach modeled the problem as a deterministic assignment model to take into account uncertainty in patients' demands without the necessity of assuming probability distributions or deriving scenarios. They classified patients into different categories based on the level of continuity of care they require and classified operators to different districts based on territory and skills, and assumed that demand of each patient during each period is an uncertain parameter with expected value and maximum value. The proposed model, which was implemented with OPL 5.1 and solved by Cplex, had better performances in terms of overtime costs and fairness in nurses' workloads.

The patient assignment problem, which we address here based on the situation of HHC system AHNH in Hong Kong, has several similar characteristics with these described in the literature. First, it aims at maximizing the level of satisfaction of both operators and patients (from the perspective of operators, it focuses on balance the workloads among operator, fairness in wage and benefit, and satisfaction of their personal preferences; from that of patients, it pays more attention on level of quality of care service, level of continuity of care, the satisfaction of patients' preference, patients' waiting time, and emotional support from operators); second, it considers patients and operators classification, and deterministic or stochastic patient demand as well. Finally, it is solved at the beginning of a planning horizon composed of several periods. Even though there exist several similarities, the model is different from these in the literature. What is more, it has several different characteristics that were not considered in the literature. First of all, overtime is forbidden to make sure that operators can provide high quality of care because of enough rest and acceptable burn out level, whereas, some papers consider that overtime is allowable with additional benefit following. However, more and more care providers request enough time to relax and enjoy their life. Second, visits can be delayed due to the unavailability of patients or operators (e.g., readmission of patient, workload of operator exceeds his/her capacity on contract) or can be brought forward due to idle time of operators and availability of patients, since operator has more active actions on the visit schedule. Once more, each patient has different priority to be visited which depends on his/her health condition.

3 Problem Definition

The patient assignment and grouping problem in AHNH requires to determine which operators will provide care services to which patients during a certain planning horizon with a predetermined objectives such as balancing the workloads of the operators, minimizing the number of delayed visits, minimizing the travel time, etc. AHNH is located in New Territories, Hong Kong. This hospital provides services, such as traditional hospitalization, home healthcare services, mostly for the elderly and post care. The home healthcare services mainly are rehabilitation services provided by physical or occupational therapists and some assistants. Approximately 30–40 patients receive their care services from operators weekly. Some key characteristics of this patient assignment and grouping problem in AHNH are shown as follows:

1. Physical therapists provide cares such as exercise therapy, mobility and walking training, chest physiotherapy, pain relief; while occupational therapists are responsible for care needs including aids prescription/training, home assessment/modification, ADL training, cognitive training; rehabilitation assistants only can help the therapists to follow the cares (when therapists' workload exceed their capacity) but cannot be in charge of cases, which means only therapists can make decision whether the cases can be closed or not.
2. The newly admitted patients should all be assigned to operators according to their needs and should all be visited in the following week after their admission. It means when weekly workload of operator exceeds his/her capacity, the newly admitted patients have higher priority to be visited than these follow-up patients, as a result, some visits to patients with lower priority will be delayed even though it is time to be visited.
3. The newly admitted patients can be distinguished into two cases, urgent or nonurgent, urgent newly admitted patients have highest priority (level 1),

followed by nonurgent newly admitted patients (level 2), while the follow-up patients are divided into three categories based on the frequency of visits and health condition. If patients request care services one time per week, their priority to be visited is set to level 3, for those request services one time per 2 weeks, their priority is set to level 4, and those request service one time per four weeks, their priority is set to level 5. However, when visits are delayed, the priorities of these cases in each period will be updated after phone traces, if patients' health conditions are stable or better and do not need to be visited as soon as possible, the priority will be the same as before, otherwise, the priority will increase according to patients' health condition.

4. Each visit to patient has a deterministic and constant duration based on the care that patient needs, which does not include travel time. Most of patients have no time preference, while some patients request certain time window to be visited due to personal reasons, in consequence, patients with time preference will be considered first when grouping patients and scheduling.
5. Only one preference operator of specific category (physical therapists or occupational therapists) can be assigned to one patient based on patients need. Patients, who need both physical and occupational therapies, are assigned twice to different categories of operators, and visits for the same patient within the same day are not allowed.
6. Each operator has workload capacity (say 30 patients per week) according to their work contract, and overtime is forbidden, while idle time is not preferred. It means two cases will occur. When current workload of operator exceeds his/her capacity, some visits may be transferred to rehabilitation assistants temporarily, or visits maybe delayed and rescheduled in next few weeks. When current workload of operator is less than capacity, some visits may be brought forward so that utilization rate of operator can approximately equal to 1.
7. Assigned patients of each operator are clustered into groups, each of which should consist of more than two patients in accordance with their priorities and home location. Because each operator can visit 3–4 patients in both morning and afternoon section, and there exists a lunch break for each operator after morning section and before afternoon section, it maybe reduce traveling time for operators if patients visited are adjacent (say 5–10 min of walk).

As frequency of visits to patients is diversified from once per week to once per 4 weeks, the planning horizon consists of several weeks t (with $t = 4$). Assignments are made on a rolling time basis: each time the problem is solved, a new planning horizon is considered and the assignment problem is solved over this planning horizon. The flow charts of this assignment problem are presented in Figs. 1 and 2, which indicate that it is necessary to propose an update scheme of patients' priority, otherwise it will be difficult to model the problem. In the initial stage of AHNH (in week 1), only newly admitted patients are assigned to operators. Start from week 2, newly admitted patients are assigned to operators, while the follow-up patients are cared by the same operators as before. When considering assignment of patients to operator, we try to use greedy heuristic method to solve it based on patients'

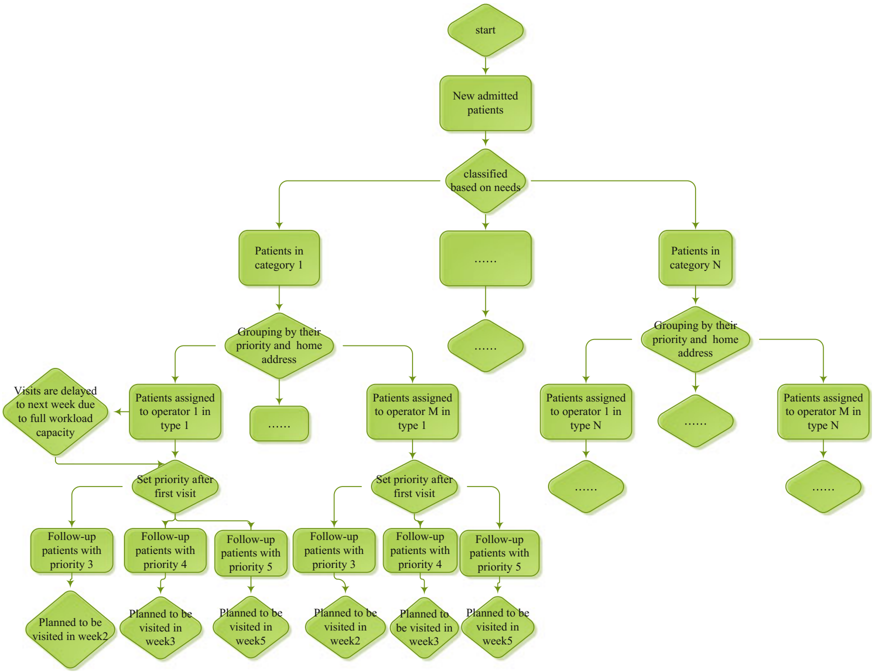


Fig. 1 Flow chart of initial week planning

locations and priority. We focus on assigning a set of patients to a certain type of operators, since the logic to deal with assignment of patients to other types of operators is the same. Table 1 shows the notations of our patient assignment and grouping problem.

4 Patient Assignment and Grouping Model

The assignment and grouping problem addressed here is different from that from the literature. The assignment considers the continuity of care case where patients are assigned to the principle operator. Once they enter the AHNH, they keep this principle operator during their length of stay. The planning horizon we consider here is 4 weeks, since most of patients request to be visited at once per week, once every two weeks, and once every four weeks. For each operator, the follow-up patients have been grouped by due week of visit, priority, and location. In theory, patients who need care once per week will appear in the planning horizon four times, patients who need care one every two weeks will appear in the patient queue twice, while those need care once every four weeks only once. However, unseen events may happen during the planning horizon, such as patients' unavailability,

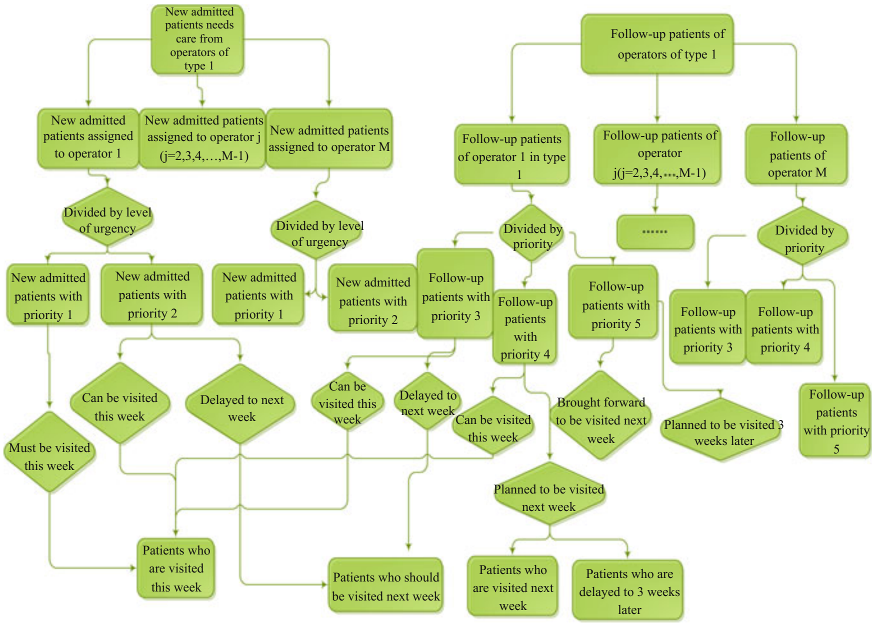


Fig. 2 Flow chart of week 2 (3, 4, …) planning for certain type of operator

Table 1 Notations

NP_t	The set of newly admitted patients in period t
$\mathrm{FP}_{i,t}$	The set of follow-up patients of operator i in period t
$\mathrm{NP}_{l,t}$	The number of newly admitted patients set to level l of priority in period t
$\mathrm{FP}_{i,l,t}$	The number of follow-up patients with level l of priority of operator i in period t
O_t	The set of operators in period t
c_i	The workload capacity of operator i
$w_{i,t}$	The current workload of operator i in period t
$l_{p,t}$	The priority of patient p in period t
$a_{p,t}$	The home address of patient p in period t
$\mathrm{fv}_{p,t}$	The frequency of visit required by patient p in period t
$d_{p,t}$	The flag of patient p who is delayed in period t
$b_{p,t}$	The flag of patient p who is brought forward in period t

which will lead to the update of due week of next visit. Figure 3 shows the follow-up patients of operator 1 in the planning horizon, and Fig. 4 shows the basic model of patient assignment and grouping problem in each week. The details about the assignment policy and priority update scheme will be indicated in following subsections.

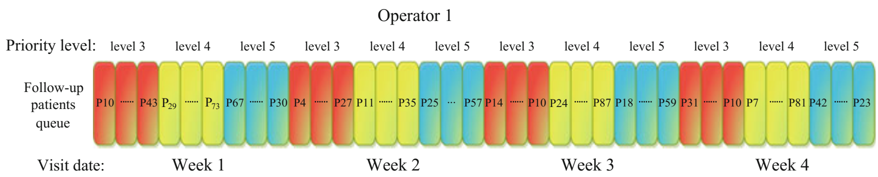


Fig. 3 Follow-up patients queue of operator

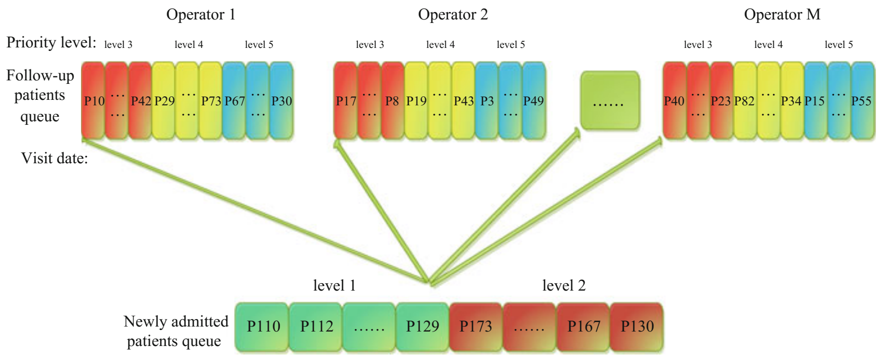


Fig. 4 Weekly assignment of newly admitted patients to operators

4.1 Priority Update Model

The level of priority of patients will be updated after each visit, for patients need and health conditions may change, or visits to patients are delayed or brought forward. In this subsection, we use Markov Chain Model to present the update of priority of patients, as shown in Fig. 5. The states in Markov chain are the level of priorities of patients, whereas the transition matrixes are based on the history data, changes depending on the patients' conditions can occur each week when a modification of the care is necessary. Starting from the initial state x_0, the patient state vector is assumed to evolve in discrete time from week k to week $k + 1$, according to the transition probability matrix $P = [p_{ij}]$. According to the frequentistic approach, the transition probability matrix is obtained from historical data for the update scheme as follows:

$$p_{ij} = \begin{cases} \frac{w_i - \sum_{j \neq i} n_{ij}}{w_i} & i = j, i \neq 0, 6 \\ \frac{n_{ij}}{w_i} & i \neq j, i \neq 0, 6 \end{cases} \tag{1}$$

where p_{ij} is the transition probability from state x_i to state x_j, n_{ij} is the total number of transitions from state x_i to state x_j and w_i is the total number of weeks spent by all the patients in the state x_i. In contrast, the state End (x_6) is defined as an absorbent

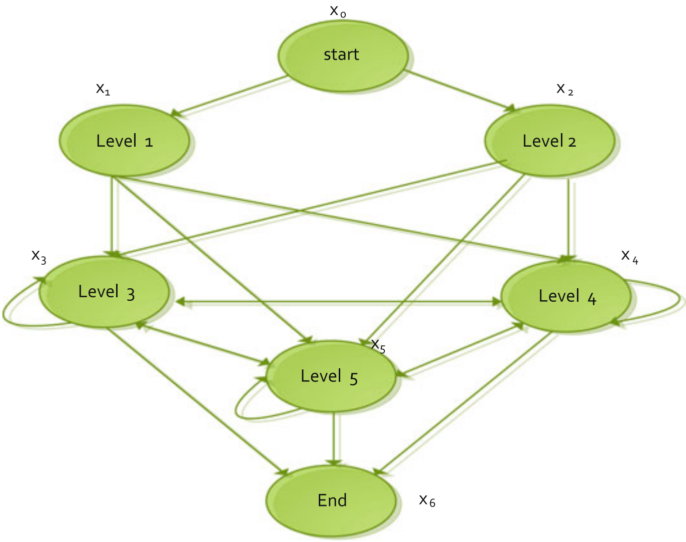


Fig. 5 Markov status graph

state, while the transition probability from state Start (x_0) to state x_1 and x_2 depend on the percentage of urgent newly admitted patients over all new-admitted patients based on history data. The proposed Markov chain is homogeneous, such as the transition probabilities are assumed to be invariant with the time, and sojourn times in each state is assumed to be geometric distribution.

4.2 Assignment Policy

In the assignment problem, it is important to maximize the operators and patients satisfaction, such as to achieve the fairness among operators (the workload balance among operators in same category), to minimize the patients waiting time (measured by the number of delayed visits in each period). Hence, the assignment policies focus on achieving these two objectives.

4.2.1 Workload Measurement

Workload balance is known as being a good measure of the performance of healthcare systems, as indicated in the literatures, almost all researchers working on operation management in HHC have considered it as an significant objective. However, workload measures used in these papers have no standard. Hughes (1999) compared activity-based measures (where the workload of each nurse mainly depends on the time spent with the patients) with dependency-based measures (where the patients are grouped into categories, and a workload is associated with each category), he made conclusion that these two types of measures are neither valid nor reliable. For this reason, Hertz and Lahrichi (2009) defined a workload measure which depends on three components: the visit load measured by the weighted sum of the visits that operator have to perform, the weight of a visit being defined according to its heaviness when compared to a witness visit; the case load which depends on the number of patients assigned to the operator in each category; the travel load which depends on the distance that the operator has to travel outside, and the number of visits required by these patients.

In this paper, we consider the similar workload measurement. Since each visit has a deterministic and constant duration (say 0.75 h), the visit load is measured as the number of visits during planning horizon; the case load depends on the number of patients assigned to the operator in each level of priority; and the travel load depends on the distance that the operator has to travel outside from ANHN to visit patients, which depends on the grouping results in terms of patients' home address. According to patients' home address, we cluster patients into different groups before and after assignment to make the travel load as low as possible. As patients with higher level of priority need more care, it will increase the imbalance if there exists significant difference of number of patients with same level of priority among different operators. As a result, to make sure workload balance, it is more important to keep the balance of case load and followed by travel load. Because the level of priority of patients has direct impact on the visit load as explained before, it is not as such important to consider visit load as a significant part.

4.2.2 Maximum Balance Policy

To keep workload balanced and reduce the number of delayed and brought forward visits due to difference between current workload and capacity, the maximum balance policy is applied in this assignment problem. The number of newly admitted patients assigned to each operator will be calculated as follows:

$$n_{\text{avg}} = \left\lceil \frac{\sum_{i=1}^{|O_t|} \text{FP}_{i,t} + \text{NP}_t}{|O_t|} \right\rceil \tag{2}$$

$$n_{\text{act}} = \left\lceil \frac{\sum_{i=1}^{|O_t|} \text{FP}_{i,t} + \text{NP}_t - \sum_{i \in \{I_t | i \in I_t, n_{\text{avg}} \le \text{FP}_{i,t}\}} \text{FP}_{i,t}}{|O_t| - |I_t|} - \text{FP}_{i,t} \right\rceil \tag{3}$$

$$n_{i,t} = \begin{cases} 0 & n_{\text{avg}} \le \text{FP}_{i,t} \\ n_{\text{act}} & \text{otherwise} \end{cases} \tag{4}$$

where $|O_t|$ is the total number of operators in week t, while $FP_{i,t}$ is the follow-up patients of operator i in week t, and NP_t is the total number of newly admitted patients that need to be assigned in week t. n_{avg} is an integer value presenting the expected number of patients of each operator, while n_{act} is the number of newly admitted patients that should be assigned to each operator whose current workload does not exceed the expected value. It can help to keep balance among operator globally. However, patients with different levels of priority need different levels of care, and the case workload of each operator will be very different if we do not consider the distribution balance of patients with different levels of priority. As a result, it will increase the imbalance of workload among operator even though they provide care to same number of patients. So it is very important to keep the case workload balance.

Newly admitted patients are classified into be urgent and nonurgent before their first visit. Even though their levels of priority as follow-up patients are unknown when being assigned, the probability of urgent cases can be calculated by frequentistic approach according to historical data, and defined as p_{01}, and the probability of nonurgent case is defined as p_{02}. Using Markov Chain Model shown in 4.1, it is enable to calculate the percentage of newly admitted patients to be set to different levels of priority (for level 3, 4, and 5). The number of new-admitted patients set to level 3 (4, 5) in period t is defined as $NP_{3,t}(NP_{4,t}, NP_{5,t})$, which is calculated as follows:

$$NP_{l,t} = p_{01} * p_{l1} * NP_t + p_{02} * p_{2l} * NP_t, \quad l = 3, 4, 5 \tag{5}$$

So the number of urgent and nonurgent patients that should be assigned to each operator can also be calculated by substituting $FP_{i,t}$ and NP_t in formulas (2)–(4) with $FP_{i,l,t}$ and $NP_{l,t}$.

Finally, it is important to deal with the travel workload balance as well. After determining the number of newly admitted patients assigned to each operator, it is necessary to consider patients' home location when certain newly admitted patient is assigned to the operator. To reduce travel load for each operator, we cluster the follow-up patients into different groups where there are at least two patients who live near to each other (e.g. it only take 5–10 min to walk). If newly admitted patient can be assigned to two operators, then he/she will be assigned to the operator if the increase of travel load is smaller. This part will focus on the grouping problem in terms of patient's home location, the details of which are not presented here.

4.3 Delay and Bring Forward Policy

Since each operator has a weekly capacity and the utilization rate should be as much approximate to 1 as possible, two cases will occur. One is that, when workload of operator in week t after aborting newly admitted patients is less than the capacity, some visits due in week $t + 1$ will be brought forward to fill the idle time. Visits are brought forward to be visited in week t, if the patients have higher priorities and live near to these patients who are planned to be visited. The level of priority of patients whose visits are brought forward will be assumed to steady. The other one is that when workload of operator in week t exceeds the capacity, some visits will be delayed to make sure the quality of care. Visits are delayed to be visited in week $t + 1$, $t + 2$, or $t + 3$ (depends on patients health condition and request), if the patients have lower priorities and live far away from these patients who are planned to visited. The level of priority of patients whose visits are delayed will be updated according to the patient request.

5 Conclusion and Future Work

The models and methods reported in the literature, show that the patient assignment problem mainly focuses on assigning newly admitted patients to operators considering continuity of care and uncertainty of patient demands and allowing overtime or outsourcing visits with certain penalty, while aiming at balance the workload among operators. These papers provide several integer programming models, most of which are solved by provided software (Borsani et al. 2006; Carello and Lanzarone 2014; Lanzarone et al. 2012), only one paper uses heuristic method (Hertz and Lahrichi 2009), while the others use simulation (Lanzarone and Matta 2012a; Lanzarone et al. 2010a, b). Most of them focus on the patient assignment problem only except the two focuses on both the patient assignment and scheduling problem (Borsani et al. 2006; Yalcindag et al. 2012). However, in practice, overtime for operators in HHC system is not allowed and visits can be delayed. With these two specific characteristics, we describe the patient assignment problem different from these in the literature and model this specific patient assignment problem. Solving it with heuristic methods, such as particle swarm optimization, genetic algorithm and tabu search, will be the major future work. This paper makes contribution to present and model a specific patient assignment problem after the report in the literatures.

References

Bertsimas, D., & Sim, M. (2004). The price of robustness. *Operations research, 52*, 35–53.

Borsani, V., Matta, A., Beschi, G., Sommaruga, F., & Ieee. (2006). *A home care scheduling model for human resources*. New York: IEEE.

Carello, G., & Lanzarone, E. (2014). A cardinality–constrained robust model for the assignment problem in Home Care services. *European Journal of Operational Research*.

Hertz, A., & Lahrichi, N. (2009). A patient assignment algorithm for home care services. *Journal of the Operational Research Society, 60*, 481–495.

Hughes, M. (1999). Nursing workload: an unquantifiable entity. *Journal of Nursing Management, 7*, 317–322.

Lanzarone, E., & Matta, A. (2012a). A cost assignment policy for home care patients. *Flexible Services and Manufacturing Journal, 24*, 465–495.

Lanzarone, E., & Matta, A. (2012b). The nurse-to-patient assignment problem in Home Care services. In *Advanced Decision Making Methods Applied to Health Care* (pp. 121–139): Springer.

Lanzarone, E., Matta, A., & Jafari, M. A. (2010a). A simple policy for the nurse-patient assignment in home care services. In (pp. 1–6): IEEE.

Lanzarone, E., Matta, A., & Sahin, E. (2012). Operations Management Applied to Home Care Services: The Problem of Assigning Human Resources to Patients. *Ieee Transactions on Systems Man and Cybernetics Part a-Systems and Humans, 42*, 1346–1363.

Lanzarone, E., Matta, A., & Scaccabarozzi, G. (2010b). A patient stochastic model to support human resource planning in home care. *Production Planning and Control, 21*, 3–25.

Yalcindag, S., Matta, A., & Sahin, E. (2012). Operator assignment and routing problems in home health care services. In (pp. 329–334): IEEE.

Part VI
Quality in Service Industry

Kitson Lee, Council Member of Singapore Quality Institute

Designing and delivering that unique and memorable customer experience is imperative in today's business. Companies seek to surpass each other through achieving excellence: in ensuing satisfaction of the customer, pursuing loyalty, and ultimately a commendable business performance.

These papers presented at the Asian Network for Quality Congress 2014 in Singapore provide us insights into the works of these authors who had demonstrated the ever-increasing requisite for customer centricity, service innovation, and service quality. One of the authors highlighted how his company's unique Customer Centricity Framework help maintain a continuous win–win relationship with their customers, thus driving up business value. Another revealed that innovation is not an option, rather essential for survival in today's turbulent business environment. One interesting study on the robustness of the conclusion of customer satisfaction surveys is also presented in this chapter.

I trust you find these shortlisted papers beneficial and thought-provoking.

About the Author

Kitson is a much sought after Presenter and Facilitator on the topics of Service Quality, Innovation, and Customer Experience. With his years of corporate and consulting experiences—spanning across both Public and Private sectors—he understands the operational and business challenges faced by employees and their organisations. Kitson also holds various appointments which include Executive Council Member & Honorary Secretary, Singapore Quality Institute (SQI); Asian Network for Quality (ANQ); GEMS Service Capability Workgroup; and Associate, Public Service Centre of Organisational Excellence. Kitson was also featured in The Straits Times and on News Radio 93.8FM for his strong entrepreneurship spirit and innovative ways of bringing organisational successes to greater heights.

Customer Centricity: Drive Sustainable Operations

Devender Malhotra

Abstract Competition and dynamic business environments have made Quality an essential part of organizational success. Earlier, Quality had more to do with process management, compliance, continuous improvement, and certifications. However, current trends show that Quality drives agility in execution. With a focus on improving process capability and people capability, Quality drive can improve productivity to enhance customer experience. From considering customers only as a source of revenue, organizations today recognize them as the raison d'être of their business. They also realize that they need to build the business around the customer. Sustainable competitive Quality extends beyond traditional boundaries, with it being increasingly linked with customer loyalty. Research shows that the customer plays a vital role in enhancing productivity and competitive Quality. Customers not only receive what an organization produces and delivers, but also directly and indirectly influence the operations and outcomes of an enterprise. It is for this that customer centricity is a bedrock of operational excellence. Wipro's **unique Customer Centricity Framework** aligns service delivery with the customer's expectations. Effective customer participation can increase the likelihood of his/her needs getting met while bringing in more customer benefits. This helps maintain a continuous, strong win–win relationship with the customer. As part of Customer Centricity Framework, we listen to the voice of customers, drive business value, and provide transparency in service operations.

Keywords Customer centricity · Agility in execution · Business value · Voice of customer · Transparency

---

D. Malhotra (✉)
Wipro IT Services, EC-2, T6, 2nd Floor, C-Wing, Keonics Electronics City, Bangalore 560100, India
e-mail: devender.malhotra@wipro.com

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_15

1 Introduction

Customer centricity is a much talked about topic in business circuits today. With recent advances in social media, data analytics and connectivity solutions, an organization is able to better understand and segment its target demographics. This allows it to differentiate itself from competitors.

Today, all IT organizations are grappling with the triad of quality, speed, and cost amid global economic uncertainty. Organizations are putting in a lot of effort to improve operational efficiency. They are trying to identify areas that can protect their profit margins or simply survive in the global marketplace. However, most of these initiatives get buried in organizational complexity and fail to align with customer expectations. Regardless of the specific processes, tools, and methods an organization adopts and the Quality experts follows, it must become customer centric to gain a competitive edge. Many Quality management programs and efforts to enhance competitiveness take a rather limited view of potential customer involvement in the process.

Until a few years ago, Quality for many organizations was all about certification and had little to do with customer experience. As the market becomes more and more competitive, firms can no longer afford to be satisfied with just meeting customer expectations. Businesses are now beginning to realize that a happy customer not only enhances profitability, but also serves as a reference for prospective clients and they are willing to build long-term relationships with their providers and to get more services. They have started to understand the importance of loyalty quotient. This is particularly important for B2B buyers since they consider peers as trusted sources of information. Such customer advocacy can be a positive differentiator and goes a long way in building trusted partnership.

In this paper, we highlight Wipro's strategy and framework and best practices for customer centricity. We begin by highlighting the importance of nurturing a customer-centric organization. We also talk about our Four-Pronged Framework through which businesses can enhance customer experience and, thereafter, we focus on some key metrics for measuring success.

2 Strategic Alignment with Customer Needs

Creation of consistent customer experiences is a complex task. A recent Forbes article suggests that achieving customer centricity might require a deep cultural change. There needs to be a centralized effort to analyze data and search for patterns that will help gain a better understanding of customers. Yet another study talks about the evolution of the customers' demands and expectations. Customers today are more tech-savvy and have diverse needs. A seamless customer experience at every point of engagement is essential. This can often be a challenge as siloed efforts by individual divisions alienate rather than delight the customer. Formulation of a centralized strategy will enable alignment among various departments.

An organization's customer centricity strategy can be based on a few basic points:

- **Understanding the Customer**: Given that customers today need individual attention and can no longer be grouped into vague segments, it is essential to understand their varying needs. Companies should use the best possible tools and techniques to create quantifiable and actionable insights into customer behavior. Such tools include cutting-edge data analytics, customer tracking, etc.
- **Decision Making Process**: Strategic decision making should be centered on customer experience management. One must analyze the impact of all decisions on customer perception of one's products and services and gauge customer satisfaction in accordance with that.
- **Organizational Culture**: Customer-focused practices must be institutionalized across the organization. Customer-facing employees should be sensitized to the fact that all customer engagement channels need to be seamlessly integrated. One must secure employee buy-in for customer-centricity initiatives and train them as and when required.

Delivering Insights: The organization should be able to draw insights from the available customer information and also predict future service behavior. It can then deliver value-added business solutions that are aligned to customer needs, and hence generate business value.

An organization that realizes the importance of customer centricity can now look at specific operational improvements to achieve such centricity. By doing so, it will be able to get closer to its customer.

3 Customer Experience: Wipro's Four-Pronged Framework

As we have already highlighted, an exceptional customer experience forms the foundation for a strong business-client relationship. To enable this, Wipro has classified services into four different levels:

- **Generic**: Carrying out assurances for the bare minimum requirements
- **Expected**: Meeting customer expectations with consistent performance/behavior
- **Differentiated**: Providing unique services that surpass customer expectations
- **Transformational**: Enhancing services enabling customers to reap unexpected benefits.

At each successive level, the customer receives more value and this helps establish a win–win relationship as illustrated in Fig. 1.

Wipro has built a robust framework that views service delivery from customers' perspective. Such an approach is intended to gain trust and, hence, build loyalty and

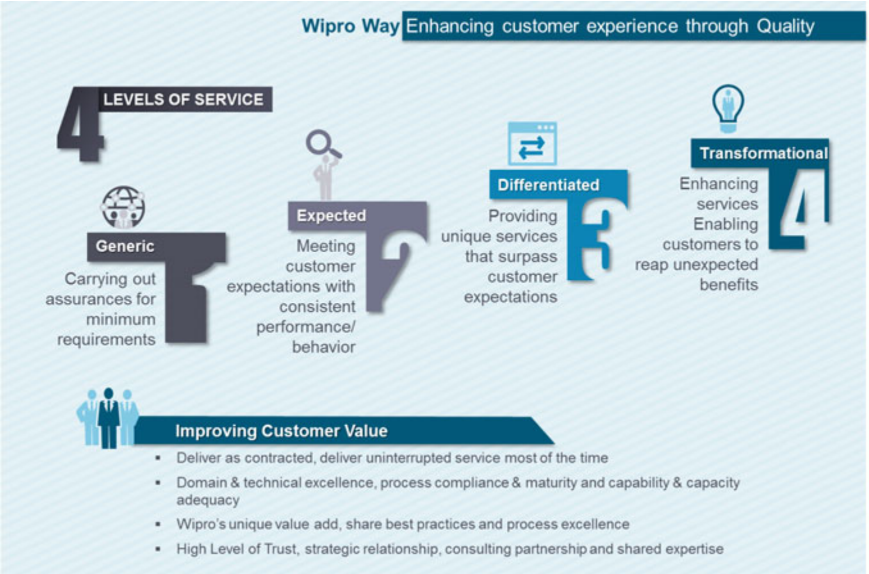


Fig. 1 Service levels and the incremental value delivered

customer advocacy. It ensures technology is implemented in a manner that maximizes the value for customers. It also establishes a monitoring and feedback system to measure the impact of each initiative and make appropriate changes as customer needs evolve. This Four-Pronged customer centric framework includes:

1. **Voice of Customers**: Maintaining an active dialogue with customers (and acting on feedback) and fostering a culture that places the customer at the heart of the decision-making process
2. **Driving business value for customers** and aligning delivery with customers' business outcomes
3. **Proactive behavior** in managing alerts/warnings and addressing potential issues
4. **Transparency in service operations** through communication and collaboration with customers.

3.1 Voice of Customers

Listening to the customer's voice at project, program, and relationship level plays a vital role in boosting customer experience. This can be done through direct feedback, informal meetings, governance meetings, and senior management interaction. Wipro's Quality organization connects with the Customers through various mechanisms.

3.1.1 Surveys

Customer satisfaction (CSAT) surveys must be conducted regularly, keeping in mind the unique characteristics of each relationship. At Wipro we conduct surveys and take feedback from customer at various milestones of projects, Quarterly relationship surveys to take pulse of senior and middle managers of customer organization. Wipro's Annual satisfaction survey ensures feedback from all customers of Wipro and with a focus on CXOs.

3.1.2 Independent Connect

Wipro Quality team has created an independent channel for customers to provide candid feedback through independent Client Connects. This plays an important role in minimizing adverse impact and restoring customer confidence.

3.2 Driving Business Value for Customers

Customer behavior and expectations are driving fundamental shifts in business models. Business complexity continues to grow and the rate at which change has to be delivered is increasing. It is very important for organizations to continuously assess and optimize the true business value that they deliver. Alignment of delivery with business outcomes contributes significantly to customer experience.

Improvements in the three focus areas for operational excellence—quality, speed, and cost—often do not proceed in tandem. For instance, any increase in speed of delivery might adversely affect the Quality and, hence, the customer experience. Better Quality might also mean higher costs. Wipro has developed a business value meter framework that overcomes such conflicts. It is designed to go beyond standard SLA measurement and capture the value delivered to customers along four dimensions:

- Growth in revenue
- Cost optimization, productivity gains, and labor arbitrage
- Improvement in end user satisfaction
- New reputational and regulatory gains.

The precise impact of acting along these dimensions is described in Fig. 2.

If an organization is able to keep track of the value delivered through this framework, it can:

- Strengthen the culture of delivering business benefits to customers
- Collaborate and partner with customers to realize business benefits
- Have a single source of information for all value delivered to customers across business segments
- Share best practices and learning.

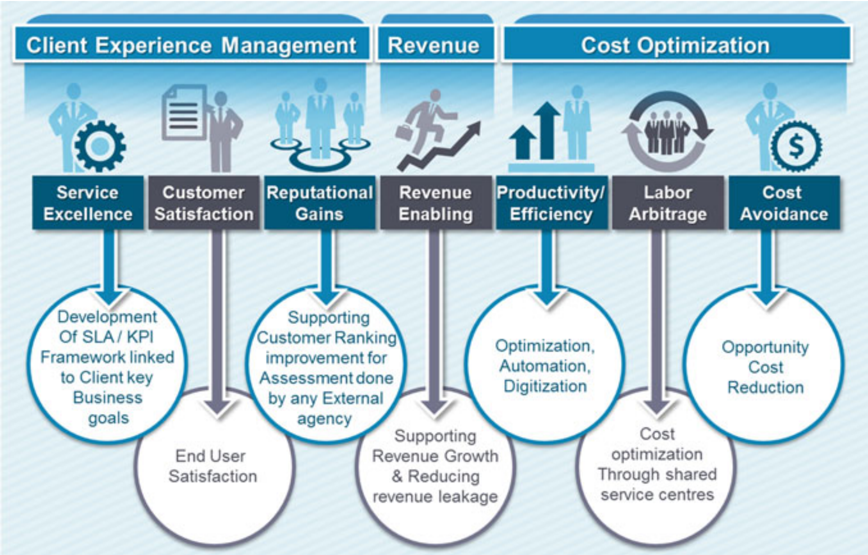


Fig. 2 Business value meter framework

We have already put the business value meter to use in IT Service Delivery, providing our customers with clear visibility of the benefits. In one instance, the customer saw end user satisfaction rise from 62 to 90%; in another case, the customer saved USD6.82 million in cost through cloud service delivery initiatives.

3.3 Proactive Behavior

Businesses must try to stay ahead of customer concerns and take quick actions on their reviews and feedback. Such proactive approach is the basis for any differentiated offering in today's competitive environment. Given that customers today (in both B2C and B2B spheres) are more technology savvy than ever before, it can be a boon for agile firms. Customer feedback can now reach the relevant departments faster than ever before. Through continuous improvements based on this feedback, organizations can address potential weaknesses before any incident is caused.

Wipro cultivates proactive behavior throughout organizations via various systems and processes. Such behavior is driven from the top management levels and pervades throughout the organization.

3.3.1 Early Warning

This process ensures proactive management of customer concerns by raising alerts in advance for potential problems that are likely to occur. This helps the project team to mitigate them through timely intervention and address the root cause. Wipro's early warning system is designed to alert the senior management at different levels depending on criticality and priority of the issue. As part of governance, reviews are enabled at various levels starting from delivery manager to CEO level to resolve issues proactively to ensure customer satisfaction.

Thus, the culture of being proactive is driven directly by C-level executives in the organization.

3.3.2 Candid Voice

Through this system, Wipro has enabled all employees to be a part of the proactive problem management culture. They can alert the senior management on customer issues that could impact customer satisfaction. Concerned stakeholders can then intervene in a timely manner to resolve the issues. While Wipro's senior management and account leadership teams continuously interact with customers at various levels to understand their expectations, Candid Voice allows each employee to promote our culture of customer centricity, which essentially makes every employee an advocate for our customers.

3.4 Transparency in Service Operations

Transparency is the counterpart of the first aspect of this Four-Pronged Framework. It is as important for a firm to communicate the details of its offerings as it is to listen to customers' voices. Wipro believes in providing transparency to its customers at all levels. Its "Crystal—Customer portal" acts as a communication and collaboration platform for enhancing customer experience while also providing visibility on service delivery performance. The portal offers information on performance metrics, status, compliance reports and statements of work (SoW). It provides a platform for customer engagement through discussion on subjects such as key risks and issues. Even as it offers a differentiated experience by providing engagements and performance views, it delights customers by posting articles on industry research, thought leadership, center of excellence (CoE) initiatives and business value meter.

4 Metrics for Success

Wipro has defined a more insightful framework for measuring the success of customer centricity programs and it is illustrated in Fig. 3. Wipro's unique action planning framework enable delivery teams to plan activities proactively to deliver value to customer and enhance their experience. Key attribute of this framework is customer involvement to ensure customer strategies and business priorities are considered. It should be noted, however, that proactive steps need continuous attention from all stakeholders and are generally harder to delineate than reactive ones. This can be made easier if the organization is able to directly involve the customer in identifying the best strategy to support the business priorities. Such collaboration forms the core of any customer-centric initiative and helps the organization stay ahead of competition.

These four categories of actions are derived through "workouts" at the account level with delivery teams. They are aligned with the four service levels mentioned in the previous section and give the teams a visibility of the direction taken by every initiative. Each category of action serves a different purpose

1. **Assured Hygiene Actions** ensure that standards are maintained and future dissatisfaction is avoided. This is the very first tier of the framework that is intended to ensure that a basic level of service is maintained.

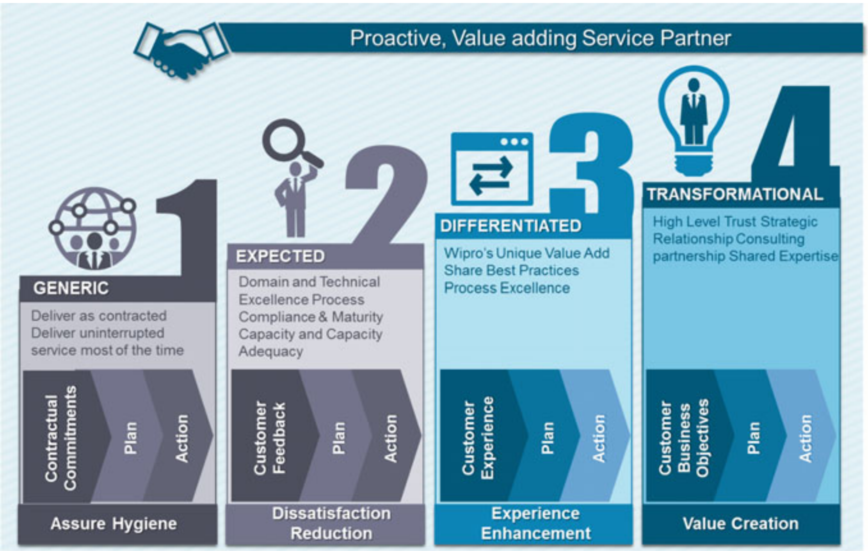


Fig. 3 Framework for customer experience measurement

2. **Dissatisfaction Reduction Actions** address stated areas of dissatisfaction. Within the confines of the contract, any shortcoming or oversight can be remedied by these actions.
3. **Experience Enhancement Actions** are related to engagements and relationships. Once we are able to ensure that contractual requirements are met at a bare minimum level, such best practices and value-add initiatives can help the organization differentiate its service offering.
4. **Value Creation Actions** generate value for customers over and above contractual levels. Finally, as the relationship evolves into a strategic partnership, these actions can prove to be transformational for the customer.

It is imperative to measure the success of an organization's customer experience initiative. It is also essential to initiate remedial action where experience levels are found to be lower than desirable. Three metrics that can measure the success of customer centricity initiatives are:

1. **Customer Loyalty and Advocacy**
   Customer loyalty can be of immense help in today's competitive industry landscape. Satisfied customers can greatly enhance marketing efforts by becoming brand ambassadors for the organization. A satisfied customer is likely to recommend others to do business with it and will also be willing to serve as references to prospective clients. Considering that B2B buyers consider peers as trusted sources, such advocacy is a positive differentiator and adds credibility to the brand.
2. **Share of Wallet**
   A greater share of wallet indicates better awareness of the customer. Again, it is more cost effective to cross-sell or up-sell to existing customers than to acquire new clients. A B2B client that already subscribes to, say, a cloud storage service will be more willing to try a new cloud solution based on past experience with the relationship. Share of wallet can be measured by dividing the revenue from a customer by their total spending in that particular category.
3. **Customer Value**
   Today, with better data gathering and analytical tools, organizations can calculate Customer Value to a high degree of precision. Inputs for this metric are the expected duration of a relationship and per-period profits from the customer.

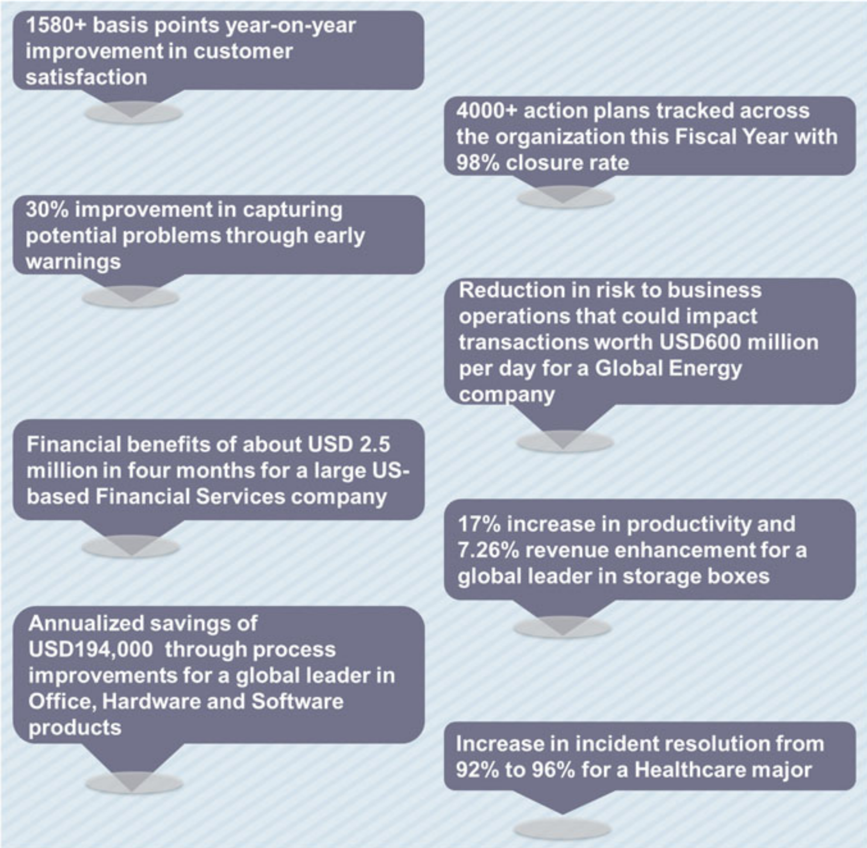
5 Conclusion

The business landscape today has become more competitive than ever before. Technology has helped knock down entry barriers and even established players cannot afford to be complacent about recent successes. Any organization—in almost any industry—needs to keep its customer at the center of its business strategy. Any initiative that is intended to improve upon service quality, speed, or cost can be sustained only if it receives customer buy-in. A comprehensive customer-centric approach would have three main facets:

1. **Customer Intuition**: With the help of cutting-edge technology, the organization must strive to gain a deeper understanding of customer behavior. Note that it is not a trivial task to get an idea of customer requirements just by asking them, as Steve Jobs famously said, "A lot of times, people don't know what they want until you show it to them." Moreover, one must ensure that customers do not feel that their privacy is being impinged upon by tracking programs. Any data-driven customer awareness initiative should be sensitive to potential privacy concerns.
2. **Utilizing the Insights**: Once the organization has a grasp on the behavior and tendencies of its target customers, it needs to modify its existing offering (as well as build new ones) based on them. This needs to be a continuous process since the competitive edge can rapidly be dulled as other firms catch up through their own customer centric initiatives.
3. **Monitoring and Feedback**: Finally, the impact of operational changes should be tracked with quantifiable metrics. Traditional monitoring mechanisms can be enhanced using predictive analytics that not only helps customers but also drives business value for the organization. All decision making processes must have a strong statistical component in an organization that strives to become customer centric. Whenever an initiative gets implemented, it should have an analytics-based backend for monitoring its performance. This helps set operational and delivery parameters like schedule, cost, and quality based on customer satisfaction. Moreover, risks can be accurately quantified and mitigated through the use of predictive models. As the types and volumes of data increase, text, and speech analytics can provide major decision support by providing unprecedented levels of insight into customer concerns.

Indeed, customer centricity can enhance both, a firm's topline as well as its profitability. Customer experience management can help achieve a stronger competitive differentiation and greater customer retention. Improvements in customer touch points, customer service, first contact resolution of customer concerns, customer intelligence, and improvements to existing products are the most common methods used for attaining the established customer experience goals. To this end, Wipro's collective knowledge assets propagate best practices for enhancing customer experience.

Outcomes/Results



References

http://monetate.com/infographic/what-does-it-mean-to-be-customer-centric/
http://www.forbes.com/sites/baininsights/2014/02/20/creating-a-customer-centric-company-six-paths-to-learning/
http://www.ey.com/Publication/vwLUAssets/The_journey_toward_greater_customer_centricity_-_US/$FILE/Customer_Centricity_Paper_29_April_Final_US.pdf
http://www.marketingprofs.com/9/customer-centricity-best-practices-customer-value-metrics-patterson.asp
http://en.wikiquote.org/wiki/Steve_Jobs

Author Biography

Devender Malhotra has over 22 years of experience in IT, BPO, engineering and automobile industries. He is a postgraduate diploma in Marketing and Financial Management with graduation in Mechanical Engineering. He is a Certified Master Black Belt with Wipro in DMAIC and Lead Black Belt in DSSS+ and TQSS Methodologies, driving Six Sigma as well as Lean Initiative as a certified Lean facilitator. He recently presented papers in ANQ and EOQ conferences.

He is currently responsible for driving Quality initiatives and is also instrumental in bringing Knowledge Management and Lean culture across Wipro.

The Relationships Between Innovation Orientation, Customer Orientation, Service Quality, and Business Performance

Gye-Soo Kim

Abstract In developing and implementing appropriate innovation related strategies, it is important understand how manager make plans to adequate competence. The aim of this study is to develop structural equation modeling techniques (SEM) of relationship that is cause and effect between innovation orientation, customer orientation, service quality, and business performance. A survey was developed to model the factors from innovation orientation activities to business performance (n = 125). Results show that innovation orientation and customer orientation factor were positively related service quality. Also, service quality was significantly impact on business performance. From the results, policy implications for innovation management are discussed and important directions for future research are suggested.

Keywords Innovation orientation · Customer orientation · Service quality · Structural equation modeling (SEM) · Business performance

1 Introduction

Business competition environment is very turbulent. Product life cycle is shorter than past because of customer need and want. The speeds of consumer trends are changing. The distance between today's generation of a product-and the next-is getting shorter. This puts innovation activities on being able to bring new goods to market quickly and stay one step ahead of the competition.

It's become a trendy managerial acronym: VUCA, short for volatility, uncertainty, complexity, and ambiguity. In this time, Innovation helps the company to deal with the turbulence of external environment and therefore, is one the key drivers of long-term success in business, and particularly in dynamic markets

G.-S. Kim (✉)
Department of Business Administration, Semyung University,
#65 Semyungro, 390-711 Jechon, Chungbuk, Korea
e-mail: gskim@semyung.ac.kr

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_16

(Baker and Sinkula 1999). Innovation orientation, customer orientation service quality, and business performance relate positively to each other. However, research that studies the interrelationships between the three concepts simultaneously is still scarce.

This research starts with a review of the literature on these topics and a description of the model proposals. Then, the article presents the design of the study to test the model and the findings of this study. In the last section, the article discusses the managerial and academic implications of this study, its implications and recommendations for future research.

2 Theoretical Background

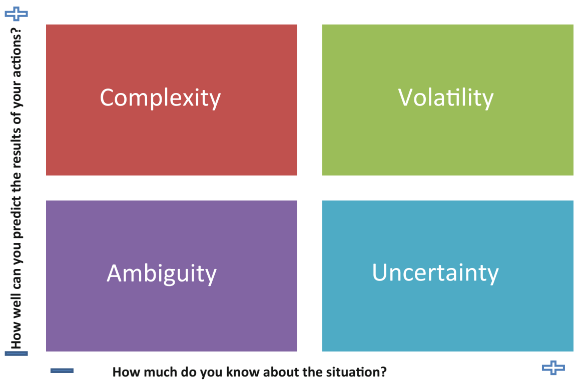
Today, companies are competing in a very different environment that they were only a few years ago.

Figure 1 VUCA World showed turbulent environment. Rapid changes such as global competition, e-business, the internet, and advances in technology have required business to adapt their standard practice. Innovation orientation and customer orientation is the critical function through which companies can succeed in this competitive landscape.

Innovativeness, the propensity to create and/or adopt new products, process, or business systems, positions a company well for developing new value for its customers. Company of Innovation orientation can better meet a customer need, thereby providing greater value. Innovation is a “key mechanism for organizational growth and renewal” (Lawson and Samson 2001). Innovation, at an aggregate level, represents the successful exploitation of ideas that are new to an adopting organization, into profitable products, processes, and/or services (Damanpour 1992).

Service Innovation can be argued that the relatively easy access to market information about trends, especially consumer habits and buying behaviors, levels the information field with respect to customer orientation (Johnson et al. 2009).

Fig. 1 VUCA World



A company is market-oriented, according to the technical definition, if it has mastered the art of listening to customers, understanding their needs, and developing products and services that meet those needs. Believing that this process yields competitive advantage, companies spend billions of dollars on focus groups, surveys, and social media. The voice of customer reigns supreme, driving decisions related to products, prices, packaging, store placement, promotions, and positioning (Dawar 2013). Market orientation, therefore, focuses the organization's ability to be responsive to customers and other relevant stakeholders in order to be profitable (Naidoo 2010).

Service orientation is perhaps the most fundamental element of market orientation to business performance, while others (Cheng and Krumwiede 2012).

In summary, Innovation orientation and service orientation help to develop and sustain a service quality, business performance. For a sustainable growth, companies should keep innovation, improve service quality. In result, companies can be survival.

3 Research Model and Research Hypothesis

Most of the broad empirical studies on the relation between innovation and performance provide evidence that this relation is positive (Thornhill 2006). Innovation orientation and service orientation can thus serve as the catalyst for innovation, since it opens up the firm to new customer needs and new business processes. In other words, Innovation orientation and service orientation can be a critical part of the initiation stage of innovation (Hurley and Hult 1998). Therefore, building on extant literature, it is hypothesized that.

H_1: Innovation Orientation relates positively to service quality.
H_2: Service Orientation relates positively to service quality.
H_3: Service Quality relates positively to Business Performance.

The three key constructs of this study are innovation orientation, customer orientation, service quality, and business performance. The survey items used in this study, were based on a five-point Likert scale format and sourced from extant literature, wherever possible(1-strongly disagree, 5-strongly agree). Table 1 provides an overview of measurement items.

The surveys were distributed between November 2013 and January 2014. SMEs (Small and Medium Enterprises) in South Korea were randomly contacted by email and visiting. 125 questionnaires were usable in this analysis. Summary statistics for the respondents are shown in Table 2.

Table 1 Measurement items used for study scales

Factors	Items	Authors
Innovation orientation	Developing new products (X1)	Johnson et al. (2009), Thornhill (2006)
	Upgrading existing product's appearance and performance (X2)	
	Innovation in production processes (X3)	
	Innovation in marketing techniques (X4)	
Customer orientation	We give close attention to after-sales service (X5)	Damanpour (1992), Hurley and Hult (1998)
	We constantly monitor our level of commitment and orientation to serving customer's need (X6)	
	Our business objects are driven primarily by customer satisfaction (X7)	
Service quality	Excellence for quality (X8)	Cheng and Krumwiede (2012), Johnson et al. (2009)
	Optimal price (X9)	
	Flexibility to delivery (X10)	
Business performance	Total market share growth (X11)	Naidoo (2010)
	Total sales growth (X12)	
	Return on sales (X13)	

Table 2 Description of samples

Demographic information	Category	Frequency (N = 125)	Percentage (%)
Gender	Male	95	76
	Female	30	24
Age	20s	25	20
	30s	40	32
	40s	50	40
	50s and over	10	8
Education	College students	100	80
	College degree	20	16
	Graduate students	2	1.6
	Graduate degree	3	2.4
Working duration of company	Less than 1 year	5	4
	More than 1 year	20	16
	More than 2 year	50	40
	More than 3 year	23	18.4
	More than 4 year	15	12
	More than 5 year	12	9.6

Table 3 Individual item reliability and AVEs for latent constructs

Factor	Variables	Standardized estimate	Error term	Construct reliability	Average variance extracted
Innovation orientation	X1	0.902	0.141	0.947	0.817
	X2	0.892	0.146		
	X3	0.909	0.153		
	X4	0.809	0.254		
Customer orientation	X5	0.832	0.217	0.882	0.717
	X6	0.855	0.208		
	X7	0.651	0.303		
Service quality	X8	0.786	0.305	0.909	0.770
	X9	0.896	0.148		
	X10	0.864	0.196		
Business performance	X11	0.878	0.17	0.896	0.700
	X12	0.825	0.188		
	X13	0.748	0.343		

4 Research Method

4.1 Analysis of Reliability and Validity

As this study used a survey method to gather data for analysis, two types of errors related to survey measurement need to be examined; random error and systemic error. Random errors are statistical fluctuations in the measured data due to the accuracy limitations of the measurement instrument. Systemic errors are reproducible inaccuracies due to any problems that occur consistently in the same direction. In general, reliability and validity analyses are used to evaluate these two errors.

Reliability is the consistency of a set of measurements. Reliability is the degree to which a variable or concept is measured consistently. Validity is the degree to which the intended variables are actually measured. Reliability and validity were examined by utilizing Cronbach's α and factor analysis. The desired lower limit for Cronbach's α is 0.7. Thus, the internal consistency of the measurement scales is verified. In other words, the various questions for each construct measured the same construct. Table 3 shows the results of the reliability analysis.

4.2 Convergent and Discriminant Validities

Reliability tests look only at the items in the scale and do not compare across constructs. To compare one variable with other variables, a validity test should be performed. Confirmatory factor analysis was performed to establish factorial

Table 4 Correlation matrix and AVE for latent variables

Component	Innovation orientation	Customer orientation	Service quality	Business performance
Innovation orientation	**0.817**[a]			
Customer orientation	0.385	**0.717**		
Service quality	0.410	0.757	**0.770**	
Business performance	0.149	0.282	0.478	**0.700**

[a]Diagonal matrix is an average variance extraction

validity in SEM as Fornell and Larcker (1981) suggested. Confirmatory factor analysis measured construct validity. After the dropout and modification of measures from the previous confirmatory factor analysis, two additional validities were employed to ensure the validity of measures. Convergent validity and discriminant validity are both considered subcategories of construct validity.

Convergent validity adopts the measure of average variance extracted (AVE) to gauge the percentage of explained variance by indicators relative to measurement errors. To establish convergent validity, AVE should be greater than 0.5 as suggested by Fornell and Larcker (1981).

AVE value can also be used to measure the amount of variance that a latent variable component captures from its indicators. Fornell and Larcker (1981) suggest AVE should be greater than 0.5 to account for 50% or more variance of indicators. Results of CFA are the followings. Chi-Square = 11.931, df = 59, RMR = 0.0271, GFI = 0.871, AGFI = 0.801, NFI = 0.929, NNFI = 0.959.

The way to establish discriminant validity is to compare the square root of the AVE of each construct to the correlations of this construct to all other constructs. Fornell and Larcker (1981) suggest that the square root of AVE should be greater than the corresponding correlations among the latent variables. The results shown in Table 4 demonstrate all latent variables exhibit high discriminant validity since each construct had a square root of AVE bigger than its correlations with other constructs. This result ensures that the measurement model has the discriminant validity.

4.3 *Structural Equation Modeling*

There are several indexes for goodness of fit. In this study, chi-square, GFI, RMR, CFI, NFI, NNFI, and RMSEA were used to test the goodness of fit of the research model as recommended. In this study, the overall fitness of the proposed research model is satisfactory on several measures for goodness of fit (χ^2/df = 1.305, GFI = 0.99, CFI = 0.99, NFI = 0.98, NNFI = 0.99, RMSEA = 0.005).

The structural model investigates the strength and direction of the relationships among theoretical latent factors. The structural model and hypotheses are tested by examining the path coefficients. In addition to the individual path tests, the explained variance in the dependent factors is assessed as an indication of the overall predictive strength of the model.

Based on the research model in Fig. 2, the three hypothesized paths are tested. Table 5 and Fig. 3 show the results of the structural model. The path coefficients indicate the strength of paths. The *T*-values were used to determine if the hypothesized relationships were significant.

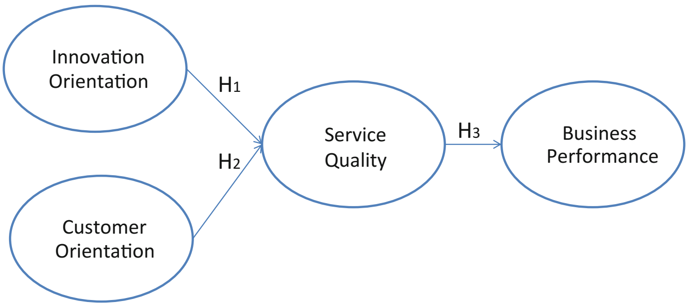


Fig. 2 Research model

Table 5 Results of path analysis

Path relationship	Estimate	Standardized error	*T*-value	Result
H_1: Innovation orientation → Service quality	0.14	0.063	2.21	Supported
H_2: Customer orientation → Service quality	0.70	0.063	11.19	Supported
H_3: Service quality → Business performance	0.48	0.080	6.01	Supported

Fig. 3 The results of the structural model

5 Conclusion

All of business sectors face turbulent business environment. VUCA (Volatility, Uncertainty, Complexity, and Ambiguity) is representative environment factor. In this circumstance, business sector should do innovation, customer orientation operation.

In empirical research, Innovation orientation and customer orientation significantly impact on service quality. And service quality significantly impacts on business performance. Through this result, company should sustain innovation, customer orientation operation. This activities significantly impact on service quality. Customer orientation innovation is not option problem but survival problem.

References

Baker, W. E., Sinkula, J. M. (1999): Market orientation, learning orientation and learning orientation on organizational performance, *Journal of Academy Market Science*, Vol. 27, No. 4, pp. 411–427.

Cheng, C. C., Krumwiede, D. (2012), The role of service innovation in the market orientation-new service performance linkage, Technovation, 32, pp. 487–497.

Damanpour, F. (1992), Organizational size and innovation, *Organization Studies*, 13(3), pp. 375–402.

Dawar N. (2013), When marketing is strategy, *Harvard Business Review*, December, pp. 101–108.

Fornell, C. and D.F. Larcker (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 48, pp. 39–50.

Hurley, R. F., Hult, T. M. (1998), Innovation, market orientation, and organizational learning: An integration and empirical examination, *Journal of Marketing*, 62(3), pp. 42–54.

Johnson, A. J., Dibrell, C. C., Hansen, E. (2009): Market Orientation, Innovativeness, and Performance of Food Companies, *Journal of Agribusiness,* Vol. 27, 1/2, Spring/Fall 2009, pp. 85–106.

Lawson, B., Samson, D. (2001), Developing innovation capability in organizations: A dynamic capabilities approach, *International Journal of Innovation Management*, 5(3), pp. 337–400.

Naidoo, V. (2010), Firm survival through a crisis: The influence of market orientation, marketing innovation and business strategy, *Industrial Marketing Management*, 39, pp. 1311–1320.

Thornhill, S. (2006), Knowledge, Innovation and firm performance in high-and low-technology regimes, *Journal of Business Venturing*, Vol. 21, No. 5, pp. 687–703.

Author Biography

Gye-Soo Kim is Chairman of Future & Vision. He is Professor for Department of Business Administration, Semyung University. He graduated from Kyung Hee University in Korea with Bachelor. He graduated from Master Degree Korea University in Korea. He graduated from Kyung Hee University in Korea Ph.D. and worked in Amore Company in the Planning and Strategy field. His interest is mainly in Structural Equation Modeling, Customer Satisfaction, and had taught many SEM lectures in both Researching Center. He also won two Best Paper Awards in 2014 KSQM Spring Congress.

Service Quality Comparison

Chin Hon Tan, Guanghua Han and Thong Ngee Goh

Abstract In studying the service quality of a service provider, customers are often asked to rate the services they have received through a questionnaire. Based on the responses received, various models/techniques that analyze the scores of different components have been proposed to evaluate overall service quality. In particular, the average score of a service provider often serves as a basis for comparison between two or more service providers. In practice, the scores associated with each response level (e.g., agree, strongly agree) are often based on intuition or some simple rule, which may not correctly reflect the differences between different response levels. Therefore, it is important to examine the robustness of a conclusion (e.g., service provider A is better than service provider B) with respect to changes in the scores assigned to each response level when studying the service quality of service providers through customer surveys. It is commonly assumed that the difference between the average scores of two service providers provides a good indication of the robustness of the conclusion. In this paper, we illustrate that this may not be true and highlight that a separate analysis on the robustness of a conclusion is necessary. We describe how the robustness of a conclusion can be represented and show how it can be computed in practice.

Keywords Quality management · Customer service · Comparison · Service · Measure of robustness

C.H. Tan (✉) · G. Han · T.N. Goh
Department of Industrial and Systems Engineering, National University of Singapore,
1 Engineering Drive 2, Singapore 117576, Singapore
e-mail: isetch@nus.edu.sg

G. Han
e-mail: isehg@nus.edu.sg

T.N. Goh
e-mail: isegohtn@nus.edu.sg

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_17

1 Introduction

The quality of consumer goods can often be measured in quantitative terms. For example, the life-span of a product can be a good indication of its quality. However, service quality is generally much harder to measure. Because services are intangible, heterogeneous and inseparable, service quality is often evaluated based on the personal experiences.

Quality service results in good consumer experience that can lead to greater customer loyalty and higher profits. Reichheld and Sasser (1990) estimate that a 5% increase in consumer loyalty, a key indicator of service quality can increase profit by 25–85%. For example, fast-food giant Taco Bell finds that its stores with low workforce turnover have more sales and 55% higher profits than stores with high turnover (Heskett et al. 1994).

Service quality depends on multiple factors (e.g., the type of service, situation, time, consumer's expectations, etc.) and various models/techniques based on the scores of different components, have been proposed to evaluate overall service quality. These include SERVQUAL (Parasuraman et al. 1988), SERVPERF (Cronin and Taylor 1994), the critical incident technique (Bitner et al. 1990), and the sequential incident technique (SIT) (Stauss and Weinlich 1997).

In the SERVQUAL model, service quality is a function of the gap between performance and expectation along five dimensions (i.e., tangibles, reliabilities, responsiveness, assurance, and empathy). However, some researchers argue that the consumer experienced performance is the only marker of perceived service quality and believe that the SERVPERF model is better at measuring service quality. In the SERVPERF model, service quality is formulated as a function of the consumer experienced performance, rather than the gap between performance and expectation. Unlike SERVQUAL and SERVPERF, which are primarily quantitative multi-attribute measurements, the critical incident technique (CIT) first identifies most satisfactory and unsatisfactory events and behaviors, before exploring the causes of the service encounter. The SIT is an alternative method that records all incidents that customers perceive in a specific service transaction. Due to the time, cost, and complexity involved when applying SIT in practice, SIT is more suitable for measuring services with clearly defined and largely standardized processes. These service quality models, which attempt to identify the relationships between salient variables, can be particularly helpful in identifying areas of improvements (Lim et al. 1999).

Service has great richness and diversity of meaning in various industries. Hence, it is not surprising that different researchers have studied service quality under different service contexts by focusing on different aspects. Table 1 lists the different aspects of service quality that were considered in some of these studies.

Clearly, a customer's perception of the service that he or she has received is influenced by multiple components. However, some components are likely to be more influential than others. Hence, an analyst may use experiential methods

Table 1 Aspects of service quality considered under different service contexts

Services	Quality measurement aspects
IT Zhu et al. (2002)	Needs of personal attention, age, personal interaction, service fee, etc.
Mobile telecommunication Hosseini et al. (2013)	Network, value-added, pricing plans, billing system, customer service, etc.
e-service Yang and Jun (2002)	Security, personalization, access, responsibility, availability, etc.
Banking Islam and Ahmed (2005)	Service on-time, Attention to clients, Tangible facilities, etc.
Hotel Wu and Ko (2013)	Conduct, expertise, problem solving, atmosphere, facility, etc.
Business consulting market Benazić and Došen (2012)	Service potential, service process, service result
Secondary school Ramseook–Munhurrun and Nundlall (2013)	Empathy, facilities, reliability, responsiveness, assurance-discipline
Supply chain Gupta and Singh (2012)	Distributor service, customer service, organization service, etc.

(e.g., Rakhmawati et al. 2013) and/or technical methods (e.g., Islam and Ahmed 2005) to select the most important and critical components in practice.

In this paper, we study the problem of comparing between the services of two service providers by asking the customers of the service providers a single question, "How would you rate the quality of service you received?"

Various kinds of rating scores have been developed to measure the differences in intensity when analyzing customer responses. For example, Likert (1932) developed the principle of measuring attitudes levels by asking people to select one of several choices (e.g., "never," "rarely," "sometimes," "frequently," and "always"). Based on the work by Likert (1932), more Likert-type scale formats, such as a 4-point scale (Erto and Vanacore 2002), a 5-point scale (Grobelna and Marciszewska 2013), a 7-point scale (Awan and Mahmood 2013), a 10-point scale (Yurtseven 2005), are proposed to deal with different problems.

The scales of each dimension are usually experientially designed and the intensities of attitudes are assumed to be linear in these studies (e.g., on a continuum from strongly agree to strongly disagree). The average score of a service provider, based on the responses received, often serves as a basis for comparison between two or more service providers.

However, it is arguable that customer responses should be seen as ordinal, rather than interval, data. This is because the relative difference between consecutive ratings is, in general, unclear. For example, it is not immediately clear if the difference between "never" and "rarely" should be the same as the difference between "rarely" and "sometimes."

It is noted that the conclusions (e.g., service provider A is better than service provider B) are generally insensitive to minor deviations in the scores associated with the different intensity levels. In this paper, we highlight the conditions, where a conclusion remains valid and the conditions where it may not.

2 Problem Description

Suppose we have two service providers X and Y. Customers of X and Y were asked to rate their service providers by choosing one of five possible levels:

- VP: very poor
- P: poor
- A: average
- G: good
- VG: very good

One thousand and eleven hundred customers were surveyed for X and Y, respectively. It is believed that the difference between VG and G is twice the difference between G and A. Similarly, it is believed that the difference between P and VP is twice the difference between A and P. Without loss of generality, a score of 1, 3, 4, 5, and 7 is assigned to VP, P, A, G, and VG, respectively. The customer responses are summarized in Table 2.

Based on their average scores, X appears to be marginally better than Y. In particular, the average score of X is only $\frac{5.98}{5.84} - 1 = 2.4\%$ greater than the average score of Y. Hence, one might suspect that the conclusion (i.e., X is better than Y) is not robust to deviations in the scores associated with the different levels.

However, the conclusion is completely robust to any deviations in scores. In particular, there does not exist a nondecreasing set of scores where the average score of X is less than the average score of Y (the skeptical reader is invited to attempt finding such a set of scores). Clearly, the difference between the average scores of X and Y is a poor indicator of the robustness of the conclusion with respect to uncertainties in the scores.

Table 2 Average score and customer responses for X and Y

	VP	P	A	G	VG	Count	Average score
Score	1	3	4	5	7	–	–
Difference	–	2	1	1	2	–	–
X	2 (0.2%)	3 (0.3%)	65 (6.5%)	400 (40%)	530 (53%)	1000	5.98
Y	40 (3.6%)	8 (0.7%)	40 (3.6%)	440 (40%)	572 (52%)	1100	5.84

Table 3 Customer responses for Z

	VP	P	A	G	VG	Count	Average score
Z	2 (0.2%)	41 (3.4%)	82 (6.8%)	620 (51.7%)	455 (37.9%)	1200	5.62

3 An Appropriate Measure of Robustness

In many cases, the ordering of the levels is known but the differences between different levels are uncertain. Without loss of generality, the scores are assumed to nondecreasing. Suppose that the true difference between the levels differ from their initial estimates by a factor of no more than $(1+\tau)$ such that:

- Difference between P and VP is between $\frac{2}{1+\tau}$ and $2(1+\tau)$
- Difference between P and A is between $\frac{1}{1+\tau}$ and $1+\tau$
- Difference between A and G is between $\frac{1}{1+\tau}$ and $1+\tau$
- Difference between G and VG is between $\frac{2}{1+\tau}$ and $2(1+\tau)$

Observe that the higher the value of τ, the higher the robustness of the conclusion with respect to uncertainties in the differences between different levels. The largest possible value of τ can be obtained using the procedure described in Tan et al. (2014).

When comparing between X and Y, τ can be infinity large. This is consistent with the observation that there does not exists a nondecreasing set of scores where the average score of X is less than the average score of Y.

Consider a third service provider Z with the survey results listed in Table 3. When comparing between Y and Z, the largest value that τ can take is 1.012. This implies that the average score of Y is at least as large as the average score of Z if the difference between P and VP, P and A, A and G, and G and VG are within [0.99, 4.02], [0.50, 2.01], [0.50, 2.01], and [0.99, 4.02], respectively. If the differences between the levels lie outside these intervals, there is no guarantee that the average score of Y is at least as large as the average score of Z.

4 Conclusion

The average scores of service providers, based on customer responses, often serves as a basis for comparison between two or more service providers. However, the differences between possible response levels are often unclear. In this paper, we described how the robustness of a conclusion with respect to these uncertainties can be compared. It is commonly assumed that the difference in average score provides a good indication of the robustness of a conclusion. In this paper, we highlight that this may not be true and a separate analysis of the robustness of a conclusion is necessary.

References

Awan, M.U., and Mahmood, K. (2013): “Development of a service quality model for academic libraries”, *Quality and Quantity*, 47(2), pp 1093-1103.

Benazić, D., and Došen, D.O. (2012): “Service quality concept and measurement in the business consulting market”, *Trziste/Market*, 24(1), pp 47–66.

Bitner, M.J., Booms, B.H., and Tetreault, M.S. (1990): “The service encounter: Diagnosing favorable and unfavorable incidents”, *Journal of marketing*, 54(1), pp 71.

Cronin J.J., and Taylor, S.A. (1994): “SERVPERF versus SERVQUAL: Reconciling performance-based and perceptions-minus-expectations measurement of service quality”, *Journal of marketing*, 58(1), pp 125–131.

Erto, P., and Vanacore, A. (2002): “A probabilistic approach to measure hotel service quality”, *Total Quality Management*, 13(2), pp 165–174.

Grobelna, A., and Marciszewska, B. (2013): “Measurement of service quality in the hotel sector: the case of Northern Poland”, *Journal of Hospitality Marketing and Management*, 22(3), pp 313–332.

Gupta T.K., and Singh V. (2012): “Systematic model development to analyze service quality in supply chain for a manufacturing organization”, *Proceedings of the National Conference on Trends and Advances in Mechanical Engineering*, October 2012, pp 19–20.

Heskett, J.L., Jones, T.O., Loveman, G.W., Earl Sasser, W., and Schlesinger L.A. (1994): “Putting the Service-Profit Chain to Work”, *Harvard Business Review*, 72(2), pp 164–174.

Hosseini, S.Y., Zadeh, M.B., and Bideh, A.Z. (2013): “Providing a multidimensional measurement model for assessing mobile telecommunication service quality (MS-qual)”, *Iranian Journal of Management Studies*, 6(2), pp 7–29.

Islam, N., and Ahmed, E. (2005): “A measurement of customer service quality of banks in Dhaka city of Bangladesh”, *South Asian Journal of Management*, 12(1), pp 37–57.

Likert, R. (1932): “A technique for the measurement of attitudes”, *Archives of psychology*, 140 (22), pp 55.

Lim, P.C., Tang, N.K., and Jackson, P.M. (1999): “An innovative framework for health care performance measurement”, *Managing Service Quality*, 9(6), pp 423–433.

Parasuraman, A., Zeithaml, V.A., and Berry, L.L. (1988): “Servqual”, *Journal of retailing*, 64(1), pp 12–37.

Rakhmawati, T., Sumaedi, S., Bakti, I., Astrini, N.J., Yarmen, M., Widianti, T., and Vebriyanti, D. I. (2013): “Developing a service quality measurement model of public health center in Indonesia”, *Management Science and Engineering*, 7(2), pp 1–15.

Ramseook-Munhurrun, P., and Nundlall, P. (2013): “Service quality measurement for secondary school setting”, *Quality Assurance in Education*, 21(4), pp 387–401.

Reichheld, F.P, and Sasser Jr W.E. (1990): “Zero Defections: Quality Comes to Services”, *Harvard Business Review*, 68(5), pp 105–111.

Stauss, B., and Weinlich, B. (1997): “Process-oriented measurement of service quality: applying the sequential incident technique”, *European Journal of Marketing*, 31(1), pp 33–55.

Tan, C., Han, G., and Goh, T. (2014): “Service quality comparison”, *Working paper*.

Wu, H.C., and Ko, Y.J. (2013): “Assessment of service quality in the hotel industry”, *Journal of Quality Assurance in Hospitality and Tourism*, 14(3), pp 218–244.

Yang, Z., and Jun, M. (2002): “Consumer perception of e-service quality - from internet purchaser and no-purchaser perspectives”, *Journal of Business Strategies*, 19(1), pp 19–41.

Yurtseven, H.R. (2005): “Service Quality of Troy: An Importance-satisfaction Analysis”, *TOURISMOS: An International Multidisciplinary Refereed Journal of Tourism*, 1(2), pp 61–75.

Zhu, F.X., Wymer, W., and Chen, I. (2002): “IT-based services and service quality in consumer banking”, *International Journal of Service Industry Management*, 13(1), pp 69–90.

Authors Biography

Chin Hon Tan is an Assistant Professor in the Department of Industrial and Systems Engineering at the National University of Singapore. He obtained his B.Eng. from the National University of Singapore (with First Class Honors) and his M.S. and Ph.D. from the University of Florida. His research interests include sensitivity analysis and decision making under uncertainty.

Guanghua Han is servicing as a Research Fellow in the Department of Industrial and Systems Engineering at the National University of Singapore. He obtained his B.E and M.S from Qingdao University, China and Ph.D. from Shanghai Jiao Tong University. He has won several awards including Scholarship Award for Excellent Doctoral Student granted by China Ministry of Education, Qingdao City Investigation Achievement Award, Hongkong Gu Guohua Ph.D. Scholarship, etc. His research interest is mainly in service quality measurement and optimization in supply chain decisions.

Thong Ngee Goh is a professor in the Department of Industrial and Systems Engineering at National University of Singapore. His previous administrative appointments include Head of Department of Industrial and Systems Engineering, Dean of Faculty of Engineering, and Director of Office of Quality Management. Prof. Goh obtained his B.E. from the University of Saskatchewan, Canada and Ph.D. from the University of Wisconsin-Madison, USA. Prof. Goh is a Founding Fellow of the Academy of Engineering Singapore. He is Academician of the International Academy for Quality (IAQ), Fellow of the American Society for Quality (ASQ) and Honorary Member of the Singapore Quality Institute (SQI). Professor Goh's recent international honors and awards include: "Educator of the Year" Award of the IEEE Engineering Management Society, the inaugural Dr. Masing Book Prize of the IAQ, the William G. Hunter Award of the ASQ Statistics Division, the Harrington-Ishikawa Medal of the Asia-Pacific Quality Organization, and the Grant Medal of the ASQ.

Part VII
Quality For Environmental Sustainability

Prof. T.N. Goh, National University of Singapore

This section comprises papers that address a growing global concern, namely environmental sustainability. Topics range from perceptions of the lay public at the ground level to the use of sophisticated mathematical models for environmental data, and contributors came from many countries that actually face environmental issues. The impression is Quality Engineering at work, rather than Management. It should be realized, however, that to win the war against environmental woes there must first of all be management at high levels such as those of corporations and governments. Unlike manufacturing, the "bottom-up" approach will not be effective and sustainable, despite the occasional usefulness of demonstrations, roadshows, etc., in bringing about citizens' awareness and concern. The solutions to environmental problems and ways for environmental protection still lie in the judicious application and development of wide ranges of management and engineering tools, some of which with analytical rigor. Through the presentations in this section, readers would gain an idea of the progress made up to the time of the Congress, and could expect more in the days to come.

About the author

The author brief biograph can be found in Chapter 1.

Air Quality Modeling via PM2.5 Measurements

Min Zhou and Thong Ngee Goh

Abstract Environmental quality for the general population is dominated by air quality. Thus modeling of air quality is the first step toward any program for quality improvement. This paper describes the use of the ARIMA (Autoregressive Integrated Moving Average) time series modeling approach, illustrated by the tracking of the daily mean PM2.5 concentration in the north region of Singapore. A framework for ARIMA forecasting revised from the general Box-Jenkins procedure is first outlined; T-test and three information criteria, namely, AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), HIC (Hannon-Quinn Information Criterion) are employed in addition to analyses on the ACF (autocorrelation function) and PACF (partial autocorrelation function). With forecasting as the primary objective, the emphasis is on out-of-sample forecasting more than in-sample fitting. It is shown that for 30 such forecasts, one-step ahead MAPE (mean absolute percentage error) has been found to be as low as 8.0%. The satisfactory result shows the classical time series modeling approach to be a promising tool to model compound air pollutants such as PM2.5; it enables short-term forecasting of this air pollutant concentration for public information on air quality.

Keywords PM2.5 · ARIMA · Singapore · Forecasting

1 Introduction

Environmental and human health issues associated with outdoor air quality especially in metropolitan regions, are always drawing concern from both policy official and the general public. Due to the burst of urban expansion in last two decades, the

M. Zhou (✉) · T.N. Goh
Industrial and Systems Engineering Department,
National University of Singapore, Singapore, Singapore
e-mail: zhoum@nus.edu.sg

T.N. Goh
e-mail: isetngoh@nus.edu.sg

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_18

world has witnessed a boom of urbanization process throughout the world. Meanwhile, the frequently breakout of the haze or smoke hit in many cosmopolitan cities has raised the public consciousness about one kind of air pollutant named particulate matters in recent year. Particulate matter (PM), also known as particle pollution, refers to a mixture of extremely small solid particles and liquid droplets found in the ambient air. Unlike other air pollutants, it comes in a variety size and a number of components. In general, the smaller and lighter a particle is, the longer it will stay in the air and the stronger its potential impact on human health,. and the fine particles (inhalable particles less than 2.5 μm in diameter), such as those found in smoke and haze, are not only responsible for the substantial loss of work and school days but also has been confirmed the strong association with the morbidity and mortality (Boldo et al. 2006; Pope III et al. 2009). The PM2.5 level has been taken as the one of the main indicator for air quality assessment in many cities. Besides the investigations for improving understanding of the potential health effects of the fine particles, building mathematical model is an indispensable way to learn the trends of the pollutant and then to assess the effectiveness of control measures.

Air quality models are essential tools to evaluate and predict the pollutant dispersion. To represent the chronological and spatial movement of the pollutants, various highly sophisticated multiparameter meteorological models or physical/chemical mechanism-based models have been developed (Christian 2001). However, it is observed that some of them fail to satisfactorily predict the air pollutants concentration fast and accurately. In fact, the uncertainty measurements on the complex meteorological conditions combining the complexity of the physical and chemical processes undergoing in the atmosphere make it very difficult for those models response quickly and nicely. Therefore, for short-term forecasting purpose, some statistical analysis techniques are appreciated. Proposed as a tool for economic time series modeling, Box-Jenkins autoregressive integrated moving average (ARIMA) model has confirmed a promising air pollutant forecasting tool (Robeson and Steyn 1990; Slini et al. 2002; Tiao et al. 1975; Ujjwal and Jain 2009). Conceptually, this nonparametric method attempts to explore the appropriate internal representation of the time series data (Ho et al. 2002). Few assumptions are needed and no a priori postulation of models or external information is required, which enables it to successfully describe most oscillatory series.

In this present study, we have developed a predictive model for tracking PM2.5 plumes and forecasting of air quality in Singapore based on classic ARIMA method. For 30 out-of-sample forecasts, one-step ahead MAPE (mean absolute percentage error) has been found to be 8.031%. The satisfactory result shows that this classical univariate time series modeling method can be effectively used for short-term air quality forecasting purpose.

2 Measured Data and Methods

2.1 Date Description

To measure and assess the ambient air quality, the Ministry of the Environment in Singapore monitors the level of air pollutants by a telemetric air quality monitoring and management system. This air quality monitoring network has 15 stations located throughout the island which are linked via the public telephone network to a central control station (CCS) at the Environment Building. Twelve of the air monitoring stations are ambient stations which monitor the general quality of the air over Singapore. Three are roadside station and they are used to assess the effectiveness of the Ministry's vehicular emission control programmes. As the increasing concern in the public about the air quality, the pollutant concentration has been published on the website of national environmental agency (NEA) since August 24, 2012. Due to the air haze stemming from the agricultural fields burning of Indonesia in June, 2013, the 24 h monitoring system has been launched.

In this present work, concentration of PM2.5 of the north region of Singapore from September, 2012 to August, 2013 has been collected from websites of NEA of Singapore (http://www.nea.gov.sg/psi/). As the PM2.5 level varies from hour to hour, the daily mean has been obtained. A total of 2 missing values have been encountered out of 365 data. In the preprocessing, 22 abnormal data have been excluded. Consequently, the sample size is reduced to 341.

Owing to its geographical location and maritime exposure, Singapore's climate has been classified into rainforest climate with uniform temperature and pressure, high humidity, abundant rainfall. Though there is no true distinct seasons, the monsoon season that happen twice each year have prevent the all-year uniformity. Therefore, there is some seasonal variation on the PM2.5 concentration. Actually, from November to March which we can loosely call Northeast Monsoon season, the PM2.5 concentration is much weaker than the Southwest Monsoon period which is from June to September. However, only one year consistent continuous data for PM2.5 concentration level was accessible when the study was conducted. Hence, seasonal characteristic may not be shown in the fitted model.

To assess the out-of-sample prediction performance, 30 new data from September 1, 2013 to September 30, 2013 have been collected afterwards.

2.2 ARIMA Modeling

2.2.1 ARIMA (p, d, q) Model

Autoregressive Integrated Moving Average (ARIMA) model basically is a mix of three processes, autoregressive, differencing and moving average. The basic concept of a AR (p) model is that for a stationary series x_t, the current value can be explained as a weighted sum of p past values of the x's, plus an added white noise process a_t. And a MA (q) model allows us to represent x_t as a weighted sum of present and past values of white noise a's. For a nonstationary series, it is assumed that the series can be stabilized by taking the dth difference.

Hence, giving an arbitrary time series z_t, $t = 0, \pm 1, \pm 2, \ldots$, the corresponding autoregressive moving average (ARIMA) model with orders (p, d, q) have the form (Box et al. 2005):

$$\phi(B)(1 - B)^d z_t = \theta(B) a_t,$$

where B is denoted as the backshift operator with $Bz_t = z_{t-1}$, then

$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p, \quad \phi_p \neq 0,$$
$$\theta(B) = 1 - \theta_1 B - \cdots - \theta_p B^q, \quad \theta_q \neq 0,$$

and white noise a_t follows normal distribution with mean zero and constant variance σ^2, respectively.

2.2.2 Autocorrelation and Partial Autocorrelation Function

In practice, the order of an ARIMA model is always unknown and need to be inferred from the available dataset. And one very useful way to glean insights is to consider the sample autocorrelation function (ACF) and partial autocorrelation functions (PACF).

Measuring the correlation between two values Z_t and Z_{t-k} which is a distant k apart, the autocorrelation coefficient is defined by

$$\rho_k = \frac{\text{Cov}(Z_t, Z_{t-k})}{\sqrt{\text{var}(Z_t) \cdot \text{var}(Z_{t-k})}}$$

and the plot of autocorrelation coefficient ρ_k versus lag k is called the autocorrelation function plot of the process. Since $\rho_k = \rho_{-k}$, in practice only the positive half of the function has been plotted as it is inevitably symmetric about zero. ACF reveals how the correlation between any two values of the series changes as their separation changes so it is associate with the terms of moving average part.

Partial autocorrelation coefficient is employed to exploit the extent of the lag in an autoregressive model. It is the amount of correlation between a variable and a lag of itself that is not explained by correlations at all. Denote ϕ_{kj} as the jth coefficient in an autoregressive process of order k (or AR(k)), then ϕ_{kj} satisfy the set of equations: $\rho_j = \sum_{i=1}^{k} \phi_{ki} \rho_{k-i}$, and the last coefficient ϕ_{kk} is defined as the partial autocorrelation of the time series z_t.

For different stationary time series processes, the ACF and PACF plots may display very different patterns. Hence they can provide the first step in model

Table 1 ACF and PACF behavior of AR, MA and ARMA process

	AR (*p*) process	MA (*q*) process	ARMA (*p*, *q*) process
ACF	Infinite (damped exponentials and/or damped sine waves)	Finite	Infinite (damped exponentials and/or damped sine waves after first *q*–*p* lags)
	Tails off	Cuts off after lag *q*	Tails off
PACF	Finite	Infinite (dominated by damped exponentials and/or damped sine waves)	Infinite (dominated by damped exponentials and/or damped sine waves after first *p*–*q* lags)
	Cuts off after lag *p*	Tails off	Tails off

identification. Box et al. (2005) has summarized some ACF and PACF behaviors of the three kind stationary ARIMA models, which is listed it in Table 1.

In general, ACF and PACF are associated with the order of MA and AR process respectively. Table 2 have summarized the ACF and PACF patterns of some particular important cases that AR process and MA process of first order and second order and the simple ARMA (1, 1) process.

Actually, with the assistant of software such as Minitab, you could just try some different combination of p and q, then selecting the one that works best. But before this, some systematic analysis on the on the ACF and PACF plots of the series may help tentatively to identify of the numbers of the terms effectively.

2.2.3 Information Criterion for Model Identification

Visual inspection of ACF and PACF provides us a useful way for tentatively identify the orders of AR and MA model. But the Box-Jenkins methodology is very subjective which depends greatly on the experience and skill of the forecasters, and it is noted there are not any theoretical model for reference when it comes the mixed ARMA with high order. Moreover, it can be quite difficult to interpret the ACF and PACF for this case. Hence some alternative methods have been proposed for assisting the order selection. Penalty function statistics, such as Akaike Information Criterion (AIC) (Akaike 1974), Bayesian information Criterion (BIC) (Schwarz 1978) and Hannan Quinn Criterion (HOC) (Hannan and Quinn 1979) are widely used three. These statistical model fit measures assess a range of models and the one with the minimal value of the defined statistic is selected.

For an ARMA model, these statistics take the forms (Meyler et al. 1998):

(1) $\mathrm{AIC} = \log\left(\frac{\mathrm{rss}}{n}\right) + \left(2 * \frac{k}{n}\right)$
(2) $\mathrm{BIC} = \log\left(\frac{\mathrm{rss}}{n}\right) + \left(\log(n) * \frac{k}{n}\right)$
(3) $\mathrm{HQ} = \log\left(\frac{\mathrm{rss}}{n}\right) + \left(2\log(\log(n)) * \frac{k}{n}\right)$

Table 2 Rules for selecting p and q (Box et al. 2005)

Order	ρ_k	ϕ_{kk}
(1, d, 0)	*exp* decay	Only ϕ_{11} nonzero
(0, d, 1)	Only ρ_1 nonzero	*exp* dacay
(2, d, 0)	*exp* or damped sine wave	Only ϕ_{11}, ϕ_{22} nonzero
(0, d, 2)	Only $\rho_{1,2}$ nonzero	*exp* or damped sine wave
(1, d, 1)	*exp* decay	*exp* decay

exp: exponential

where

rss residual sum of squares.
n number of observations.
k number of the parameters to be estimated.

Actually, each of these criterions has their own advantages and disadvantages. HIC is usually cited but seems have seldom been used in practice. AIC shows theoretical advantage over BIC but usually lead to over-fitting as the penalty on number of parameters is quite weak. So it is recommended to carefully examine all of these criterions instead of relying on any single one.

2.2.4 T-Test for Model Identification

Besides the information criterion method, hypothesis testing is another objective way to decide the specified order for p and q.

Given a stationary time series $x(t)$ which can be described by a typical ARMA (p, q) model, then we have

$$x(t) = \beta_1 x(t-1) + \cdots + \beta_p x(t-p) + e(t) + \beta_{p+1} e(t-1) \\ + \cdots + \beta_{p+q} e(t-q) + \bar{u}.$$

Here β_i, $i = 1, \ldots, p$ and β_i, $i = p+1, \ldots, q+p$ are the coefficient of AR term and MA term respectively. As $\widehat{\beta}_i$, $i = 1, \ldots, p+q$, is the unbiased estimators of the true coefficient of the each terms which is regarded as normally distributed, we can test the contribution of each term to the overall model with $H_0 : \beta_i = 0$ VS $H_1 : \beta_i \neq 0$ by employing the test statistic $T_0 = \frac{\widehat{\beta}_i}{\mathrm{se}(\widehat{\beta}_i)}$. Where $\mathrm{se}(\widehat{\beta}_i)$ is denoted as the standard error of $\widehat{\beta}_i$. And the null hypothesis is rejected if $|t_0| > t_{\frac{\alpha}{2}, n-p-q}$ or the P-value greater than the significant level α, which indicates that that term should be included into the model.

Actually, to carry out T-test procedure, it is better to adopt forward selection method which begins with no variables in the model and adds only one at a time since no available rule for setting upper bound on the order of p or q and the addition of the high order may change the significance of the lower ones. As we have discussed before the autocorrelation at lag of high order may be merely due to the propagation of the lower ones.

3 Results and Discussion

Figure 1 outlines the ARIMA modeling and forecasting strategy in our work. Based on general Box-Jenkins procedure, model evaluation has been considered as well. It is worthwhile to note that model building is not a simple sequential process but may involve iterative looping. Moreover, testing and checking should be carried across every stage to ensure the model is reasonable.

3.1 Model Identification and Parameter Estimation

In ARIMA modeling, the first and most important step is to decide the differentiation order d that successfully transforms the original data into a stationary series. Traditionally, two ways can be employed to assess the series. The first is to plot the series against index or time to see if it fluctuates around a constant mean. And the alternative one is by checking the ACF plot. Usually, a series with a fairly fast decreasing ACF should be regarded stationary. If the series is stationary, then $d = 0$ otherwise we should decide the correct amount of differences that yields a stationary series. Generally, performing the first or second difference will effectively transform the original series to a stationary one.

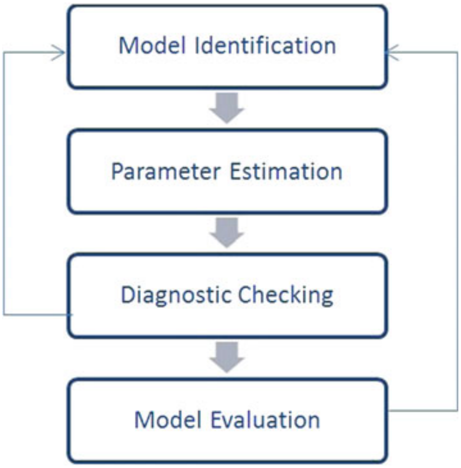


Fig. 1 ARIMA forecasting procedure

Figure 2 is time series plot of the PM2.5 daily concentration from September 1, 2012 to August 31, 2013. We can observe that there are some seasonal variations due to the changes in meteorological conditions which we have explained in Sect. 2.1. Regardless of the seasonal variation and some small range fluctuation, we fail to detect any significant decrease or increase trend when it comes to the whole sequence. So we can tentatively consider the series as stationary. It is noted that the autocorrelation plot in graph 2 has positive autocorrelations with high number which suggests nonstationarity. However, by inspecting the PACF plot, there is a significant spike only at lag 1, which indicates that the high order of the autocorrelation in ACF just merely due to the procreation of the correlation at lag 1. Therefore, it does not violent the previous stationary assumption (Fig. 3).

If the stationarity of the series is still in doubt, the new series can be obtained by taking the first order differenced to helping decide the value of *d*. The corresponding time series plot and probability plot of the new series have been shown in Fig. 4. It can be confirmed that the original series is stationary since the new series fits the normal distribution with the mean zero very well. The same conclusion can also draw from the ACF plot in Fig. 5.

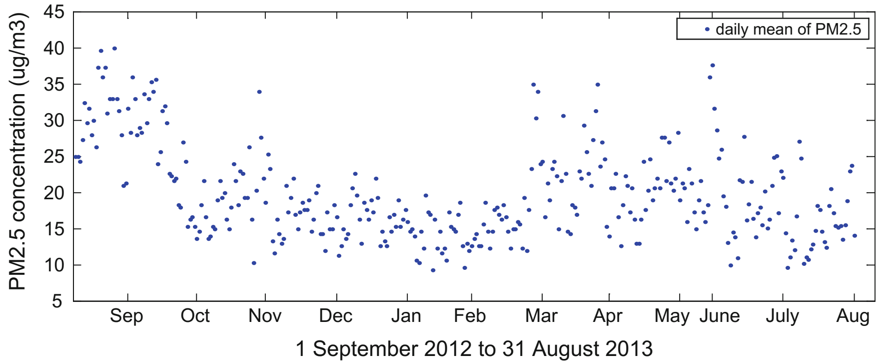


Fig. 2 Sample time series plot

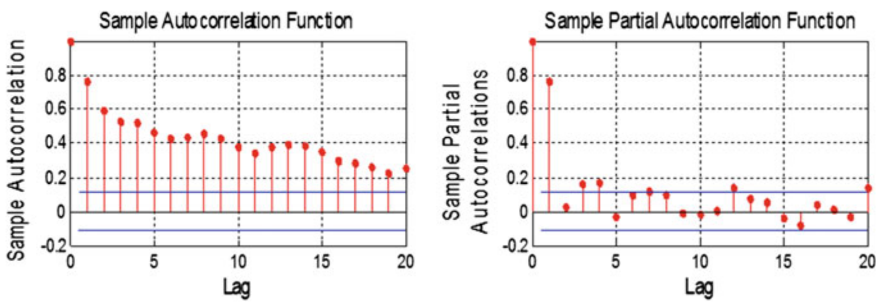


Fig. 3 Sample ACF and PACF plots

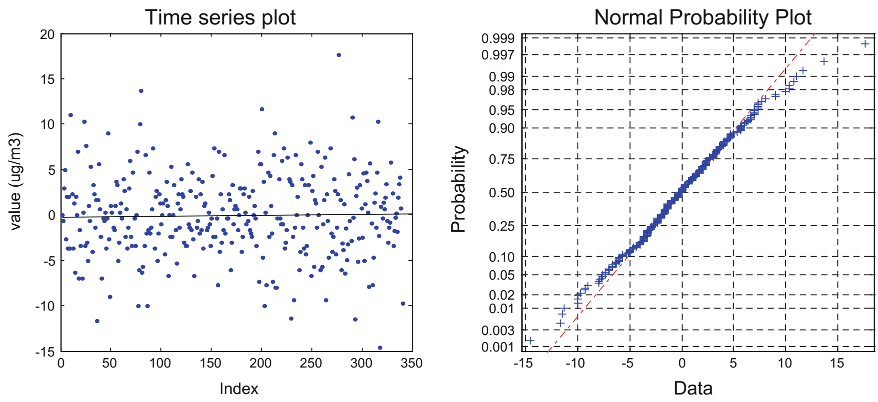


Fig. 4 Time series plot and normal probability plot of differentiated sequence

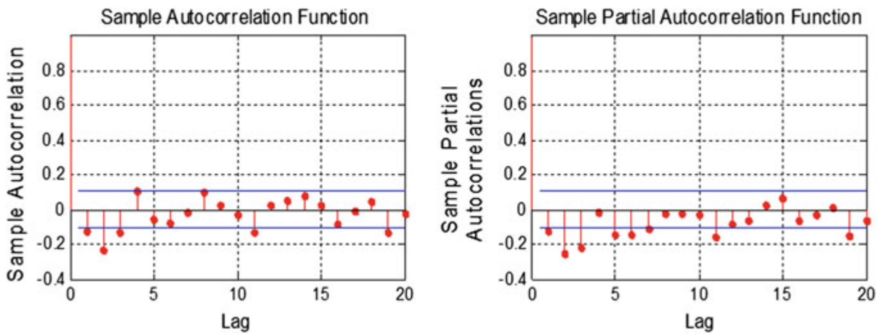


Fig. 5 Smaple ACF and PACF of differentiated sequance

After determining the differentiation order, the next step is to identify the number of AR and MA terms. Based on ACF and PACF plots of our air pollutant data, a mixed ARMA process should be proposed. But it is quite difficult to interpret the specified model from the plots. Therefore, some objective methods introduced in Sect. 2 should be considered to identify order p and q. In our study, both T-test method (Sect. 2.2.3) and information criterion based method (Sect. 2.2.2) have been employed. For information criteria-based method, we set the maximum order as 5 for AR and MA term which is suggested by Minitab. The T-test procedure is started with ARIMA (1, 0, 1) and the significance level has been set as 0.05. The results are displayed in Table 3. We can see that the HIC and BIC criteria come to the same model with $p = 1$ and $q = 3$. And T-test picks a balancing one with the same number of AR and MA terms. With weak penalty on the parameter number, AIC generates a more meticulous model which shows the best in-sample fitting performance with the small root mean square errors (RMSE).

Table 3 Selected models and the corresponding RMSE

	AIC	HIC	BIC	T-Test
Model	(3, 0, 5)	(1, 0, 3)	(1, 0, 3)	(2, 0, 2)
RMSE	3.746	3.803	3.803	4.053

3.2 Diagnostic Checking

Since the models are decided, the next step is to check the model adequacy formally. If the model correctly describes the data, then the residuals should be uncorrelated random variable with mean zero and constant variance. There are quite lot of diagnostic tools help for assessing a model is satisfactory or not. Residual plotting is the very rigorous one. It is frequently helpful to identify the outliers or any anomalies that ruin the fitness of the model. If the sample sizes are quite big, the histogram is meaningful. Otherwise, a normal probability plot is more preferred. Still, plot the residuals (1) versus the fitted value (2) in time sequence can point out any possible problems as well. A second check is to plot the ACF of the residuals. If the model is adequate, the plot should be die out from lag 1.

The normal probability plot and histogram plot of the residuals derived from ARIMA (2, 0, 2) have been exhibited in Fig. 6. As most of the data fall into a straight line in the normal probability plot, there is no evidence shows the model violent the normality assumption. And the ACF and PACF confirm that it is reasonable to consider the residuals as white noise. By conduct the same procedure, the validation of other two models are both confirmed.

It can be noted that in the normal probability plot in Fig. 6, there are some residuals with extreme large value which undoubtedly have a large effect on both the fitting performance and the parameter value. By investigation these corresponding data manually, we find that they are the start points of level shifts whose values are consistent with the latter. Therefore, we kept all of them and the parameters have not been recomputed.

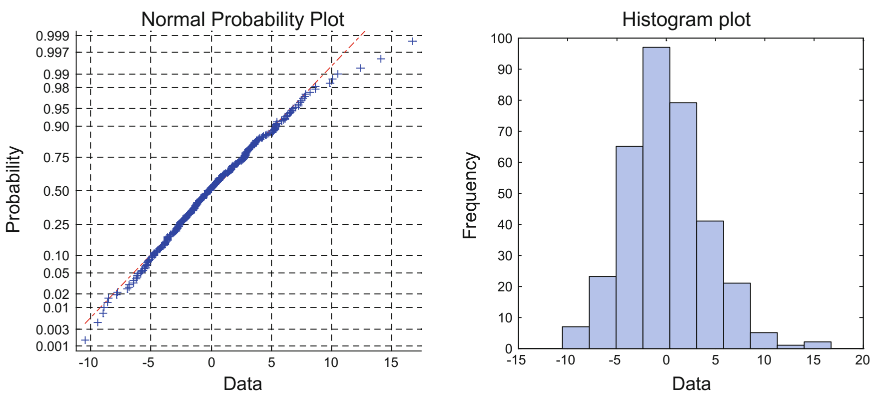


Fig. 6 Normal Probability plot and histogram plot of residuals of ARIMA (2, 0, 2)

3.3 Model Evaluation

Due to underlying principles of the parameter estimation, the in-sample fitting performance sometimes may not guarantee the out-of-sample forecasting ability no matter what modeling paradigm or technique is being used. Therefore, model evaluation is a very important step if the modeling procedure is being utilized for prediction purpose. To assess the out-of-sample performance, it is advised to collect some new observation starting from the end of the sample period. For this purpose, the sequence of daily PM2.5 concentration of September 2013 has been obtained. In this study, we just assess the one-step ahead forecasting performance ability. Besides the RMSE, Mean absolute percentage error (MAPE) has been also adopted as the quantitative measurement for evaluating the models. Scores of these two statistics are displayed in Table 4. It can be seen that ARIMA (1, 0, 3) shows the best forecasting performance with the both the smallest RMSE and MAPE. Nevertheless, the ARIMA (3, 0, 5) which owns the best fitting performance but yields the largest RMSE and MAPE. Compared to the in-sample fitting, ARIMA (2, 0, 2) shows better out-of-sample performance since both RMSE and MAPE are very close to ARIMA (1, 0, 3).

Actually, in ARIMA modeling, an adequately parsimonious model is preferred as it usually comes to better forecasting performance. Based on the RMSE and MAPE, both ARIMA (1, 0, 3) and ARIMA (2, 0, 2) are quite reasonable for describing the studied series, and here we just take be ARIMA (1, 0, 3) due to its optimal in-sample fitting and out-of-sample forecasting performance. Figure 7 depicts the plots of observed and predicted concentrations of PM2.5 where the predicted sequence is generated by ARIMA (1, 0, 3). We can see that the one-step ahead forecasted sequence can catch the trend of the observed series quite well as the two lines are reasonable close to each other (Fig. 8).

The conducted research confirms this classical time series modeling technique, a promising tool to mimic the variation of particulate matter pollutant in metropolitan regions such as Singapore. Based on the above analysis, few observations on ARIMA modeling are highlighted below:

1. If the stationarity of the series is in doubt, take the first difference and draw the conclusion based on the new series.
2. Even though there are some traditional guidelines or instruction for orders selection, such as examining the sample autocorrelation and partial autocorrelation plots graphical, most of them are ambiguous and subjective. Consequently, the performance of the chosen model may depend on the experience and skill of the researchers.

Table 4 Forecasting performance of the models for 30 out of sample forecast

	AIC	HIC	BIC	T-Test
Model	(3, 0, 5)	(1, 0, 3)	(1, 0, 3)	(2, 0, 2)
RMSE	3.746	2.974	2.974	2.985
MAPE (%)	9.873	7.791	7.791	8.031

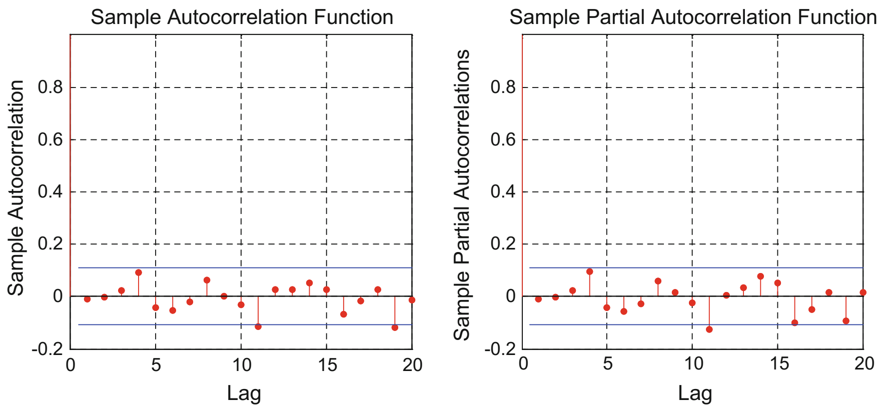


Fig. 7 ACF and PACF plots of residuals of ARIMA (2, 0, 2)

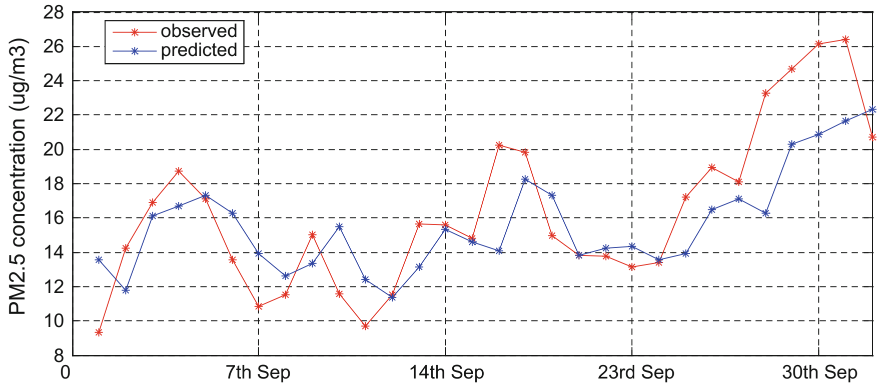


Fig. 8 Plots of observed and predicted concentration of PM2.5

3. Penalty function statistics, such as Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and so on, have been introduced to assist model selection. But different criteria may result in very different models. Still the criteria-based methods may yield meticulous models with a large number of AR or MA terms which maximizes the in-sample performance but the poor predictive power to the out-of-sample. Hence the adequate thresholds for p and q are quite critical.
4. The underlying theory of ARIMA modeling is to utilize the past value and previous error for the purpose of forecasting, therefore the models are intrinsically looking backward. Actually we may see from Fig. 7 that the trend of the predicted sequence is always one-step lagging behind the observed one. Still, the forecasting performance is quite poor in the jump points whose value is very different from the previous one.

5. Graphically examining data is important and useful in every step of ARIMA modeling. However, it is quite subjective and the interpretations vary.
6. ARIMA method does consider any external factors but only the internal relationship of the series itself, hence it may do not able to catch some variance caused by the external factors.

4 Conclusion

This work has considered autoregressive integrated moving average (ARIMA) method to model the PM2.5 variation in north region of Singapore. A framework for ARIMA is outlined with the main steps as follows: model identification, diagnostic checking, and model evaluation. As model identifying is the core part, three approaches, namely visual inspection on ACF and PACF plot, Information criterions and hypotheses testing have been considered to determine adequate models. A more detailing procedure can be summarized as follows:

Step 1 Plot data to identify any seasonal variation, structural breaks and possible outliers
Step 2 Select the order of (p, d, q) of an ARIMA (p, d, q) model
Step 3 Estimate the ARIMA coefficients
Step 4 Diagnose the residuals of the selected model
Step 5 Evaluate forecasting ability with out-of-sample
Step 6 Repeat Step 2–Step 5
Step 7 Choose the model with the best performance for forecasting.

Eventually, a relatively parsimonious ARIMA model with order (1, 0.3) has been found to be both optimal the in-sample fitting and out-of-sample forecasting performance. The satisfactory results concerns the statistical analysis of this classical time series modeling approach, as proven by its application to model compound air pollutant such as PM2.5. Limited by the data, we can only consider the nonseasonal ARIMA model even though some seasonal variation has been identified, which may have some impact on the model performance. The pitfall of ARIMA models have been points out by many researchers and we have discussed some of them. And the weakness of similar models is mainly concentrated on the forecasting part. Actually the prediction performance decreases dramatically if we increase the forecasting step even from one ahead to two. Moreover, it fails to catch the variation which is not caused internally. Consequently, external factors such as some the forecasted meteorological information could be taken into accounts to improve the prediction capability both in precision and time span.

References

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on, 19*(6), 716–723.

Boldo, E., Medina, S., Le Tertre, A., Hurley, F., Mücke, H.-G., Ballester, F., & Aguilera, I. (2006). Apheis: Health impact assessment of long-term exposure to PM2. 5 in 23 European cities. *European journal of epidemiology, 21*(6), 449–458.

Box, G. E., Hunter, J. S., & Hunter, W. G. (2005). *Statistics for experimenters: design, innovation, and discovery* (Vol. 2): Wiley Online Library.

Christian, S. (2001). Current Status of Air Quality Models for Particulate Matter. *Journal of the Air & Waste Management Association, 51*. doi:10.1080/10473289.2001.10464383

Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 190–195.

Ho, S., Xie, M., & Goh, T. (2002). A comparative study of neural network and Box-Jenkins ARIMA modeling in time series prediction. *Computers & Industrial Engineering, 42*(2), 371–375.

Meyler, A., Kenny, G., & Quinn, T. (1998). Forecasting Irish inflation using ARIMA models.

Pope III, C. A., Ezzati, M., & Dockery, D. W. (2009). Fine-particulate air pollution and life expectancy in the United States. *New England Journal of Medicine, 360*(4), 376–386.

Robeson, S., & Steyn, D. (1990). Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations. *Atmospheric Environment. Part B. Urban Atmosphere, 24*(2), 303–312.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics, 6*(2), 461–464.

Slini, T., Karatzas, K., & Moussiopoulos, N. (2002). Statistical analysis of environmental data as the basis of forecasting: an air quality application. *Science of the total environment, 288*(3), 227–237.

Tiao, G. C., Box, G. E. P., & Hamming, W. J. (1975). Analysis of Los Angeles Photochemical Smog Data: A Statistical Overview. *Journal of the Air Pollution Control Association, 25*. doi:10.1080/00022470.1975.10470082

Ujjwal, K., & Jain, V. K. (2009). ARIMA forecasting of ambient air pollutants (O3, NO, NO2 and CO). *Stochastic Environmental Research and Risk Assessment, 24*. doi:10.1007/s00477-009-0361-8

Authors Biography

Min Zhou received the B.S. degree in Automation from University of Science and Technology of China and now she is a Ph.D. student in Department of Industrial and Systems Engineering at National University of Singapore. Her interests include quality management, big data analytics, statistical process control, monitoring, and diagnosis for complex manufacturing process.

T.N. Goh is Professor of Industrial and Systems Engineering, specializing in Quality Engineering. He obtained his B.E. from the University of Saskatchewan and Ph.D. from the University of Wisconsin–Madison. He is an Academician of the International Academy for Quality and Fellow of the American Society for Quality.

Environmental Quality: A Case Study of Public Perceptions and Scientific Reality

Jing Li, Thong Ngee Goh and Min Zhou

Abstract It is widely acknowledged that environmental quality and protection should not be ignored in the development of a society. Environmental quality actually depends not only on government programs and policies but to a large extent awareness and actions from the public. However, the latter would then relate to human behavior which could be subjective, irrational, and impressionable as well. In recent years, the fast development of Internet technology has propelled the boom of social media and instant communication tools which enable the public to express their opinions on current events more conveniently and quickly. Nevertheless, the majority of swift comments tend to be emotional or one-off knee-jerk reaction which may fail to appreciate the actual impacts behind. With this background, the importance of critical and statistical thinking in the Web era has been illustrated in this paper. One case study, which relates a recent actual incident (discarding of rubbish) on Tiananmen Square in the capital of China, is used to underscore the need for rational opinions and the circumspect in spreading the message for enhancing environmental quality and handling related behavioral issues. It has been found that typical comments on the Internet in terms of this particular environmental issue reveal the usual emotional reactions and lack of statistical knowledge among the public. It is concluded that for advances in the importance quality of the environment, guidance of critical and statistical thinking is indispensable.

J. Li (✉)
Mechanical Engineering Department, National University of Singapore, Singapore, Singapore
e-mail: lijing8822@gmail.com

T.N. Goh · M. Zhou
Industrial and Systems Engineering Department, National University of Singapore, Singapore, Singapore
e-mail: isetngoh@nus.edu.sg

M. Zhou
e-mail: zhoum@nus.edu.sg

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_19

Keywords Quality of environment · Human behavior · Statistical thinking · Systems thinking

1 Introduction

It is common knowledge that after the Industrial Revolution, economic development accelerated in many areas till what is seen in the modern world. Together with this is the inevitable series of environmental issues, for instance, climate changes, wastes of resources, environment pollution, and so on.

With a view to protecting the environment, governments around the world have set up various forms of regulations to curb pollution, among other measures. Thus, for example, to restrict pollution from vehicles, strict emission regulations were put in place. However, it is clear that legal means alone would not achieve the purpose; the moral imperative should also be established. In fact many environmental issues are rooted in human behavior (Steg and Vlek 2009). Therefore, public awareness on environmental protection is essential; educational campaigns have been launched aiming at changing individual behavior, e.g., (Feng and Reisner 2011), where a survey was done to understand the factors that affect environmental protection behavior. Statistical analysis of the results showed that the effect of individually perceived environmental protection and environmental resource conservation is more important than public assertion of desired behavior. It is thus necessary that awareness of the populace should be of a higher priority. In fact in daily life, wastes are more related to public behavior, especially in terms of municipal solid waste.

As is well known, this is an era with massive data on the Internet. It is mentioned in the book The World is Flat that ten forces flattened our world (Friedman 2005). Among those forces, the prevalence of personal computers and Internet accessibility are milestones in the process of the world "becoming flat" as people could acquire information easily and conveniently in a way never before possible. They can forward and comment on news, blogs on a variety of subjects. Furthermore, people can build their own blogs to express and propagate their personal ideas. Facing such huge information sources, one not only needs to have good judgment on available massive information, but take the responsibility of filtering rumors or immature opinions. Thus both critical and statistical thinking are increasingly essential in facing issues related to environmental protection.

Critical thinking is one way to stay away from subjectivity when one faces a new situation. It requires people setting proper probing questions to evaluate a statement or a problem, etc., in an objective way. There are three basic activities involved in the process of critical thinking; they are investigation, interpretation, and judgment (Ruggiero 2011). As mentioned, people today are in a world full of massive information and can hardly not be influenced by herd opinions and mass culture. It is necessary that people acquire the ability to judge statements on Internet. Recently, one Chinese blogger, Qin Huohuo, was sentenced to 3 years in jail because of spreading rumors on Weibo, the most popular Chinese microblogging website (Bischoff 2014). In fact as long as readers could think in a critical way, rumors will be nipped in the bud. For example, Qin alleged that Lei Feng, a moral model known as a poor but selfless soldier in China, was actually quite wealthy. In critical thinking, one who read the statement should pose the following questions:

1. Is this true?
2. How does Qin know this? Has he obtained enough evidence?

Once the readers slightly think about those questions, they may not simply spread the unveracious posting.

If critical thinking is more like a qualitative way to evaluate a statement, statistical thinking could be considered as a quantitative counterpart that examines the statement by some objective and quantitative measures. Statistical thinking could be explained as a manner to examine facts preferably by objective quantitative measures rather than simple talks. Though statistics has been included in the curricula of many schools, the emphasis is often on the mechanics of number crunching. And it is still a long way for citizens in society to use benefit from the statistical thinking in face of real problems. Goh illustrated how to think statistically when one faces with a statement involving data (Goh 1986). Number as the medium to convey information, needs to be expressed in a precise way hence avoiding unclear meaning.

In this paper, a case related to environmental issues caused by human behavior that comes from the most populous nation on earth is studied to show the importance of critical and statistical thinking.

2 Case Study on Litter Issue

Given the speed of economic development in China, environmental issues are numerous. At the level of daily life, littering is one environmental issue that is closely related to human behavior and it certainly lowers the quality of life in general. No doubt the educated men and women would be aware of the consequence of littering; however littering seems impossible to eliminate. Some people are unconcerned about the morality of it, and at the same time most people (more than 60%) feel shy to dissuade rubbish littering (Yan et al. 2012). The case detailed here is to demonstrate that with popular awareness, what roles critical and statistical thinking might play in a high-profile incident.

2.1 Show of Patriotism

On October 1, 2013, the 64th anniversary of the founding of the People's Republic of China, a traditional ceremony, namely national flag raising, was held early in the morning at Tiananmen Square, acknowledged to be the heart of the nation. An estimated crowd of more than 110,000 from all parts of the country gathered to witness the ceremony. After the high-profile ceremony was over, it was reported that more than 5 tons of rubbish was left at the square (Ying 2013), evoking a vigorous nation-wide debate on the “Quality of our Countrymen”.

2.2 *Hot Debate Arouse on the Internet*

Debates were evoked on the Internet. In this section, all the points of view put forward are listed here for further discuss.

2.2.1 The Human Behavior Point of View

The most typical is the eye-catching headline: “Tiananmen Square: 5 Tons of Rubbish! At worst spots, even the ground could not be seen!” With such highlighting of the amount of litter strewn in an open public place—and the nation’s heart at that—criticisms of the low standards of ethics among the Chinese people began to emerge, while some blamed the media for exaggerations and misleading descriptions. Table 1 lists these two points of view.

2.2.2 The Statistical Point of View

Average

It would be useful to examine what it really means when 5 tons of rubbish is left by 110,000 people, assuming that these are good estimates. As can be seen from Table 2, the figures translate to only 45.45 g of rubbish per person. This adds substance to the accusation that the media did exaggerate the littering behavior of the public.

Table 1 Two points of view

	Blame the people (Pan 2013)	Blame the media
1	Littering about reflects the poor public spirit of Chinese countrymen	Rubbish had to be put down on the ground as there is no litter bin at Tiananmen Square (Anonymous 2013c)
2	Littering about is an unpatriotic behavior	Littering about is unrelated to the love for the country; however attending the flag-raising ceremony is patriotic behavior (Anonymous 2013b)

Table 2 Calculation of rubbish from each person

Number of people	Weight of rubbish (g)	Rubbish (g/person)
110,000	5,000,000	45.45

Comparison

In addition to calculating the per capita figures, the anti-media individuals went further with comparisons of extents of littering with situations abroad. They pointed out that 5 tons of rubbish is insignificant when contrasted to 20 tons in 2 days in the much smaller Tsimshatsui Square in Hong Kong during a Christmas celebration event, or 40 tons during a New Year countdown night at the New York City Times Square (Anonymous 2013a).

Another way of comparing was also put forward that contrasted this year's amount of rubbish with those in the past years on the same occasion. Based on the figures in Table 3, some argued that "Instead of condemning the 5 tons of rubbish, this is precisely a figure that reflects a tremendous improvement: the amount last year was 8 tons, and during the 2005 'golden week', the daily rubbish was as much as 19 tons. One can see that the amount of littering has gone down tremendously, and the public spirit of tourists on the other hand has been growing!"—In other words, the facts point to a steady improvement of public ethics by the year which of course is a positive sign rather than something to be criticized.

2.3 Critical Thinking

Table 1 lists the idea which is about the defense for the people who litter. If one thinks critically, some questions could be raised:

1. What is littering?
2. Is littering a virtue or vice?
3. Is the "put down" action out of the littering scope?
4. Why do people "put down" the rubbish on the ground?

Following are the corresponding answers:

Table 3 Weight of rubbish in recent years (Anonymous 2013a)

Date of report	Rubbish
2005.05.07	19 t/day
2006.10.08	222 t/week
2007.05.07	146 t/week
2009.10.08	220 t/week
2010.10.01	20 t/day
2012.10.02	8 t/day
2013.10.01	5 t (after the flag-raising ceremony)

1. It is explained that litters are things that have been thrown away and that are lying on the ground in a public place (Merriam-Webster 2014).
2. There is no doubt that it is undesirable to litter.
3. It is stretching too far to say that because rubbish bins could not be found, 5 tons of rubbish was “put down” on the ground and not “thrown away” in an act of littering. If such behavior is an acceptable one, then anyone could “put down” rubbish anywhere and anytime when there is no rubbish bin in sight. To spread a message like this is inviting wanton littering not with a tinge of shame, but with full justification and dignity.
4. This leads to the question why the crowd “put down” rubbish the way they did. It has been claimed that security considerations accounted for the absence of rubbish bins on Tiananmen Square. However, in a crowded event, it was ruled that one should not go around freely and if there was rubbish, it should be disposed off on the spot. Such “rules” simply forced people to “put down” rubbish on the ground! As a matter of fact, this is a perennial problem: since 2005 (or earlier), the Chinese media have been harping on the tonnage of rubbish in public places. However, the authority seemed oblivious to this and, in the case of flag ceremony, it could simply provide rubbish disposing bags to the public, which has a dual purpose of reducing the workload of cleaners called in subsequently to clear the place.

2.4 Statistical Thinking

2.4.1 Rubbish Generation and Removal

It has been pointed out that 150 cleaners were able to remove 5 tons of rubbish within 30 min, pointing to a very high efficiency. Table 4 lists the speed of cleaning, from which it can be deduced that every cleaner must perform at the rate of 1.63 m^2 per second, thus guaranteeing a pleasant environment for China’s leaders to offer their flowers at the People’s Martyrs Memorial (Huo and Wu 2013). Actually from another point of view, it is hard work for anyone cleaning at this speed for 30 min nonstop; Table 5 contrasts the per capita amounts of rubbish generation and removal. It can be seen from the table that the 45.45 g of rubbish from each person actually led to 733 times (33,333/45.45) more work for the cleaners. In that sense, if a person has a heart for the cleaners, he would consciously take away his rubbish instead of leaving it behind to the cleaners.

Table 4 Speed of removal of rubbish

Area of square (m^2)	Time to clean up (s)	Workers (person)	Speed (m^2/s/person)
440,000	1800	150	1.63

Table 5 Rubbish distributed by workers and citizens

	Workers	Citizens
Total rubbish in gram	5,000,000	
Number	150	110,000
Average per person (g)	33,333.33	45.45

2.4.2 Categories of Rubbish

On first look 45 g is indeed a small amount. However, from the photographs in news reports (Fig. 1), it can be seen that most of the rubbish comprises discarded newspapers, plastic bags, empty water bottles, and so on. Table 6 exhibits some typical indicative numbers.

In fact 45 g can be made up of the following:

(a) A sitting person takes up an area of $0.5 \times 0.3 = 0.15$ m^2. One piece of A4 paper has an area of approximately $0.210 \times 0.297 = 0.062$ m^2, i.e., $2.4 = 0.15/0.062$ sheets of such paper will satisfy the sitting needs.
(b) On the average one bottle of water is consumed during the waiting period prior to flag raising.
(c) Before their removal, the paper and the bottle of drinks could generate about 33 g of rubbish; the remainder of 12 g of rubbish may arise from the plastic bags or food packaging, and possibly rain water as well.

2.4.3 Traps of Statistics

As mentioned before, some commentators took it as an improvement of the fact that the amount of rubbish generated seemed decreasing, as shown in Table 3, but they have ignored the fact that the 5 tons of rubbish in 2013 appeared after the flag-raising ceremony, while the information for previous years refers to daily figures: in other words, there is no common basis of comparison.

Fig. 1 Rubbish on the ground

Table 6 What comprise the 45 g rubbish?

Category	45.45 g equals	Unit weight (g)
Paper	9 pieces	5 (A4 size)[a]
Empty bottles	2	20.67[b]
Plastic bags	10	4.5[c]

[a]Available at http://en.wikipedia.org/wiki/Paper_density. It states that typical office paper has 80 g/m^2, therefore a typical A4 sheet (1/16 m^2) weight 5 g

[b]Available at http://sports.china.com.cn/2010yayun/2010-10/10/content_21093290.htm. It states that a bottle for tea or juice weighs 22–25 g, and for purified water it weighs 15 g. Here the average value of 22, 25, and 15 is taken, that is 20.67 = (22 + 25 + 15)/3

[c]Available at http://www.plasticbagfacts.org/Main-Menu/Fast-Facts/. It states that a typical plastic bag weighs 4–5 g. Here, we take the weight for each plastic bag as 4.5 g, which is got from (4 + 5)/2

In the book How to Lie with Statistics, it is illustrated how samples in reports tend to have built-in bias. (Huff 1954). It appears that most people would choose data favorable to their own viewpoints when arguing a case—perhaps not a surprising situation but nonetheless a habit not to be condoned. It is seen that in Table 3, no data from 2011 was mentioned. On investigation as to what happened in 2011, it was found that on Chinese National Day of 2011, 120 thousand people came to the flag-raising ceremony, resulting in 0.94 tons of rubbish (Wang and Wang 2011). The reports in 2011 and 2013 both use the rubbish volumes after the ceremony; hence the comparison is more meaningful, suggesting that ordinary folks are capable of not "putting down" rubbish anywhere, possibly just taking them away. It can be noted that those not in favor of such an observation would claim data incompatibility or ignore unfavorable data when expressing their disagreements; there are even reports that used inappropriate data from Table 3 for comparisons, claiming objectivity in their conclusion that there is a general improvement of quality of behavior of the people (Zhu 2013).

3 Concluding Remarks

One can see how a single straightforward incidence could arouse public reactions that range from emotional responses to making use of data that is favorable to some preconceived or desirable conclusions—especially in environmental matters where there seem to be "morally right" attitudes, believes, and answers. The fact is, for a given issue, most of the time the complete information in the desired format is unavailable. It makes more sense to tread carefully, expressing objective albeit inconclusive views or pointing out various possibilities, even perhaps reserve judgment. To create healthy and rumors-free publicity, both critical and statistical thinking are necessary qualities that all citizens need to acquire. In this sense, more generally, one could not but remember what H G Wells admonished almost a century ago, "Statistical thinking will one day be as necessary for efficient citizenship as the ability to read or write." (Huff 1954).

References

Anonymous. (2013a): *14 tons less rubbish than that in the year of 2005 during golden week*. Accessed October 9, 2013. http://news.sina.com.cn/c/2013-10-02/104628350755.shtml.

Anonymous. (2013b): *Heavier rubbish than that in tiananmen square*. Accessed October 9, 2013. http://blog.sina.com.cn/s/blog_9f679c850101fdmg.html.

Anonymous. (2013c): *The truth hiden behind the media*. Accessed October 9, 2013. http://blog.sina.com.cn/s/blog_613d53180101b4rd.html.

Bischoff, P. (2014): *Yet another chinese blogger jailed for allegedly spreading rumors on weibo*. Accessed May 6, 2014. http://www.techinasia.com/chinese-blogger-jailed-allegedly-spreading-rumors-weibo/.

Feng, W. and Reisner, A. (2011): *Factors influencing private and public environmental protection behaviors: Results from a survey of residents in shaanxi, china*. Journal Of Environmental Management 92 (3): 429–436.

Friedman, T.L. (2005). *The world is flat: A brief history of the twenty-first century*. Farrar, Straus and Giroux, New York.

Goh, T.N. (1986): "*Knowledge of a very meagre and unsatisfactory kind*", The Industrial Engineer: Yearbook: pp 84–85.

Huff, D. (1954). *How to lie with statistics*. WW Norton & Company, New York.

Huo, X.G. and Wu, J. (2013): *The community party leaders presented flowers to the people's heroes monument*. Accessed October 10, 2013. http://news.xinhuanet.com/politics/2013-10/01/c_117583078.htm.

Merriam-Webster. (2014): *Merriam-webster's collegiate dictionary*. Accessed January 15, 2014. http://www.merriam-webster.com/dictionary/littering.

Pan, H.Q. (2013): *5 tons of rubbish on tiananmen square is a mirror.* Accessed October 9, 2013. http://news.xinhuanet.com/local/2013-10/02/c_117583962.htm.

Ruggiero, V.R. (2011). *Beyond feelings: A guide to critical thinking*. 9th ed. McGraw-Hill, New York.

Steg, L. and Vlek, C. (2009): *Encouraging pro-environmental behaviour: An integrative review and research agenda*. Journal of Environmental Psychology 29 (3): 309–317.

Wang, P. and Wang, D.P. (2011): *120 thousand tourists witness national flag-raising ceremony on the 1st october*. Accessed October 10, 2013. http://news.ifeng.com/mainland/special/2011guoqing/content-4/detail_2011_10/02/9621575_0.shtml.

Yan, G., Kang, J., Wang, G., Lin, H., Zhu, J., Liu, C., Sun, W., Li, Y. and Jin, T. (2012): *Change trend of public environmental awareness in shanghai (2007 to 2011)*. Energy Procedia 16, Part B (0): 715–721.

Ying, Z. (2013): *5 tons of rubbish left after national flag-rasing ceremony by 100 thousand tourists*. Accessed October 9, 2013. http://news.163.com/13/1001/09/9A3F2A640001124J.html#p=9A3NETCA00AN0001.

Zhu, D.Q. (2013): *Understandable square "rubbish" and unreasonable "rubbish men" mentality*. Accessed October 9, 2013. http://sd.dzwww.com/sdnews/201310/t20131002_8956822.htm.

Part VIII
Reliability Practices

Prof. Cher Ming Tan, Chang Gung University, Taiwan

As mentioned in Chapter 2, reliability activities in Asia is no longer limit to qualification and pass/fail testing. Detail reliability analyses are being done in Asia, due possibly to the increasing trend of shifting the R&D to Asia in the past decade. However, detail reliability analysis requires complex statistics, and one can see that almost all the work done on reliability analyses are limited to Universities. Although some reliability analyses are reported by companies in ANQ2014, their methods are not rigorously correct if they did it by themselves. For those companies where their methods are accurate, they are done in collaboration with Universities.

While all the detail reliability analyses are done with or by Universities, the requirements for real field data in the analyses are noted. Interestingly, the reported reliability analyses are mostly related to the parts in automotive. This is because of the high reliability requirements of automotive industry.

Besides component reliability in automotive, system reliability is also addressed in terms of maintenance and the awareness of system design that may compensate components degradation, rendering direct applications of components reliability to predict system reliability inapplicable.

Author biograph can be found in Chapter 2.

Quality Decision for Overcharged Li-Ion Battery from Reliability and Safety Perspective

Feng Leng, Cher Ming Tan, Rachid Yazami, Kenza Maher and Robert Wang

Abstract During charging of Lithium-ion battery (LiB), the charging cutoff voltage (COV) may exceed the manufacturers' specification because of incorrect monitoring of the charging control circuit, either due to the aging of the control circuit or the design/manufacturing errors of the control circuit. In fact, it is found that overcharging LiB cell is a common abuse. This work shows the effect of excessive COV on cell's discharging ability, and the use of a novel nondestructive method to evaluate if the damage made in the cell by the excessive COV is rendering the cell from further safe usage or it is still acceptable with minor degradation in reliability and safety, thus providing a basis for quality consideration of the cell. The method also enables battery manufacturers to identify the internal components for their cells that are most vulnerable to the excessive COV so that quality improvement of their batteries can be designed and produced. This method also alerts electric vehicles user on the hidden safety issues of their battery pack,

F. Leng · R. Wang
School of Electrical Electronics Engineering, Nanyang Technological University, Blk S2.1, 50 Nanyang Avenue, Singapore 639798, Singapore
e-mail: fleng001@e.ntu.edu.sg

R. Wang
e-mail: wangqingyun123@yahoo.com

F. Leng · R. Yazami
TUM CREATE PTE LTD, 1 Create Way, #10-02 Create Tower, Singapore 138602, Singapore
e-mail: YAZAMIrachid@ntu.edu.sg

C.M. Tan (✉)
Department of Electronic Engineering, Chang Gung University, No. 259, Wenhua 1st Road, Guishan District, Taoyuan City 333, Taiwan
e-mail: cherming@ieee.org

R. Yazami · K. Maher
School of Materials Science and Engineering and Research Institute at Nanyang (ERIAN), Nanyang Technological University, Research Techno Plaza, X-Frontier Blk, 50 Nanyang Drive, Singapore 637553, Singapore
e-mail: MaherKMaher@ntu.edu.sg

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_20

and enables battery management system to perform reliability balancing, a new patented technique to ensure the safe and reliable operation of battery pack.

Keywords Lithium-ion battery · Electrochemistry-based electrical model · Overcharge · Battery safety · Reliability balancing

1 Introduction

Rechargeable Lithium-ion battery (LiB) has been widely used in mobile phones, and notebook computers due to its high energy densities, high power densities, long cycle life, and safe operation. It is also an enabler that makes pure electric and hybrid electric vehicles a reality (Ye et al. 2012).

However, several abusive conditions can happen to a LiB cell during its operation, and overcharge is considered as a common abuse that may lead to its severe degradation and even catastrophic failure via thermal runaway and subsequent fire hazard with possible explosion in extreme cases as reported (Tobishima and Yamaki 1999). This becomes a primary concern for battery users, and hence a nondestructive method is required to access the damaging effect of the overcharging in order to decide if one should change the cell from the reliability and safety perspective. From battery manufacturers point of view, if one can identify the internal components in a cell that are most vulnerable to excessive cutoff voltage COV, they can improve their cells so as to make the cell more robust against this abuse. Unfortunately, such a method is not available in the market.

With the recent development of electrochemistry-based electrical (ECBE) model (Leng et al. 2014), the degradation of each individual components inside a cell can be determined through its discharging curve (i.e., terminal voltage versus time during discharging) alone. In this work, we use the ECBE method to examine the effect of excessive COV on the degradation of each component experimentally.

Figure 1a shows the ECBE model and the physical meaning of each components as shown in Fig. 1b where R_e is the ohmic resistance of electrodes, C_{dl} is the double layer capacitance, R_{ct} is the total charge transfer resistances at the electrodes, K is the charge transfer rate constant at the electrode embedded in the Butler–Volmer impedance (Z_{BV}), and R_n and C_w are the Warburg element in the electrolyte.

2 Experiments

In this work, the discharging curves of LiB cells are obtained experimentally. LiB coin cells (LIR2032) rated $\sim$44 mAh as shown in Fig. 2 are used for experiments. The electrodes materials are graphite anode and lithium cobalt oxide (LCO) cathode, and it is the most commonly used cathode material in commercial LiB (Park et al. 2010).

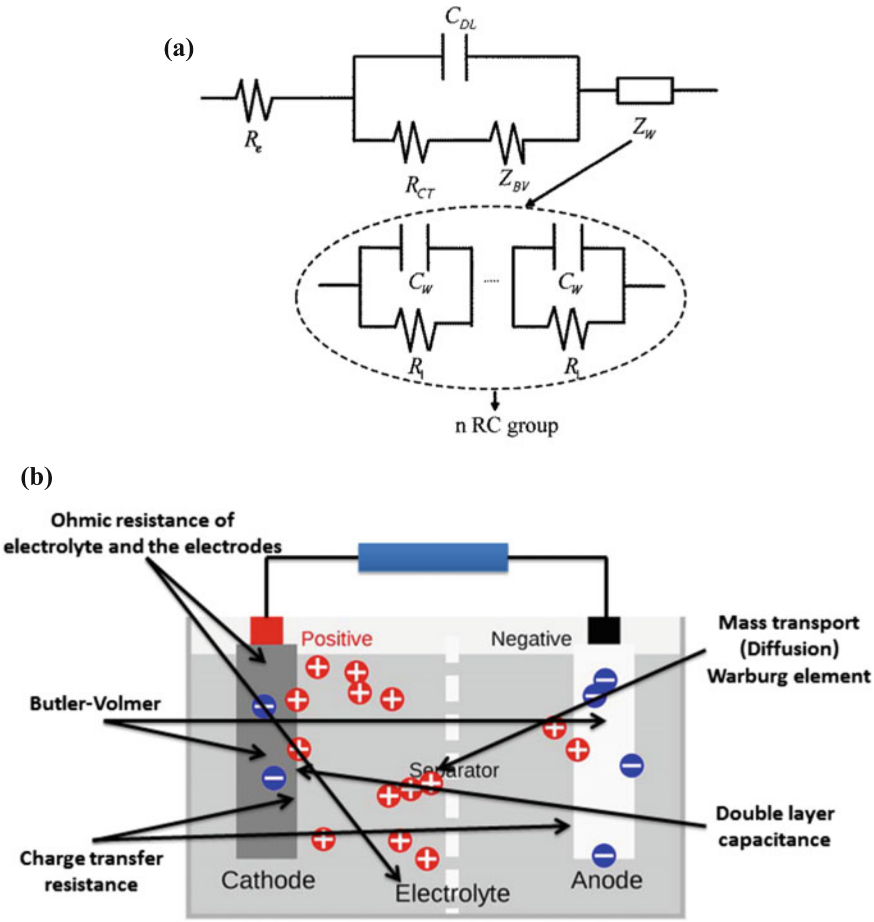


Fig. 1 **a** Electrochemistry-based electrical (ECBE) model (Leng et al. 2014). **b** Real physical parts of the inner battery represented by ECBE model parameters

Arbin Instruments battery cycler is used for the overcharging experiments, and the set up for experiments in this work is programmed as follows (Maher and Yazami 2013):

- All fresh battery cells are firstly discharged to a cutoff voltage of 2.75 V
- They are then charged abusively at C/4-rate (i.e. 11 mA) to COV between 4.2 and 4.9 V (specification of max COV is 4.2)
- They are rested for 160 s after charging for the their overvoltage stabilization
- For the discharging experiments, they were discharged at a constant current of C/4 to a COV of 2.75 V at constant room temperature of 23 °C unless otherwise stated.

Fig. 2 The coin cell (LIR2032) used in this work

- The terminal voltages and charging/discharging currents of the entire tested cells are continuously monitored and recorded at every 20 s interval.

3 Results and Analysis

Figure 3 shows the discharging curves of the LiB cells after excessive COV. From the discharging curves, we can see that the higher the COV, the longer the discharging time to reach 2.75 V, expect when the COV is above 4.7 V. The rates of change of the terminal voltages for all cases are almost the same. Thus, this may indicated a benefit of having higher COV as cell can be used for longer time per charging.

However, it is reported that when cell is overcharged at excessive COV, it is indeed stressed, and its life will be reduced, and both the cell reliability and safety are compromised (Buchmann 2001). As LiB cells use volatile and flammable organic electrolytes, irreversible decomposition of the electrolyte may occur and this will trigger the gas evolution reaction at high COV, and leads the continuously swell and rupture of the cells eventually. The overcharged cell behaves as a resistor at high voltages, dissipating excess energy as heat. The locally accumulated heat due to increase of cell resistance as a result of degradation at electrodes, leading to the catastrophic failure such as fire hazard and possible explosion of battery by thermal runaway at extreme conditions (Maher and Yazami 2013; Li et al. 2011; Lin et al. 1865; Ohsaki et al. 2005). These accidents resulted from the extreme COV as shown in Fig. 4. This is attributed to the following degradation of the internal components in the LiB cells as revealed by our ECBE method as follows.

Figure 5 shows that the maximum initial capacity of a cell increase with COV, and this explains the observed "enhanced" runtime of the discharging curves with higher COV as observed in Fig. 3. This increase is a result of more lithium ions are extracted from electrode due to overcharging.

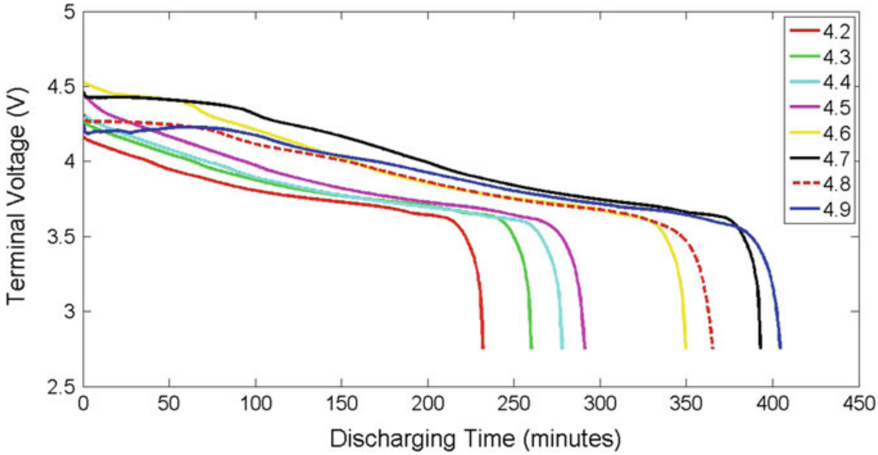


Fig. 3 Discharging curves of LIR2032 cells subjected to different charge cutoff voltages (COV)

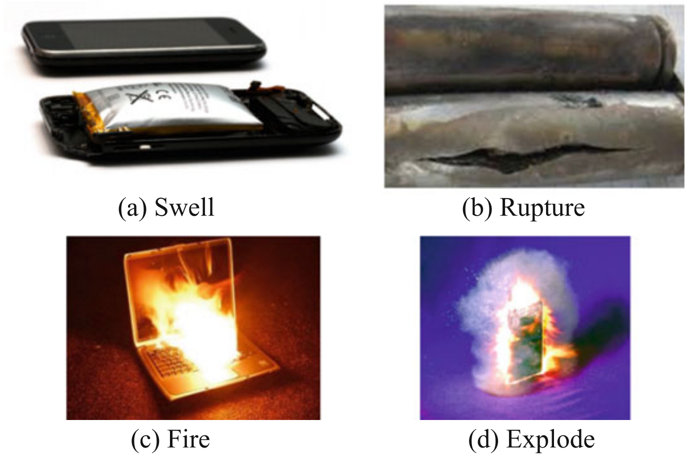


Fig. 4 The accidents caused by extreme overcharging abuse

The m_1 and m_2 represent efficiencies of anode and cathode in providing their stored Li-ion for discharging, respectively. The decrease in m_1 with excessive COV can be seen in Fig. 6, and this implies that anode electrode is degraded when COV is excessive. However, the performance of the graphite electrode is insignificant against excessive COV as seen in Fig. 7. Here we have to take note that graphite electrode is anode during charging, and as our results are extracted from discharging curve, the anode mentioned from the extraction using ECBE is referred to the LCO electrode.

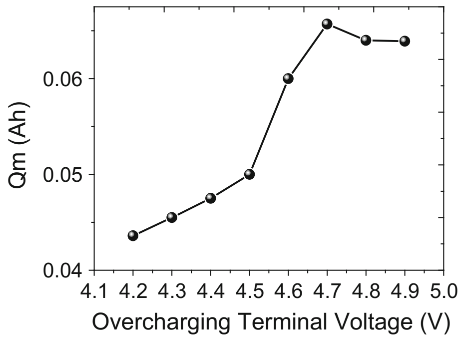


Fig. 5 Maximum initial capacity versus excessive COV

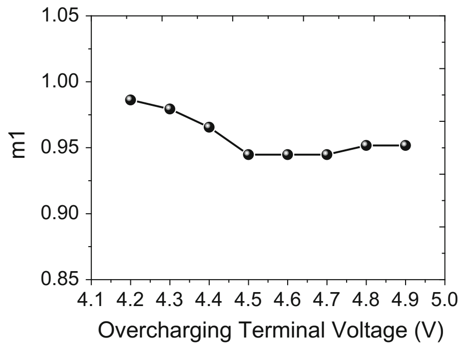


Fig. 6 M_1 of cobalt-oxide electrode (*anode*) versus excessive COV

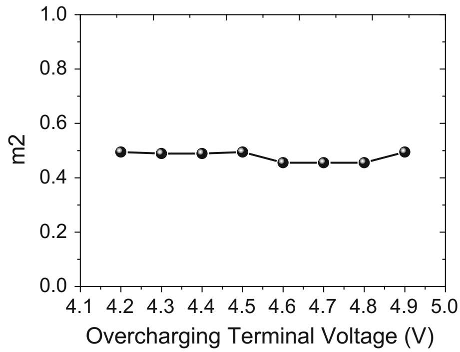


Fig. 7 M_2 of graphite electrode (*cathode*) versus excessive COV

The sum of ohmic resistance of electrode and charge transfer resistance decreases with COV and then increases for COV above 4.3 as shown in Fig. 8. While the chemical kinetic explanation of the phenomena is beyond the scope of this work, the increase of $R_e + R_{ct}$ increases the localized heating of the anode due to Joule heating. Thus the localized temperature of anode is expected to be higher for higher COV above 4.3 V, and this could explain the degradation observed in anode as in Fig. 6. This localized high temperature at anode could also be one of the

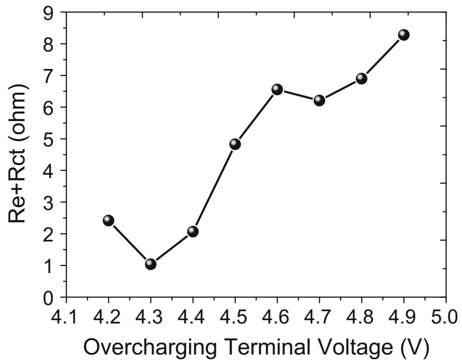


Fig. 8 The sum of ohmic resistance of electrode and charge transfer resistance versus excessive COV

reasons for higher Q_m as the chemical reaction rate is going to be higher at higher temperature according to the Arrhenius relation.

As COV is further increased to beyond 4.5 V, the localized high temperature can become so high that dissociation of the electrolyte next to electrode occurs, and the formation of solid electrolyte interface (SEI) is enhanced (this is also reflected as an increase in R_{ct}). The much higher localized temperature at the electrode is also evidenced by the higher rate of increase in Q_m with respect to COV when COV is above 4.5 V. With formation of thicker SEI, the charge transfer at the electrode will be retarded, and hence one can see a sharp decrease in K values as shown in Fig. 9 when COV is above 4.5 V. This thick SEI also tends to protect the anode from the electrolyte and thus its degradation stop when COV is above 4.5 V as shown in Fig. 6. However, with thicker SEI, most part of the charging voltage will be dropped across the SEI, and the electric field in the SEI can become too high at COV above 4.5 V that causes it to breakdown, trigger the full cell thermal runaway (Belov and Yang 2008). It is to be noted that the above degradation happen with only one excessive COV charging. One could expect that with continuous excessive COV charging, the electrode resistance will increase continuously, and the localized temperature at anode will continue to increase. While we enjoy the seemingly longer runtime with higher COV, the cell reliability is degrading, and at some point in time, the high localized temperature can trigger some safety issue. In

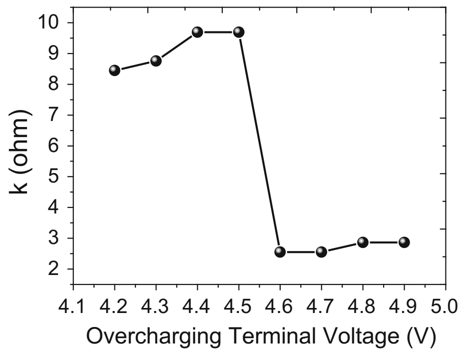


Fig. 9 Charge transfer rate constant versus excessive COV

a mild case, we will see a swelling of the Li-ion cell in our hand phones due to larger thermal expansion, but in a severe case, fire and explosion can occur as have been observed in electric vehicles incident (Lamb et al. 2014).

As our method can provide real-time online nondestructive assessment of the internal components in a Li-Ion cell, battery manufacturers can identify the weakest components that are most vulnerable to the excessive COV. This method can also be incorporated into battery management system (BMS) and once a cell is in an unsafe situation as defined by the $R_e + R_{ct}$, a bypass of the cell can be initiated according to the new technology of reliability balancing developed recently (Tan and Tan 2013). Thus, our method enables both battery manufacturers and BMS producers to improve their products quality over time and safety.

4 Conclusion

4.1 Achievement

We have successfully identified the seeming gain of Li-Ion cell by having it to charge with high COV, and revealed the hidden issues associated with charging the cell with high COV, with the use of recently developed ECBE battery model. Therefore, the misconception of higher COV to enhance the runtime is clarified. With our method, we are able to provide real-time online nondestructive assessment of the internal components in a Li-Ion cell to evaluate its reliability and safety with respect to this overcharging abuse. We are able to characterize the degradation level of each internal electrochemical component in LiB and consequently identify the weakest material that is most vulnerable to the excessive COV. This will be beneficial for battery manufacturers to design and produce more robust cells against overcharging. Our method can also be incorporated into existing battery management system so that a more advanced safety technology to ensure safe operation of the battery pack with enhanced system reliability can be possible.

Acknowledgments The authors a would like to acknowledge the support from Nanyang Technological University, School of Materials Science and Engineering and Research Institute at Nanyang (ERIAN) in providing testing facilities.

References

Y. Ye, Y. Shi, N. Cai, J. Lee, and X. He, "Electro-thermal modeling and experimental validation for lithium ion battery," *Journal of Power Sources,* vol. 199, pp. 227–238, 2/1/ 2012.

S.-i. Tobishima and J.-i. Yamaki, "A consideration of lithium cell safety," *Journal of Power Sources,* vol. 81–82, pp. 882–886, 9// 1999.

F. Leng, C. M. Tan, R. Yazami, and M. D. Le, "A practical framework of electrical based online state-of-charge estimation of lithium ion batteries," *Journal of Power Sources,* vol. 255, pp. 423–430, 6/1/ 2014.

Y. Park, N. H. Kim, S. B. Cho, J. M. Kim, G. C. Kim, M. S. Kim, *et al.*, "Characterization of the passivating layer on Li[Ni0.31Co0.32Mn0.28Al0.09]O2 cathode in the overcharge state," *Journal of Molecular Structure,* vol. 974, pp. 139–143, 6/16/ 2010.

K. Maher and R. Yazami, "Effect of overcharge on entropy and enthalpy of lithium-ion batteries," *Electrochimica Acta,* vol. 101, pp. 71–78, 7/1/ 2013.

I. Buchmann, *Batteries in a portable world*: Cadex Electronics Richmond, 2001.

S. L. Li, L. Xia, H. Y. Zhang, X. P. Ai, H. X. Yang, and Y. L. Cao, "A poly(3-decyl thiophene)-modified separator with self-actuating overcharge protection mechanism for LiFePO4-based lithium ion battery," *Journal of Power Sources,* vol. 196, pp. 7021–7024, 8/15/ 2011.

C.-K. Lin, Y. Ren, K. Amine, Y. Qin, and Z. Chen, "In situ high-energy X-ray diffraction to study overcharge abuse of 18650-size lithium-ion battery," *Journal of Power Sources,* vol. 230, pp. 32–37, 5/15/ 2013.

T. Ohsaki, T. Kishi, T. Kuboki, N. Takami, N. Shimura, Y. Sato, *et al.*, "Overcharge reaction of lithium-ion batteries," *Journal of Power Sources,* vol. 146, pp. 97–100, 8/26/ 2005.

D. Belov and M.-H. Yang, "Investigation of the kinetic mechanism in overcharge process for Li-ion battery," *Solid State Ionics,* vol. 179, pp. 1816–1821, 9/30/ 2008.

J. Lamb, C. J. Orendorff, K. Amine, G. Krumdick, Z. Zhang, L. Zhang, *et al.*, "Thermal and overcharge abuse analysis of a redox shuttle for overcharge protection of LiFePO4," *Journal of Power Sources,* vol. 247, pp. 1011–1017, 2/1/ 2014.

C. M. Tan, Y. Q. Tan, "Battery Reliability Balancing," presented in International Symposium of next generation electronics, 2013.

Authors Biography

Mr. Feng Leng obtained his Ph.D degree from Nanyang Technological University (NTU) of Singapore. He is also Executive Director of Innovation and Reliability Solution Pte. Ltd. He graduated from NTU and worked in various companies like Seagate International and UMC in the reliability field. His interest is mainly in Battery safety and reliability.

Prof. Cher Ming Tan is the Director of Centre for Reliability of Sciences and Technologies (CReST) in Chang Gung University, Taiwan and also a Professor in the Electronic department under the college of Engineering. He is also a Chair Professor in Ming Chi University of Technology, Taiwan. He was a faculty member of the Nanyang Technological University, Singapore for more than 19 years before he joined the current position. He started his reliability work in 1984 with Fairchild Semiconductor where the reliability and failure analysis works were needed for the US Defence. He later pursued his master and Ph.D. in University of Toronto, Canada. After graduation, he joined Lite-On Power Semiconductor, Taiwan, continued his work on quality, reliability and failure analysis. He is still active in providing training and consultation to companies that need reliability worldwide. He is an Editor of IEEE Transaction on Device and Materials Reliability, Editor of Scientific Report, Chief Editor of Springer Brief on Reliability, Advisory board member of Microelectronic Reliability, and IEEE Distinguished lecturer of EDS in Reliability. He was the course coordinator of certified reliability program of Singapore Quality Institute.

Prof. Rachid Yazami is a Professor in the School of Materials Science and Engineering at the Nanyang Technological University, Singapore, on leave from the French National Center for Scientific Research (CNRS) where he has held a Research Director position since 1985. Prof. Yazami spent 10 years at the California Institute of Technology (Caltech) where he cofounded a CNRS-Caltech joint laboratory on materials science for energy storage applications, including in lithium-ion batteries and in hydrogen storage. Prof. Yazami is the author of over 200 published scientific papers, book chapters and reports. He is the inventor involved in over 50 patents related to battery technology. In 1980 while starting his Ph.D. research at the Grenoble National Polytechnic Institute (INPG), Prof. Yazami invented the graphite anode (negative electrode) used in the 3-billion lithium-ion batteries produced in 2010. This discovery was the turning point in rechargeable lithium battery technology. In the mid-80s Prof. Yazami contributed to a new graphite fluoride cathode invention, which after further research led him to start a company (Contour Energy Systems, Inc.) in Azusa, California. The company commercializes high performance primary lithium batteries and carries out R&D on rechargeable fluoride ion batteries discovered by Prof. Yazami. In addition to materials science research, Prof. Yazami's interest focused recently on battery safety and life extension. He developed a technology based on thermodynamics measurements being used to accurately assess the state of health and the state of charge of batteries. This technology is licensed by Caltech to a new start up company founded by Prof. Yazami in Singapore, KVI LTP, LTD and addresses mobile electronics and electric vehicles markets for the most.

Dr. Kenza Maher is Research Fellow and Project Leader at Joint Future Mobility Research Lab with BMW Group and Nanyang University, and TUM CREATE Center for Electromobility, Singapore. Dr. Kenza Maher received her Ph.D. from Cadi Ayyad University (Marrakech, Morocco) in January 2011. Her Ph.D. focused on development of new phosphates for high energy lithium-ion batteries. After her Ph.D., she joined the Energy Research Institute @ Nanyang Technological University (ERI@N, Singapore) as Research Fellow for 1 year and 4 months; she worked on the development of a new method to analyze the commercial Li-ion batteries to determine their state of health (SOH) based on the entropy and enthalpy profiles. Her research focuses on Electrochemistry, Thermodynamics and Safety of Li-ion Batteries using Electrochemical Thermodynamic Measurement System (ETMS), Electrochemical Impedance Spectroscopy (EIS) and Accelerating rate calorimetry (ARC). She is working on determination of Li-ion batteries SOC, SOH and SOS.

Mr. Robert Wang is software engineer ShowBox Pte. Ltd and CTO of Innovation and Reliability Solution Pte. Ltd. He graduated from NTU with M. Science Degree. His interest is mainly in web development, mobile apps development, computational methods, and large server operation.

Statistical Inference for Grouped Field Failure Data

Piao Chen and Zhi-Sheng Ye

Abstract The field failure data is important in the reliability field. Component failure data, instead of system failure data, are usually collected by the system owners. However, in many cases the individual component time-to-failure data are not recorded. Instead, cumulative operation time and the number of failure times are recorded. This can be explained by the fact that exponential distribution is often used to fit the data. Unfortunately, the assumption of exponential distribution is not valid in many cases. In our paper, Gamma distribution and Inverse Gaussian (IG) distribution are used to fit the data. The point estimate for parameters of these two distributions can be obtained easily by general computer software. We then proposed methods to obtain the confidence interval (CI) for parameters of gamma distribution and IG distribution when the data are merged. In particular, for the rate parameter and mean of a gamma distribution, no efficient methods have been found to get the interval estimation. We borrow the idea of General Pivotal Quantity to obtain the interval estimation for the rate parameter and mean of a gamma distribution. The simulation study shows our method outperforms Wald method in terms of coverage probability. At last, an illustrative example is given to demonstrate the applicability of proposed models.

Keywords Generalized pivotal quantity · Grouped data · Inverse Gaussian distribution · Gamma distribution · Confidence interval

P. Chen (✉) · Z.-S. Ye
Department of Industrial and Systems Engineering (ISE),
National University of Singapore (NUS), Singapore, Singapore
e-mail: cp@nus.edu.sg

Z.-S. Ye
e-mail: yez@nus.edu.sg

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_21

1 Introduction

Observed failure data from fielded systems are valuable as they contain useful information about product quality and reliability. These data can provide insight into the design and maintenance of reliable systems. Therefore, it is extremely important to evaluate the reliability of such systems based on the observed data. However, a complex system may have a lot of components and it is costly to test such a system. As an alternative way, component reliability is often evaluated as a premise of system reliability.

In the ideal case, we can collect all the individual failure time of the components. However, in reality, it is impossible to record every individual failure time due to cost constraints and a lack of sufficient collecting instruments. Instead, the actual individual failure time is often not available. The observed data are grouped and presented as a total number of failures for a number of identical components in a system and a cumulative operation time of the system. A typical example of grouped data can be found in the largest publicly available source of reliability data (Denson et al. 1991) collected by the Reliability Analysis Center (RAC). In this data source, field failure data for a large variety of electrical and non-electrical components are available. Most individual component failure time is not recorded while these data are recorded in the way of cumulative operation time and a total number of failures. An illustrative example is presented in Table 1. There are six systems (air planes) and each system has several lights. As we can see, only the number of failures (N_i) within a certain period (T_i) is recorded for each system.

Except for the cost constraint, this lack of information can be explained by the fact that exponential distribution is often used to model individual time-to-failure and in this case the distribution parameter can be estimated with total number of failures and total operation time sufficiently. However, Coit and Dey (1999) stated that exponential distribution is not always appropriate for this kind of data. They developed a hypothesis test to evaluate the validity of an exponential distribution for the grouped data. The hypothesis rejected the exponential distribution assumption when the data were generated from a Weibull distribution. This finding showed that the suitability of individual time-to-failure distribution should be

Table 1 Air plane indicator light failure data

System ID i	Number of failures N_i	Cumulative operation time T_i (1000 h)
1	2	51
2	9	194.9
3	8	45.3
4	8	112.4
5	6	104
6	5	44.8

evaluated. As a start work to deal with the grouped data, Coit and Jin (2000) developed a gamma distribution to fit the data and maximum likelihood estimates (MLE) for the gamma distribution were developed. A simulation study was then carried out to show that parameters of the gamma distribution can be accurately estimated with reasonable sample size.

To the best of our knowledge, no other study has been carried out to deal with the grouped data. To address this important problem, we first use the gamma distribution and Inverse Gaussian (IG) distribution to fit the data. Methods of constructing CI for parameters of gamma distribution and IG distribution in terms of the grouped data are then adopted. At last, random effects are incorporated into the gamma and IG models to represent the system-to-system heterogeneity. The EM algorithm is used to find the MLE for these two random effects models.

The IG distribution plays an important role in the area of probability and physics because of its close relation with the Brownian motion (Seshadri 1993). It has a wide application in many areas, especially in the area of applied statistics. For example, IG distribution and its corresponding IG processes are well received in degradation modeling (Wang and Xu 2010; Ye and Chen 2013) and risk analysis (Ter Berg 1980). More applications of IG distribution can be found in the book written by Chhikara (1988). In our paper, we use IG distribution to fit the individual failure time and the MLE of parameters can be obtained easily.

Another main contribution of this paper is to propose methods of constructing CI for parameters in the gamma distribution and IG distribution when the data are grouped. When data are not grouped, that is, all the individual failure times are observed, the methods of constructing CI for parameters are widely studied and we will review them in Sects. 3 and 4. However, no research has been done to construct CIs when the data are grouped. In our paper, we show that the method of constructing CI for the shape parameter of a gamma distribution and both parameters of an IG distribution in the case of grouped data is quite similar to the case of non-grouped data. In terms of rate parameter and mean of gamma distribution, even when all the individual data are collected, no efficient methods have been found to construct the CIs. In this paper, generalized pivotal quantity (GPQ) (Weerahandi 1995; Hannig et al. 2006) is used to obtain the interval estimation of the rate parameter and mean of the gamma distribution. The simulation results show that our methods of constructing CI have a much better performance than the Wald method in terms of the coverage probability.

The remainder of the paper is organized as follows. Section 2 lists the notations which will be used in this paper. Section 3 introduces how to use gamma distribution to fit the grouped data and methods of constructing CI for gamma distribution are adopted. In Sect. 4, IG distribution is used to fit the grouped data and also the methods of constructing CI are adopted. An illustrative example is provided in Sect. 5. Section 6 concludes the paper.

2 Notations

The following notations are used in the remaining part of this article. We assume the individual component time-to-failure are i.i.d. and is distributed in accordance with a certain distribution. Besides, we assume that each system has only one data record.

i index of system, $i = 1, \cdots, n$
N_i number of observed failures in system i
T_i cumulative operation time in system i
N total number of observed failures of all the systems, $N = \sum_{i=1}^{n} N_i$
T total cumulative operation time of all the systems, $T = \sum_{i=1}^{n} T_i$
X_{ij} jth component failure time in system i, where $i = 1, \cdots, N_i$

3 Gamma Distribution Model

For a random variable X following a gamma distribution with shape parameter k and rate parameter θ, the probability density function (PDF) of X is

$$f(X) = \frac{\theta^k}{\Gamma(k)} X^{k-1} e^{\theta X}$$

when $X > 0$ and 0 otherwise, where $\Gamma(k)$ is the gamma function and $\Gamma(k) = \int_0^\infty x^{k-1} e^{-x} \mathrm{dx}$.

Assume $X_{ij} \sim \Gamma\ (k, \theta)$, where k and θ are shape parameter and rate parameter. Based on the properties of a gamma distribution, we know T_i follows a gamma distribution with shape parameter N_{ik} and rate parameter θ.

We can then obtain the log-likelihood function as follows (omitting some constant parts).

$$l = Nk \log \theta - T\theta - \sum_i [\log \Gamma(N_i k) - N_i k \log T_i]$$

It is easy to get the estimations of parameter k and θ using quasi-Newton search.

3.1 Confidence Interval (CI) for Parameters of Gamma Distribution

3.1.1 Review on Interval Estimation for Gamma Distribution

Some studies have been done to construct the CI for parameters of a gamma distribution when all the individual data, say $x_1, \ldots, x_n$ are observed.

1. Shape Parameter

   The arithmetic mean $\bar{x} = \sum_{i=1}^{n} x_i/n$ and geometric $\tilde{x} = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$ are joint complete sufficient statistics for the gamma distribution. Denote $S = \log(\bar{x}/\tilde{x})$ and $W = 2\, nkS$.

   Bartlett (1937) showed that S has an approximately distribution of

   $$\frac{1 + (1 + \frac{1}{n})/6k}{nk} \chi^2(n-1)$$

   where $\chi^2(n-1)$ is χ^2 distribution with n^{-1} degrees of freedom. The CI for k can then be obtained based on a constructed table of CI. Based on this approximation, Linhart (1965) stated that the quality of the approximation improves with increasing k while almost independent of n.

   Bain and Engelhardt (1975) posed a two-moment χ^2 approximation for Bartlett (1937) showed the cumulant generating function of S and then the expectation and variance of W can be obtained (as a function of k). The authors tried to formulate an approximation distribution for W with correct first and second moments through $c\,(k, n)\, W\, \chi^2\, [v\,(k, n)]$. Afterwards $c\,(k, n)$ and $v\,(k, n)$ can be obtained as a function of k and n. Consequently, the CI for k can also be obtained.

   Hwang and Huang (2002) used the squared sample coefficient of variation V_n^2, where $V_n = s_n^2/\bar{x}$, as a statistic to construct the approximate CI for the shape parameter k. The authors did not show the accurate approximation distribution but stated that V_n^2 looks like a gamma distribution for any parameters and n based on some simulation results. The authors then used the simulation results to construct approximated 75:5, 95 and 98% CI for k.

2. Rate Parameter

   Some literatures have considered the inferences for the rate parameter θ with k unknown.

   Engelhardt and Bain (1977) constructed the uniformly most powerful unbiased (UMPU) test for the rate parameter θ based on the conditional density of $\bar{x}$ given $\tilde{x}$, which is independent of k. By letting $W_n = \bar{x}/\tilde{x}$ and $G_n = \theta\,\tilde{x}$, the anthers provided percentage points $u_\gamma(g)$ such that $P\left\{\sqrt{ng}[W_n - E(W_n|g)] \leq u_\gamma(g)|g\right\}$. They also derived the expression of $u_\gamma(g)$ when $g \to \infty$ and $n \to \infty$. In their next paper (Engelhardt and Bain 1978), they derived an approximate form for $u_\gamma(g)$ by deriving an approximate two-moment chi-square distribution for statistic $Z_n = 2\,ng(W_n - 1)$.

   Kulkarni and Powar (2010) obtained a approximate chi-square distribution for statistic $T_2 = Z + \frac{2n\hat{k}S}{c}$, where $Z = 2\,n\bar{x}\theta$ and $\hat{k}$ is the MLE of k. The authors stated that the performance of Z alone in terms of controlling Type 1 error rates is quite poor when k is small (≤ 1). They also proposed an alternative way to construct CI for θ, using simulation results based on the concept of parametric bootstrapping.

3. Mean
   Some papers have focused on the inference for the mean of a gamma distribution, say μ = k = θ (Grice and Bain 1980; Shiue and Bain 1990; Jensen 1986; Jensen and Kristensen 1991; Wong 1993). Kulkarni and Powar (2010) gave us a detailed review. It is noted that all these methods are based on the statistic $2nk\bar{x}/\mu$, which follows a χ^2 distribution with 2 nk degrees of freedom. In addition, the MLE of k is often substituted to get the CI for μ.

3.1.2 Proposed Method for Interval Estimation

Recall that $T_i \sim \Gamma(N_i k, \theta),\ i = 1, \cdots, n$. Let $X_i = T_i/N_i$ and then $X_i \sim \text{IG}(N_i k, N_i \theta)$. Denote $\tilde{T} = \left(\prod_i X_i^{N_i}\right)^{1/N}$ and $\bar{T} = \sum_i T_i/N$.

1. Shape Parameter
   Let $S_0 = \tilde{T}/\bar{\text{T}}, S_1 = \log S_0, S_2 = S_0^{N_k}$ and $W = -2NkS_1$. We first note that S_0 does not depend on θ as it is a scale invariant statistics. Besides, as any scale invariant statistics is independent of $\sum T_i$ (Pitman 1937), S_0 is independent with $\bar{T}$. Therefore, we can simply let θ = 1.
   We derive a two-moment chi-square approximation for the distribution of W like Bain and Engelhardt (1975). Assume $W \sim c\chi^2(v)$, where c and v are functions of k and n. Moreover, c and v satisfy equations $E(W) = cv$ and $\text{Var}\,(W) = 2c^2 v$ to make the first two moments accurate, where E (W) and Var (W) can be obtained by the moment-generating function (MGF) of W.
   It is not hard to show the following two statements

   (a) when $k \to 0$, then $c \to 1$ and $v \to 2(n-1)$
   (b) when $k \to \infty$, then $c \to 1$ and $v \to 2(n-1)$

   Having this approximate distribution, a $1-\alpha$ lower confidence bounds k_l can be found by

$$-2Nk_l\ \log\frac{\tilde{T}}{\bar{T}} = c(k_l, n)\chi^2_\alpha[v(k_l, n)]$$

   A $1-\alpha$ equal-tails level CI for θ, say (k_l, k_u) can be found by

$$-2Nk_l \log\frac{\tilde{T}}{\bar{T}} = c(k_l, n)\chi^2_{\alpha/2}[v(k_l, n)]$$

   and

$$-2Nk_u \log\frac{\tilde{T}}{\bar{T}} = c(k_u, n)\chi^2_{1-\alpha/2}[v(k_l, n)]$$

2. Rate parameter

   As for the rate parameter, it is well known that $2T\theta \sim \chi^2$ (2 Nk), where $T = \sum_i T_i$. If k is known, then it is easy to construct the CI for parameter θ. However, when k is unknown, the accurate form cannot be obtained easily. In the literature, $\hat{k}$ is usually used to replace k and then the CI for θ can be obtained. This method is easy to implement while not so accurate. We proposed a new method here based on the idea of GPQ. The basic idea is if we know the exact distribution of k, then $\chi^2(2Nk)/2T$ can be regarded as a GPQ of θ. We can simply generate a large number of random variables from the distribution of k and then a series number of value of θ can be derived. We can obtain the CI for θ at any level based on this series of value. Unfortunately, the exact distribution of k is unknown. We used the approximate distribution of k, $W = -2Nk \log\frac{\tilde{T}}{T} \sim c\chi^2(v)$, to construct the confidence interval for k. The precise procedures are as follows.

   Step 1 Use the MLE of k, denoted as $\hat{k}$, to get the value of c and v, denoted as $\hat{c}$ and $\hat{v}$;

   Step 2 Based on the approximate distribution $W = -2Nk \log\frac{\tilde{T}}{T} \sim c\chi^2(v)$, generate B random variables from the distribution $\hat{c}\chi^2(\hat{v})$ and consequently we can get B value of k, denoted as $k_1, \ldots, k_B$;

   Step 3 $\chi^2(2Nk_b) = 2T, b = 1 \ldots, B$ forms a series value of q and we can get CI for θ at any level α.

3. Mean

   Sometimes, the mean of a gamma distribution is of more interest. Constructing the confidence interval for the mean $\mu = k/\theta$ is quite similar to the rate parameter. As we know $\frac{2Tk}{\mu} \sim \chi^2(2Nk)$, the idea of GPQ can then be used.

   Step 1 Use the MLE of k, denoted as $\hat{k}$, to get the value of c and v, denoted as $\hat{c}$ and $\hat{v}$;

   Step 2 Based on the approximate distribution $W = -2Nk \log\frac{\tilde{T}}{T} \sim c\chi^2(v)$, generate B random variables from the distribution $\hat{c}\chi^2(\hat{v})$ and consequently we can get B value of k, denoted as $k_1, \ldots, k_{\mathrm{B}}$;

   Step 3 $2Tk_b/\chi^2(2Nk_b), b = 1, \cdots, B$ forms a series value of μ and we can get CI for μ at any level α.

3.2 Simulation Results for Interval Estimation of Gamma Distribution

We do a simulation to verify our methods for gamma distribution by comparing the performance with Wald method in terms of coverage probability.

For the shape parameter, the Wald interval can be obtained by the approximate distribution

$$\sqrt{-\frac{\partial^2}{\partial k^2}\log L(k,\theta)|_{(k,\theta)=(\hat{k},\hat{\theta})}}(\hat{k}-k) \sim N(0,1)$$

For our proposed method, as parameter θ does not affect the result, we simply let $\theta = 1$. Scenarios with $k = 0.05$, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 and $n = 5$; 10; 20 are studied, respectively. In each scenario, a level α test of $H_0 : k \leq k_0$ versus $H_1 : k > k_0$ is used and the coverage probability is obtained by 10,000 replications.

Tables 2 and 3 show the coverage probability of k in the case of $\alpha = 0{:}05$ and $\alpha = 0{:}1$, respectively. As can be seen, the proposed method outperforms the Wald method in terms of coverage probability in any scenario. In addition, the coverage probability by our proposed method is very close the nominal one, i.e., $1 - \alpha$, even when the sample size is very small ($n = 5$). Based on the results, we believe it is appropriate to treat the proposed approach as optimal in practice.

For the rate parameter θ, the Wald interval can be obtained by the approximate distribution

Table 2 Coverage probability of k by proposed method (CP_p) and Wald method (CP_w) in the case $\alpha = 0.05$

k	$n = 5$		$n = 10$		$n = 20$	
	CP_p	CP_w	CP_p	CP_w	CP_p	CP_w
0.05	0.9384	0.9056	0.9465	0.9185	0.9481	0.9310
0.1	0.9398	0.8301	0.9449	0.8749	0.9453	0.9106
0.2	0.9485	0.7257	0.9467	0.7950	0.9463	0.8507
0.3	0.9475	0.6725	0.9470	0.7434	0.9497	0.8230
0.4	0.9491	0.6231	0.9501	0.7064	0.9514	0.7815
0.5	0.9535	0.6002	0.9434	0.6733	0.9510	0.7486
0.6	0.9523	0.5671	0.9518	0.6406	0.9465	0.7380
0.7	0.9575	0.5430	0.9479	0.6341	0.9499	0.7112
0.8	0.9605	0.5442	0.9491	0.6002	0.9499	0.6943
0.9	0.9627	0.5200	0.9484	0.5993	0.9519	0.6780
1	0.9643	0.5002	0.9587	0.5836	0.9523	0.6749

Table 3 Coverage probability of k by proposed method (CPp) and Wald method (CPw) in the case $\alpha = 0.1$

k	$n = 5$		$n = 10$		$n = 20$	
	CP_p	CP_w	CP_p	CP_w	CP_p	CP_w
0.05	0.8885	0.8160	0.8960	0.8451	0.8978	0.8614
0.1	0.8911	0.7526	0.8966	0.8030	0.8980	0.8202
0.2	0.8890	0.6660	0.9017	0.7416	0.9000	0.7634
0.3	0.9001	0.6153	0.8922	0.6930	0.8991	0.7222
0.4	0.9037	0.5694	0.8976	0.6621	0.8994	0.6886
0.5	0.9058	0.5484	0.9023	0.6389	0.9015	0.6633
0.6	0.9058	0.5352	0.8986	0.6150	0.8976	0.6512
0.7	0.9118	0.4989	0.8955	0.5937	0.9010	0.6247
0.8	0.9218	0.4937	0.9080	0.5779	0.8982	0.6161
0.9	0.9236	0.4801	0.9098	0.5679	0.9040	0.6011
1	0.9326	0.4670	0.9043	0.5624	0.9049	0.6026

$$\sqrt{-\frac{\partial^2}{\partial\theta^2}\log L(k,\theta)}\bigg|_{(k,\theta)=\left(\hat{k},\hat{\theta}\right)}\left(\hat{\theta}-\theta\right)\sim N(0,1)$$

Scenarios with $k = 0.5,\ 1,\ 3$ and $\theta = 0.5,\ 1,\ 3$ are studied. We set $B = 10{,}000$ and repeat 10,000 times simulation in each scenario (under certain k and θ). Similar to the test for shape parameter k, a level α test of $H_0 : \theta \le \theta_0$ versus $H_1 : \theta > \theta_0$ is used.

Tables 4 and 5 are the simulation results by proposed method and Wald method in terms of coverage probability. As we can see, our proposed approach performs much better than the Wald method and can provide a quite accurate coverage probability even when the sample size is limited. In fact, we proposed a new method of constructing the CI for rate parameter of a gamma distribution as non-grouped data is a special case of grouped data, which performs much better than any other existing method.

For the mean μ, the log-likelihood function for a gamma distribution with shape parameter k and mean parameter μ is as follows.

$$l(k,\mu)=\sum_i\left[(N_ik-1)\log T_i-\ \log\Gamma(N_ik)-N_ik\ \log(\mu/k)-kT_i/\mu\right]$$

The Wald interval for μ can be constructed by the following approximate distribution

$$\sqrt{-\frac{\partial^2}{\partial\mu^2}\log L(k,\mu)}\bigg|_{(k,\mu)=\left(\hat{k},\hat{\mu}\right)}(\hat{\mu}-\mu)\sim N(0,1)$$

Scenarios with $k = 0.5$, 1, 3 and $\mu = 0.25$, 0.5, 0.5, 1, 1.5, 3 are studied. We also let B = 10,000 and repeat simulation 10,000 times in each scenario. A level α test

Table 4 Coverage probability of θ by proposed method (CPp) and Wald method (CPw) in the case α = 0.05

k	θ	$n = 5$		$n = 10$		$n = 20$	
		CP_p	CP_w	CP_p	CP_w	CP_p	CP_w
0.5	0.5	0.9555	0.7744	0.9594	0.9038	0.9604	0.9703
	1	0.9557	0.6762	0.9589	0.8190	0.9550	0.9169
	3	0.9555	0.5452	0.9575	0.6810	0.9586	0.7739
1	0.5	0.9527	0.7244	0.9545	0.875	0.9540	0.9540
	1	0.9517	0.6359	0.9550	0.7867	0.9564	0.8898
	3	0.9545	0.5060	0.9542	0.6330	0.9565	0.7507
3	0.5	0.9500	0.6548	0.9518	0.8025	0.9508	0.9038
	1	0.9513	0.5602	0.9502	0.7203	0.9506	0.8195
	3	0.9558	0.4659	0.9520	0.5820	0.9549	0.6726

Table 5 Coverage probability of θ by proposed method (CPp) and Wald method (CPw) in the case α = 0.1

k	θ	$n = 5$		$n = 10$		$n = 20$	
		CP_p	CP_w	CP_p	CP_w	CP_p	CP_w
0.5	0.5	0.9081	0.7000	0.9103	0.8542	0.9088	0.9364
	1	0.9065	0.6076	0.9108	0.7662	0.9106	0.8632
	3	0.9037	0.4869	0.9121	0.6185	0.9187	0.7139
1	0.5	0.9103	0.6693	0.9048	0.8273	0.9089	0.9057
	1	0.9045	0.5755	0.9076	0.7272	0.9134	0.8291
	3	0.9073	0.4637	0.9053	0.5839	0.9090	0.6822
3	0.5	0.8977	0.5944	0.8979	0.7551	0.9082	0.8428
	1	0.8967	0.5149	0.8967	0.6546	0.9027	0.7486
	3	0.9003	0.4297	0.8992	0.5490	0.9050	0.6330

of $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$ is used and the simulation results are shown in Tables 6 and 7 where α = 0:05 and 0:1, respectively. As expected, our methods outperform the Wald methods in terms of the coverage probability.

4 IG Distribution Model

The PDF of an IG distribution is given by

$$f(X) = \sqrt{\frac{\lambda}{2\pi X^3}} \exp \frac{-\lambda (X - \mu)^2}{2\mu^2 X}$$

for $X > 0$, where μ and λ are mean and shape parameters, respectively.

Table 6 Coverage probability of μ by proposed method (CPp) and Wald method (CPw) in the case α = 0.05

k	μ	n = 5		n = 10		n = 20	
		CP_p	CP_w	CP_p	CP_w	CP_p	CP_w
0.5	0.25	0.9520	0.9668	0.9515	0.9682	0.9503	0.9743
	0.5	0.9453	0.9667	0.9519	0.9686	0.9508	0.9726
1	0.5	0.9472	0.9491	0.9515	0.9544	0.9493	0.9645
	1	0.9542	0.9531	0.9462	0.9571	0.951	0.9632
3	1.5	0.9503	0.9279	0.9472	0.9419	0.9537	0.9538
	3	0.9491	0.9286	0.9505	0.9439	0.9511	0.9532

Table 7 Coverage probability of μ by proposed method (CPp) and Wald method (CPw) in the case α = 0.1

k	μ	n = 5		n = 10		n = 20	
		CP_p	CP_w	CP_p	CP_w	CP_p	CP_w
0.5	0.25	0.8949	0.9232	0.8944	0.9333	0.8985	0.9285
	0.5	0.9034	0.9244	0.8986	0.9364	0.9071	0.9296
1	0.5	0.9000	0.9066	0.9017	0.9172	0.8899	0.9227
	1	0.8996	0.9036	0.8951	0.9163	0.9026	0.9125
3	1.5	0.8971	0.8727	0.8987	0.8938	0.9000	0.9096
	3	0.8991	0.8786	0.8966	0.9013	0.9066	0.9016

Assume $X_{ij} \sim IG(\mu, \lambda)$ and then $T_i \sim IG\left(N_i\mu, N_i^2\lambda\right)$. Similar to the case of gamma distribution, we can obtain the log-likelihood function as follows (omitting some constant part)

$$l = \sum_i \left[\frac{1}{2}\log\lambda - \frac{\lambda(T_i - N_i\mu)^2}{2\mu^2 T_i}\right]$$

the MLEs of μ and λ can be derived directly as $\hat{\mu} = \bar{T}$ and $\frac{1}{\hat{\lambda}} = \frac{1}{n}\left[\sum_i \frac{N_i^2}{Ti} - \frac{N}{\bar{T}}\right]$.

4.1 Confidence Interval; for Parameters of IG Distribution

Assume n grouped data $T_1, \ldots, T_n$ are observed, and $T_i \sim IG\left(N_i\mu, N_i^2\lambda\right)$. Denote $V = \sum_i \frac{Ni^2}{T_i} - \frac{N}{\bar{T}}$ where $N = \sum_i N_i$ and $\bar{T} = \sum_i T_i/N$ and then it can be easily proved that $\lambda V \sim \chi^2(n-1)$. Based on this distribution, the two-sided 100 (1−α) CI for λ can be

$$\left(\frac{\chi^2_{\alpha/2}(n-1)}{V}, \frac{\chi^2_{1-\alpha/2}(n-1)}{V}\right)$$

On the other hand, for the mean parameter μ, it can be shown that statistic $H = \frac{\sqrt{N(n-1)}(\bar{T}-\mu)}{\mu\sqrt{\bar{T}V}}$ follows a weighted student's t distribution. That is

$$f_H(h) = \frac{\left[1 + h^2/(n-1)\right]^{-n/2}}{\sqrt{n-1}\beta\left(\frac{1}{2}, \frac{n-1}{2}\right)} q(h)$$

where

$$q(h) = 1 - \frac{h}{\Gamma(n/2)} \int_0^\infty \frac{w^{(n-1)/2} e^{-w} dw}{\sqrt{h^2 w + 2N\lambda(h^2 + n - 1)/\mu}}$$

and $w = \frac{u}{2}\left(1 + \frac{h^2}{n-1}\right)$.

Based on this statistics, a two-sided 1-α CI for μ can be

$$\left(\frac{\bar{T}}{1 + \frac{t_{1-\alpha/2}\sqrt{\bar{T}V}}{\sqrt{N(n-1)}}}, \frac{\bar{T}}{1 - \frac{t_{1-\alpha/2}\sqrt{\bar{T}V}}{\sqrt{N(n-1)}}}\right)$$

5 Illustrative Example

We use the example of the air plane light as listed in Table 1. Coit and Jin (2000) used a gamma distribution to fit the data and MLEs can be efficiently found by quasi-Newton search.

Table 8 shows the point and interval estimate for parameters of simple gamma and IG distributions.

We then assess the goodness of fit of these models by a simple graphical method. For the gamma distribution if $X \sim \Gamma(k, \theta)$, where k and θ are shape and rate parameter, Wilson and Hilferty (1931) stated that $\sqrt[3]{X\hat{\theta}/\hat{k}}$ is approximately normally distributed with mean $1 - 1/9\hat{k}$ and variance $1/9\hat{k}$. For the IG distribution, Wang (2010) noted that if $X \sim IG(\mu, \lambda)$, then $\lambda(X-\mu)^2/(\mu^2 X) \sim \chi^2$ (1). The MLEs of μ and λ, $\hat{\mu}$ and $\hat{\lambda}$, can be substituted into the equation to assess the adequacy of the IG distribution. Figure 1 shows the quantile–quantile (Q–Q) plot based on above approximate distribution. As can be seen, gamma distribution provides a better fit to this grouped data though the system number is too small to get a more convincing conclusion.

Table 8 MLEs and CIs for gamma and IG distribution

Gamma		IG	
k	θ	μ	λ
0.73	0.048	14.536	7.829
(0.12849, 1.47441)	(0.00498, 0.10638)	(8.95978, 38.50372)	(1.08461, 16.74454)

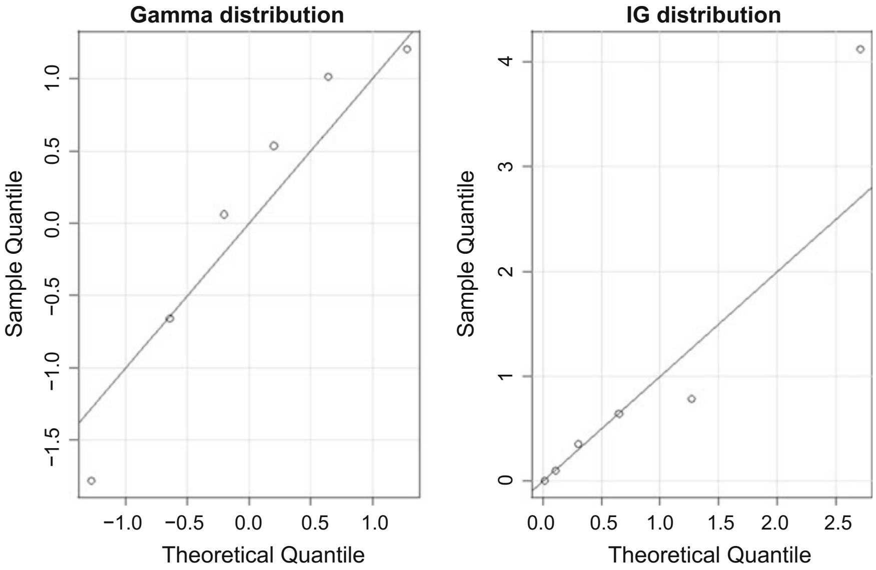


Fig. 1 Goodness of fit

6 Conclusion

When the observed field failure data are grouped, represented as a total number of failures and cumulative operation time, traditional techniques to deal with incomplete data are not applicable. In this paper, we used gamma distribution and IG distribution to fit the merged data. Method of constructing confidence intervals for parameters of these two distributions when data are merged was proposed. It is noted here that even when the data are not merged, that is, all the individual failure time are collected, no efficient method has been found to construct CI for the rate parameter of a gamma distribution. Our proposed method is proved to efficiently solve this problem even when the data are merged. On the whole, this study has provided detailed statistical inference and systematically investigated this non-standard form incomplete database.

It is recommended that further research be undertaken in the following areas. First, other distributions, like wiener or lognormal distribution can be taken into consideration to fit the grouped data. This is a more difficult problem as the merged data do not follow the distribution of individual time-to-failure any more. In addition, nonparametric inference for the grouped data is of interest. As the exact individual failure time is not available, the Kaplan–Meier method is not applicable for the nonparametric estimation. Furthermore, as the data are recorded by the number of failures and cumulative operation time, the cumulative failure time will be less than the cumulative operation time if there is no failure at the end of observation. This is not uncommon and should be considered in the future work.

References

Bain, L. and Engelhardt, M. (1975). A Two-Moment Chi-Square Approximation for the Statistic Log. 70(352):948–950.

Bartlett, M. S. (1937). Properties of Sufficiency and Statistical Tests. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 160(901):268–282.

Chhikara, R. (1988). The Inverse Gaussian Distribution: Theory: Methodology, and Applications, volume 95. CRC Press.

Coit, D. and Jin, T. (2000). Gamma distribution parameter estimation for field reliability data with missing failure times. IIE Transactions.

Coit, D. W. and Dey, K. A. (1999). Analysis of grouped data from field-failure reporting systems. Reliability Engineering & System Safety, 65(2):95–101.

Denson, W., Chandler, G., Crowell, W., and Wanner, R. (1991). Nonelectronic parts reliability data 1991.

Engelhardt, M. and Bain, L. (1977). Uniformly most powerful unbiased tests on the scale parameter of a gamma distribution with a nuisance shape parameter. Technometrics, 19(1):77–81.

Engelhardt, M. and Bain, L. (1978). Construction of optimal unbiased inference procedures for the parameters of the gamma distribution. Technometrics, 20(4):485–489.

Grice, J. V. and Bain, L. J. (1980). Inferences concerning the mean of the gamma distribution. Journal of the American Statistical Association, 75(372):929–933.

Hannig, J., Iyer, H., and Patterson, P. (2006). Fiducial Generalized Confidence Intervals. Journal of the American Statistical Association, 101(473):254–269.

Hwang, T. and Huang, P. (2002). On new moment estimation of parameters of the Gamma distribution using its characterization. Annals of the Institute of Statistical Mathematics.

Jensen, J. and Kristensen, L. (1991). Saddlepoint approximations to exact tests and improved likelihood ratio tests for the gamma distribution. Communications in Statistics-Theory and Methods, 20(4):1515–1532.

Jensen, J. L. (1986). Inference for the mean of a gamma distribution with unknown shape parameter. Scandinavian journal of statistics, pages 135–151.

Kulkarni, H. V. and Powar, S. K. (2010). A new method for interval estimation of the mean of the Gamma distribution. Lifetime data analysis, 16(3):431–47.

Linhart, H. (1965). Approximate confidence limits for the coefficient of variation of gamma distributions. Biometrics, 21(3):733–738.

Pitman, E. J. (1937). The closest estimates of statistical parameters. In Proceedings of the Cambridge Philosophical Society, volume 33, pages 212–222. Cambridge Univ Press.

Seshadri, V. (1993). The inverse Gaussian distribution: a case study in exponential families. Clarendon Press Oxford.

Shiue, W. and Bain, L. J. (1990). Simple approximate inference procedures for the mean of the gammma distribution. Journal of Statistical Computation and Simulation, 34(2–3):67–73.

Ter Berg, P. (1980). Two pragmatic approaches to loglinear claim cost analysis.

Wang, X. (2010). Wiener processes with random effects for degradation data. Journal of Multivariate Analysis, 101(2):340–351.

Wang, X. and Xu, D. (2010). An Inverse Gaussian Process Model for Degradation Data. Technometrics, 52(2):188–197.

Weerahandi, S. (1995). Generalized confidence intervals. Exact Statistical Methods for Data Analysis, (1961).

Wilson, E. B. and Hilferty, M. M. (1931). The distribution of chi-square. Proceedings of the National Academy of Sciences of the United States of America, 17(12):684.

Wong, A. (1993). A note on inference for the mean parameter of the gamma distribution. Statistics & probability letters, 17(1):61–66.

Ye, Z. and Chen, N. (2013). The inverse Gaussian process as a degradation model. Technometrics, pages 1–28.

The Correlation Between Device Aging and System Degradation

Song Lan and Cher Ming Tan

Abstract An electronic system contains millions of components on a single chip today, and the failure of any of these components may lead to the system failure. Conventionally, the system reliability is ensured by having a high reliable subsystem or component, and thus efforts are placed to study the reliability of individual components in electronic system. However, the failure mechanisms of a system is lot more complex, and the correlation between the system failure and component failure is not straight forward, depending on the system configuration, as demonstrated in this work. Therefore, electronic system reliability may not be derived from its components' reliability.

Keywords Component reliability · System degradation · System configuration · Feedback · Component importance

1 Introduction

Reliability is the probability that a product or system can perform its intended function for a specific period of time (O'Connor et al. 2002). Conventionally, system reliability is ensured by having high reliable subcomponent or subsystem within it, and the components' reliability is obtained based on time-to-failure model. However, with the improvement in technology, the time-to-failure of components is becoming longer, and the test time for conventional time-to-failure method will be impractically long. As such, today's reliability study for electronics is performed by monitoring its performance degradation (Song and Cher Ming 2013). The

S. Lan
Department of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore

C.M. Tan (✉)
Department of Electronic Engineering, Chang Gung University, Taoyuan City 33302, Taiwan
e-mail: cmtan@mail.cgu.edu.tw

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_22

performance degradation path with time is defined as $D(t)$. When $D(t)$ exceeds, the maximum allowed degradation limit D_f, the component is treated as failure. Comparing to the conventional time-to-failure model, the degradation study provides more physical information related to component's failure/degradation mechanism. Moreover, the performance degradation can also be used to estimate component time-to-failure distribution (Song and Cher Ming 2013).

The performance degradation has been widely used to study the reliability of single component such as transistor/battery lifetime (Chenming et al. 1985; Chung et al. 1991) where the lifetime of a single component is usually defined as certain percentage drift of its parameters. Unfortunately, the lifetime of a single component is not directly related to the electronic system degradation since different systems have different immunity to its internal component degradation. In this case, the system strength index should be defined in order to study its reliability.

In this work, the correlations between components' degradation and system degradation will be introduced. We begin with a simple resistor network, followed by a more complex circuit where feedback configuration is used.

2 Resistor Network Degradation

Resistor network is widely used for voltage divider, current divider, and digital-to-analog convertor, to name a few. In this work, three resistor network configurations will be studied, and the impact of the resistors' degradation on network performance will be presented. In this work, the system performance is defined by the total impedance of the resistor network R_{sys}. The correlation between individual resistor degradation ΔR_{ind} and ΔR_{sys} is discussed and simulated using LTspice.

A series configuration resistor network is shown in Fig. 1, and it is frequently used as voltage divider. The total impedance R_{sys} is expressed as

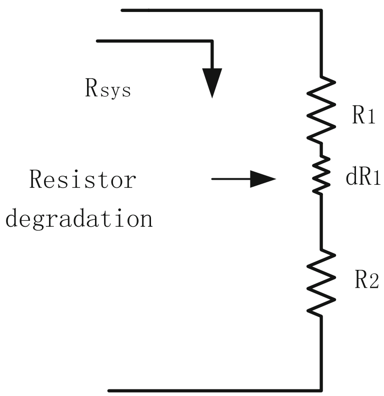


Fig. 1 Series configuration resistor network

$$R_{\mathrm{sys}} = R_1 + R_2 \tag{1}$$

If R_1 is the resistor that will degrade with stress time, while R_2 remains unchanged, the impact of R_1 degradation on the system performance can be estimated by differentiating Eq. (1) by R_1 and dividing both sides by R_{sys} as given below.

$$\frac{dR_{\mathrm{sys}}}{R_{\mathrm{sys}}} = \frac{dR_1}{R_1 + R_2} \tag{2}$$

Given that $R_2 = mR_1$, where m is the resistor ratio, the correlation between system degradation and R_1 degradation can be expressed as

$$\frac{dR_{sys}}{R_{sys}} = \frac{dR_1}{(1+m)R_1} \tag{3}$$

From Eq. (3), it can be inferred that the system is more immune to the resistor degradation if the resistor ratio is large. In other words, with the m $\rightarrow \infty$, the R_1/R_{sys} is close to zero, and the impact of R_1 on the system performance is negligible. The simulation result is shown in Fig. 2, and one can see that the impact of resistor degradation on system performance is inversely proportional to resistor ratio m.

Parallel resistor network is another configuration that is widely used as current divider. A classical configuration of the parallel resistor network is shown in Fig. 3. Similar to the series configuration resistor network, R_1 is assumed to be the resistor that degrades with time. The system performance is defined by the total impedance R_{sys} given below.

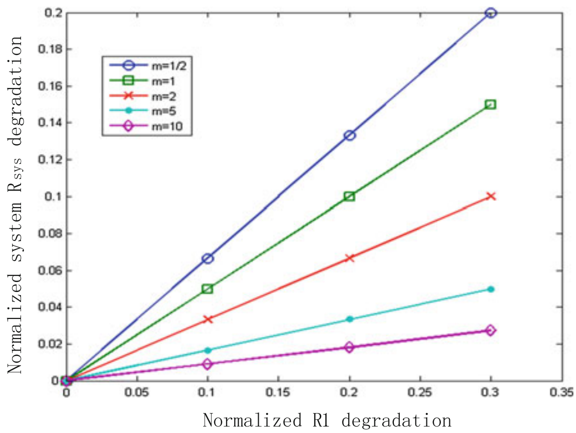
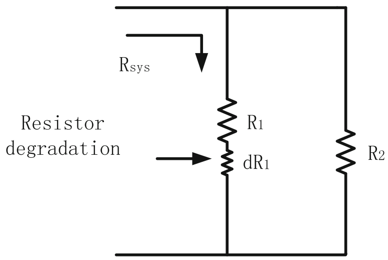


Fig. 2 System degradation versus resistor degradation for series configuration

Fig. 3 Parallel configuration resistor network



$$R_{sys} = \frac{R_1 R_2}{R_1 + R_2} \quad (4)$$

The impact of R_1 degradation on system deterioration is calculated by differentiating Eq. (4) by R_1 and dividing both sides by R_{sys} as shown in Eq. (5)

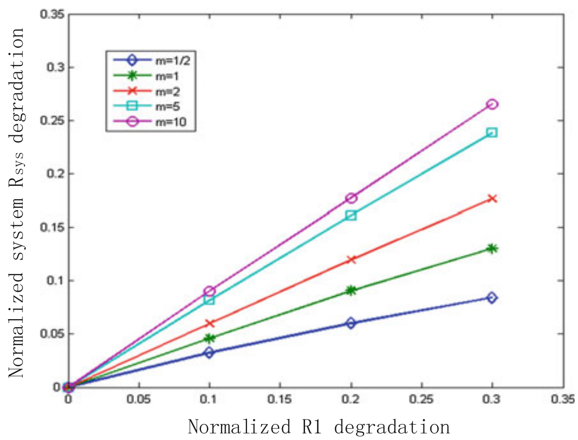
$$\frac{dR_{sys}}{R_{sys}} = \frac{R_2 dR_1}{(R_1 + R_2)R_1} \quad (5)$$

With $R_2 = mR_1$, Eq. (5) becomes

$$\frac{dR_{sys}}{R_{sys}} = \frac{dR_1}{\left(1 + \frac{1}{m}\right)R_1} \quad (6)$$

Unlike the series configuration, the impact of resistor degradation on system performance is proportional to the resistor ratio *m*. For parallel-connected resistor network, the total impedance is mainly affected by the resistor with smaller impedance. As *m* increases, the system will be more sensitive to the change in R_1.

Fig. 4 System degradation versus resistor degradation for parallel configuration



The simulation result of the parallel configuration resistor network is shown in Fig. 4, and we see that the sensitivity increases with the value of m as expected.

Besides series and parallel configuration, delta connected resistor network is another configuration used in three-phase power system as shown in Fig. 5a. Comparing to the series/parallel configuration, the analysis of the delta resistor network is more complicated. To simplify the analysis, the impact of the degradation of the center resistor R_c is studied in this work.

Usually, the delta configuration network is transformed to Y configuration network to calculate the overall impedance as shown in Fig. 5b. The resistances of R_1, R_2, R_3, and R_{sys} can be expressed as

$$R_1 = \frac{R_c R_{y1}}{R_{y1} + R_{y2} + R_c} \tag{7}$$

$$R_2 = \frac{R_c R_{y2}}{R_{y1} + R_{y2} + R_c} \tag{8}$$

$$R_3 = \frac{R_{y2} R_{y1}}{R_{y1} + R_{y2} + R_c} \tag{9}$$

$$R_{sys} = \frac{(R_{x1} + R_1)(R_{x2} + R_2)}{R_{x1} + R_1 + R_{x2} + R_2} + R_3 \tag{10}$$

By substituting Eq. (7)–(9) into Eq. (10), the total impedance of the system is expressed as

$$R_{sys} = \frac{\left(R_{x1} + \frac{R_c R_{y1}}{R_{y1} + R_{y2} + R_c}\right)\left(R_{x2} + \frac{R_c R_{y2}}{R_{y1} + R_{y2} + R_c}\right)}{R_{x1} + \frac{R_c R_{y1}}{R_{y1} + R_{y2} + R_c} + R_{x2} + \frac{R_c R_{y2}}{R_{y1} + R_{y2} + R_c}} + \frac{R_{y2} R_{y1}}{R_{y1} + R_{y2} + R_c} \tag{11}$$

The sensitivity of the system performance to the degradation of R_c is estimated by differentiating Eq. (11) by R_c. With $a = R_{y1}/R_{x1}$ and $b = R_{y2}/R_{x2}$, Eq. (11) becomes

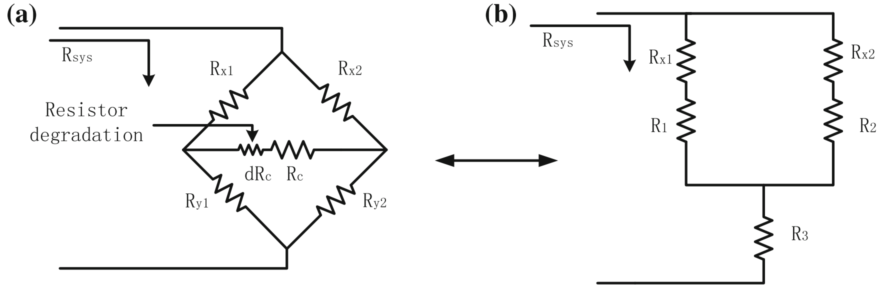
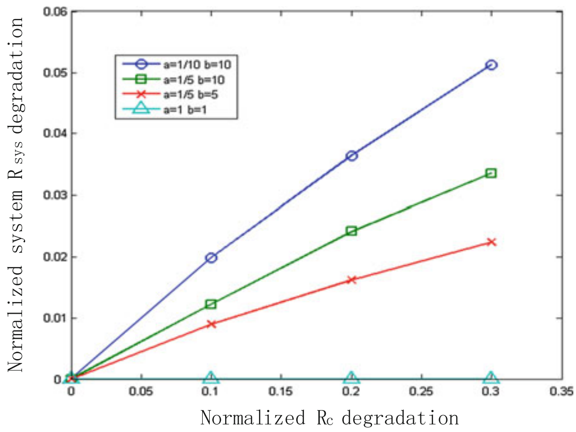


Fig. 5 **a** Delta configuration network **b** Equivalent Y configuration

Fig. 6 System degradation versus resistor degradation for delta configuration



$$\frac{dR_{\text{sys}}}{dR_c} = \frac{R_{x1}^2 R_{x2}^2 (a-b)^2}{\left[R_c(R_{x1}+R_{x2}+aR_{x1}+bR_{x2})+aR_{x1}^2+bR_{x2}^2+R_{x1}R_{x2}(a+b)\right]^2} \tag{12}$$

It can be found that the sensitivity of the system performance is proportional to the term (a–b), which is the resistor ratio difference of R_{y1}/R_{x1} and R_{y2}/R_{x2}. With $a = b$, the voltage difference between R_c is zero and no current can follow through the resistor. In this case, the system performance is independent of the degradation of R_c. Thus, the system is completely immune to R_c degradation with $a = b$, and any improvement in the reliability of R_c will not result in any improvement in the reliability of the network. The simulation result is shown in Fig. 6, and the sensitivity of the system degradation is proportional to the difference of a and b.

When $a = b$, R_{sys} is immune to R_c as observed. However, this does not imply that R_{sys} is immune to the degradation of other resistors' values. It only indicates that R_c is not a component of important for the reliability of the network for $a = b$, and one has to identify other components of important. There are several methods such as Birnbaum's measure and Risk reduction worth, to name a few, can be used to identify the components of important other than the difference of a and b (Rausand and Høyland 2004).

From the above simple illustrations of three resistor networks, one can see that the correlations between system degradation and the internal resistor degradation is related to the network configuration and the resistor ratio in the resistor network. There is no generic formula that can be used to predict the correlation between them.

Let us now go to a more complex circuit where feedback is used for circuit performance stabilization. A LED driver is used as an example in this work.

3 The Impact of Transistor Degradation on Driver Performance

Due to the higher efficiency, LED is expected to replace the traditional incandescent light. However, LED cannot be driven by AC power supply and an LED driver is needed to provide constant voltage/current (Lan et al. 2014). The LED driver is found to be a weak element in LED luminaries (Lan et al. 2012).

The LED driver considered in this work is a linear-mode LED driver. Its equivalent circuit and the I–V sweep characteristics are shown in Fig. 7 (Song and Cher Ming 2013). The I–V sweep characteristics shown in Fig. 7b include two key indices that indicate the performance of the LED driver. The first index is the current level I_{out} and the second index is the knee point V_{knee}.

From the equivalent circuit, the linear-mode LED driver regulates the output current through the feedback circuit. The current level determines the output lumen of the LED. Due to the feedback network, the output current is always regulated to

$$I_{out} = \frac{V_{ref}}{R_{fb}} \tag{13}$$

Therefore, the output current is independent of the transistor degradation. On the other hand, the knee point voltage is observed to increase with time. This knee point is the minimum required output voltage for the proper operation of the LED driver. It has been proved that the knee point degradation is caused by Hot Carrier Injection (HCI) of the output transistor (Song and Cher Ming 2013; Lan et al. 2012, 2014). However, most studies on HCI are based on single-transistor degradation (Chenming et al. 1985; Ang and Ling 1998, 1999; Cham et al. 1987). The correlation between transistor aging and driver performance deterioration needs therefore to be

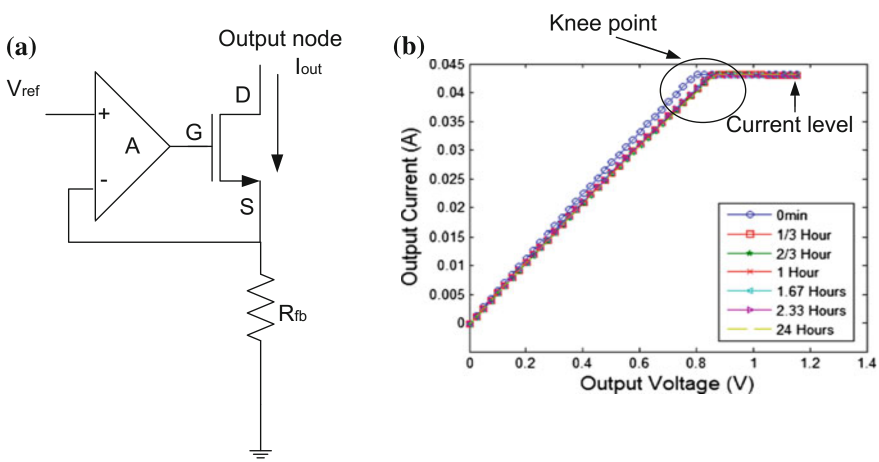


Fig. 7 **a** Equivalent circuit of LED driver **b** Knee point degradation

investigated. As discussed in our previous work (Song and Cher Ming 2013), V_{knee} is function of the output impedance $R_{out\text{-}bef\text{-}knee}$ and output current I_{out} is given by

$$V_{knee} = I_{out} R_{out-bef-knee} \tag{14}$$

As the output impedance consists of the transistor impedance $R_{trans.}$ and R_{fb} in series, the knee point can be expressed as follows:

$$V_{knee} = I_{out}(R_{trans.} + R_{fb}), \tag{15}$$

where $R_{trans.}$ is the on-resistance of the transistor and can be expressed as

$$R_{trans.} = \frac{1}{K_n\left(V_{gs} - V_{th}\right)} \tag{16}$$

The degradation of the knee point voltage is caused by the increase in V_{th} and reduction of K_n (Song and Cher Ming 2013). Therefore, the correlation between the transistor degradation and that of the knee point voltage is similar to that of series configuration resistor network. The sensitivity of knee point degradation to transistor degradation can be calculated by differentiating Eq. (15) with $R_{trans.}$ as shown below:

$$\frac{dV_{knee}}{V_{knee}} = \frac{dR_{trans.}}{R_{trans.} + R_{fb}} \tag{17}$$

Thus, the knee point sensitivity is related to the ratio of R_{trans}/R_{fb}.

To verify this correlation, the linear-mode driver is simulated using different output transistors with different R_{fb}. The overall I–V sweeps of these circuits are about the same as they have similar output impedance. The degradation of the transistor is modeled by reducing transistor parameter K_n value, which are equivalent to the mobility reduction due to HCI. The transistor and feedback resistor R_{fb} parameters are summarized in Table 1. The transistor degradation is set to be from 0 to 30%.

The knee point degradation is extracted based on I–V curve as shown in Fig. 8a–c. The normalized knee point degradation versus that of the transistor is shown in Fig. 8d. One can see that the knee point degradation is more sensitive to the degradation of transistor 1, since its impedance is comparable to the nondegraded R_{fb}. On the other hand, the degradation of transistor 3 is negligible as R_{fb} is much

Table 1 Transistor parameters

	K (A/V^2)	$R_{trans.}$ (Ω)	R_{fb} (Ω)	$R_{out\text{-}bef\text{-}knee}$ (Ω)
Transistor 1	0.025	10	10	20
Transistor 2	0.05	5	15	20
Transistor 3	2	0.125	20	20.125

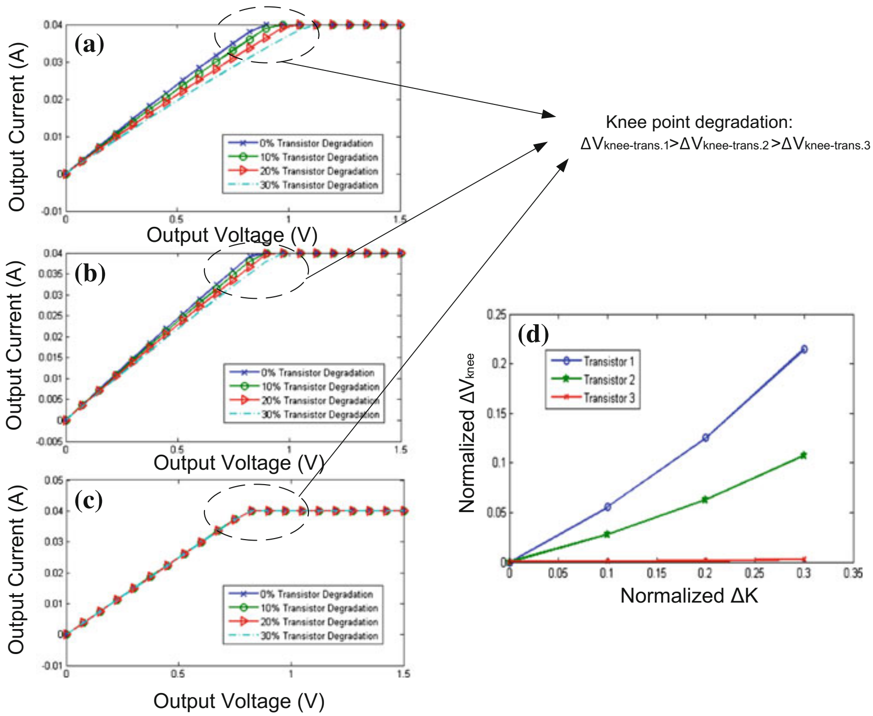


Fig. 8 Knee point degradation of different transistors **a** *Transistor 1* **b** *Transistor 2* **c** *Transistor 3* **d** Knee point degradation versus transistor degradation

larger than the impedance of transistor. The results clearly confirm that the same percentage degradation of transistor with different transistors has different impact on the circuit performance, which is similar to that of series connected resistor network.

For the case of transistor 3, the system (i.e., driver circuit) performance does not degrade even if transistor 3 is degrading. However, this does not imply that the circuit will not degrade. It simply implies that its performance is robust against the output transistor. However, it should be noted that as HCI continues, the gate oxide becomes weak and it can lead to gate oxide breakdown, causing system hard failure. Moreover, the system performance can also be affected by other part of the circuit such as Electromigration of the output interconnection.

4 Conclusion

In this work, the correlation of component degradation and system degradation is discussed. Three types of resistor network are used for illustrations. The simulation result shows that the component degradation cannot indicate the system degradation. Instead, the correlation is related to circuit configuration and resistor ratio. For some specific configuration, the system is immune to its internal component degradation.

We also examine the correlation of component degradation and circuit degradation for circuit with feedback network. A linear-mode LED driver is used as a case study. The simulation shows that the output current of the circuit is immune to the transistor degradation due to the feedback. On the other hand, the transistor aging can lead to increase in knee point voltage. As such, the knee point voltage is a better health index to represent the circuit strength. The correlation between transistor degradation and that of the knee point is similar to the series configuration resistor network, which depends on the values of feedback resistance and the transistor on-resistance.

In short, component reliability may not be able to predict the system reliability and the system strength index should be identified before the reliability study, instead of simply believing that degraded components will lead to degraded system performance. Thus, one may invest a lot to make sure the components are very reliable, and this may not bring significant improvement (gain) in the corresponding system reliability.

References

P. D. T. O'Connor, D. Newton, and R. Bromley, *Practical Reliability Engineering*, Chichester: John Wiley & Sons Ltd, 2002.

L. Song, and T. Cher Ming, "Reliability Evaluation and Improvement for High Power LED Driver", Presented in ICMAT 2013.

H. Chenming, C. T. Simon, H. Fu-Chieh *et al.*, "Hot-Electron-Induced MOSFET Degradation - Model, Monitor, and Improvement," *Solid-State Circuits, IEEE Journal of,* vol. 20, no. 1, pp. 295–305, 1985.

J. E. Chung, P.-K. Ko, and C. Hu, "A model for hot-electron-induced MOSFET linear-current degradation based on mobility reduction due to interface-state generation," *IEEE Transactions on Electron Devices,* vol. 38, no. 6, pp. 1362–1370, 1991.

M. Rausand, and A. Høyland, *System reliability theory: models, statistical methods, and applications/ Marvin Rausand, Arnljot Høyland*: Hoboken, NJ: Wiley-Interscience, c2004. 2nd ed., 2004.

S. Lan, C. M. Tan, and K. Wu, "Methodology of reliability enhancement for high power LED driver," Microelectronics Reliability, vol. 54, no. 6–7, pp. 1150–9 , 2014.

S. Lan, C. M. Tan, and K. Wu, "Reliability study of LED driver – A case study of black box testing," *Microelectronics Reliability,* vol. 52, no. 9–10, pp. 1940–4, 2012.

D. S. Ang, and C. H. Ling, "On the time-dependent degradation of LDD n-MOSFETs under hot-carrier stress," *Microelectronics Reliability,* vol. 39, no. 9, pp. 1311–1322, 1999.

K. M. Cham, J. Hui, P. Vande Voorde *et al.*, "SELF-LIMITING BEHAVIOR OF HOT CARRIER DEGRADATION AND ITS IMPLICATION ON THE VALIDITY OF LIFETIME EXTRACTION BY ACCELERATED STRESS," *Annual Proceedings - Reliability Physics (Symposium).* pp. 191–194, 1987.

D. S. Ang, and C. H. Ling, "Unified model for the self-limiting hot-carrier degradation in LDD n-MOSFET's," *IEEE Transactions on Electron Devices,* vol. 45, no. 1, pp. 149–159, 1998.

Author Biography

Song Lan received B.E. degree from Nanyang technological University, Singapore in 2011. He obtained his Ph.D degree from Nanyang Technological University in 2015. His research interests are on the circuit reliability, device modeling, and data processing using statistical method.

Cher Ming Tan is the Director of Centre for Reliability of Sciences and Technologies (CReST) in Chang Gung University, Taiwan and also a Professor in the Electronic department under the college of Engineering. He is also a Chair Professor in Ming Chi University of Technology, Taiwan. He was a faculty member of the Nanyang Technological University, Singapore for more than 19 years before he joined the current position. He started his reliability work in 1984 with Fairchild Semiconductor where the reliability and failure analysis works were needed for the US Defense. He later pursued his master's and Ph.D. in University of Toronto, Canada. After graduation, he joined Lite-On Power Semiconductor, Taiwan, continued his work on quality, reliability, and failure analysis. He is still active in providing training and consultation to companies that need reliability worldwide. He is an Editor of IEEE Transaction on Device and Materials Reliability, Editor of Scientific Report, Chief Editor of Springer Brief on Reliability, Advisory board member of Microelectronic Reliability, and IEEE Distinguished lecturer of EDS in Reliability. He was the course coordinator of certified reliability program of Singapore Quality Institute.

Lifetime Prediction of Vehicle Components Using Online Monitoring Data

Chiharu Kumazaki, Watalu Yamamoto and Kazuyuki Suzuki

Abstract In order to set an adequate lifetime target for each market, quantitative evaluation of variation of lifetime characteristics is required. In particular, the lifetime of vehicle unit depends heavily on customer's usage (e.g., gross vehicle weight, road gradient, and acceleration operation). We thus have developed an online monitoring system that continually collects some information such as usage and environmental conditions. A method has been developed for predicting vehicle component lifetimes using data from an online monitoring system that collects an extensive amount of data during vehicle operation. The linear model used for prediction takes into account variations in usage conditions and models' data as covariates. The prediction procedure was generalized to enable it to make predictions using a new data sample. The large amount of information on usage and environmental conditions obtained with the online monitoring system enabled the usage of each sample to be quantified and treated as a stratification factor. A stratified analysis produced fairly accurate results, meaning that using online monitoring data should be useful for lifetime prediction.

Keywords Cluster analysis · Cumulative damage model · Linear regression model · Principal component analysis · *S–N* curve

C. Kumazaki (✉) · W. Yamamoto · K. Suzuki
Department of Informatics, University of Electro-Communications,
1-5-1, Chofugaoka, Chofu, Tokyo 182-8585, Japan
e-mail: chi.kumazaki@hino.co.jp

K. Suzuki
e-mail: suzuki@se.uec.ac.jp

C. Kumazaki
Department of Power Train Experiment, Hino Motors, Ltd.,
3-1-1, Hino-Dai, Hino, Tokyo 191-8660, Japan

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_23

1 Introduction

The lifetimes of vehicle components greatly depend on the usage conditions. This is particularly true for commercial vehicles as they are often operated under severe conditions. Improvement in lifetime design for each market requires an understanding of the relationship between the lifetime characteristics and usage conditions in the actual market. If it is not done correctly, we will not prevent the unexpected failure and further trouble. However, the conventional market research requires on the spot measurement of actual vehicles. Therefore, the research has been conducted with both a small number of samples and non-random sampling, and we did not completely understand the market. We thus have developed an online monitoring system that continually collects vehicle information during operation by in-vehicle communication terminal. This system enables an extensive amount of vehicle driving data, with a larger sample size and a longer observation period, to be easily collected. The data can be used to clarify the actual usage conditions.

The purpose of this study is to feed back to the appropriate lifetime design using these online monitoring data. To systemize the scheme to clarify the usage and modeling the relationship between lifetime and usage, this study continues to analyze the data. Meeker and Hong (2014) described several cases in which online monitoring data were used for reliability purposes. Suzuki and Tsubaki (2013) generalized a structure for online monitoring data and stated the importance of online monitoring based on the physical and scientific mechanism. According to this paper, to prevent the unexpected troubles, it is necessary that the failure mode is observed. If the failure mode cannot be observed directly, we may be able to predict it from an alternative characteristic value for failure. For example, the major alternative characteristic value for many vehicle components is the load frequency (torque transfer to the driveline, i.e., input torque). If a failure mechanism is revealed, the failure mode can be reliably predicted theoretically. However, in practice, even the measurement of the load frequency is difficult in many cases. This means that various usage and environmental conditions during operation are important pieces of information for describing the characteristics of failure. Use of online monitoring to collect such information facilitates the collection of a large amount of data. Moreover, identifying the relationships between a fault and these data would enable many more lifetime samples to be collected. In a case study, Kumazaki et al. (2012) predicted product lifetimes using covariates based only on usage data.

We have developed a method for predicting vehicle component lifetimes using data from an online monitoring system that collects an extensive amount of data during vehicle operation. The linear model used for prediction takes into account variations in usage conditions and models data as covariates. However, since its model represents only the average effect of covariates, there is a limit to the linear model to try to predict the overall market. Then, the prediction procedure was generalized to enable it to make predictions using a new data sample. A stratified analysis produced fairly accurate results, meaning that using online monitoring data should be useful for predicting component lifetimes.

2 Materials

2.1 Target Component

The component targeted was the transmission (*T/M*) and the focus was on gear surface failure due to pitting or spalling. The relationships among failure mode, failure mechanism, and usage/environmental conditions for the *T/M* gears are illustrated in Fig. 1.

The mechanism of gear surface failure is fatigue damage, so lifetime is generally evaluated using the cumulative damage model and the stress versus number of cycles (*S*–*N*) curve (Fig. 1b). These measures are explained in detail in JSME S 002 (1994) and Koide (2005). The vertical axis in the *S*–*N* curve graph represents input torque *S*, and the horizontal axis represents the number of gear rotations *n*.

2.2 Online Monitoring Data

The online monitoring data included the input torque $[N \cdot m]$ as the alternative characteristic value (Fig. 1e). The usage data (Fig. 1a, d) included the vehicle speed (km/h), the gross vehicle weight (GVW) (ton), the road gradient (%), the accelerator position angle (%), the engine speed (rpm), the current gear position, the number of

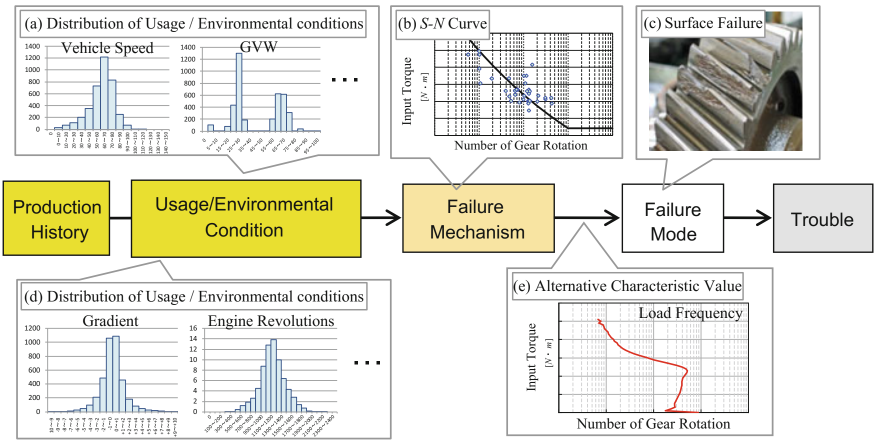


Fig. 1 Usage/environmental conditions, failure mechanism, and failure mode

upshifts, the number of downshifts, and the number of vehicle starts. These data were obtained as frequency distributions. For each one, we obtained a covariate, either the average value or the standard deviation of the distribution.

2.3 Data Source

The data were collected from 12 transport vehicles operating in a certain area. Each vehicle was operated by one of four participating transport companies (2–4 vehicles per company). The vehicles had the same type of *T/M*, and the gear to be analyzed was the same for each vehicle.

3 Lifetime Prediction Using Alternative Characteristic Value

3.1 Degree of Gear Damage and Predicted Lifetime

We first discuss the prediction of gear lifetime using the alternative characteristic value (load frequency). The load frequency distribution shown in Fig. 1c can be illustrated as shown in Fig. 2. This figure shows the number of gear rotations $\{n_1, n_2, \ldots, n_k\}$ for each torque range $\{s_1, s_2, \ldots, s_k\}$ (less than s_1 is the fatigue limit, which is not considered damage).

When the load frequency was obtained after operating t (km), the gear damage is calculated as

$$y(t) = \sum_{i=1}^{k} n_i / N_i \tag{1}$$

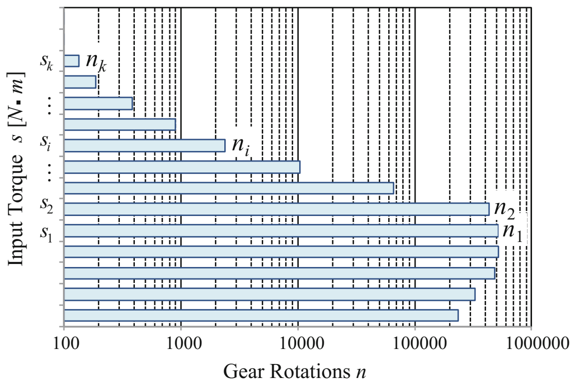


Fig. 2 Distribution of load frequency

with a cumulative damage rule. When the degree of damage reaches 1, the gear is assumed to have failed. Predicted lifetime L is calculated using

$$L = t/y(t). \tag{2}$$

3.2 Improved Prediction Accuracy Using Condition Monitoring

As mentioned in Sect. 1, in the conventional market research, on the spot measurement in actual vehicles is required. Therefore, there is a problem that measurement period cannot be obtained only a few days. We thus predicted the lifetime from 40 measurements made during operation for 1 day or 1 week. Histograms of these data are shown in Fig. 3. For the one-day measurement period, the degree of damage varied from day to day due to the effects of load capacity, availability, route, and so on for each day; so the standard deviation (sd) was quite high (7.28e+k). In contrast, for the 1-week measurement period, the sd was much smaller (1.61e+k).

Using data from the online monitoring system thus facilitated long-term measurement and enabled lifetime prediction with a small variation.

3.3 Visualization of Damage Situation Using Cumulative Graph

The cumulative graph in Fig. 4 illustrates the state of progress (accumulation) of the degree of damage. The vertical axis represents cumulative damage y for each measurement point, and the horizontal axis represents cumulative mileage t (km).

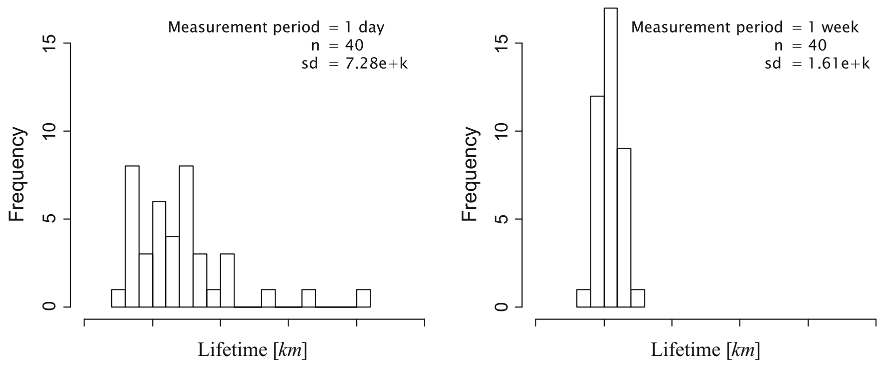


Fig. 3 Histogram of lifetime a (k: constant)

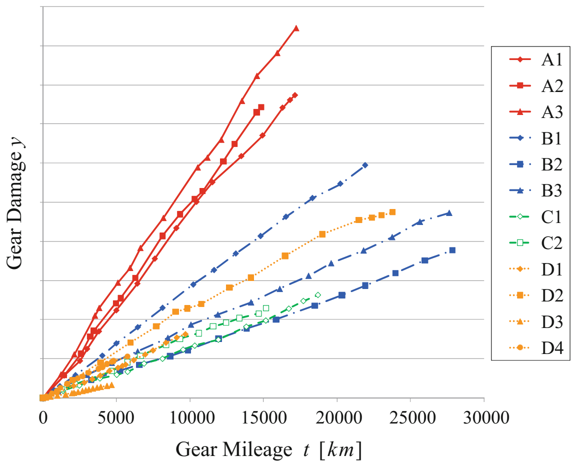


Fig. 4 Cumulative degree of gear damage

Each line represents one sample path, for example, $A1$ means vehicle 1 of company A.

A change in the long-term trend of a line would indicate a possible large change in the operating conditions, such as a change in the driver or a change in the usage environment. The lines for our samples remained fairly linear.

3.4 Lifetime Prediction Using Cumulative Graph

As mentioned in Sect. 3.1, when the degree of damage y reaches 1, the gear is assumed to have failed. This point can be easily predicted by extrapolation using linear regression:

$$y = a \cdot t + \varepsilon. \tag{3}$$

4 Lifetime Prediction Using Usage Data

4.1 Model of Degree of Damage

In the previous section, we described prediction of the degree of damage and lifetime from the load frequency as the alternative characteristic value for failure. Here we describe prediction of the degree of damage from the usage data. First, the degree of damage for measurement point j of sample i is given as

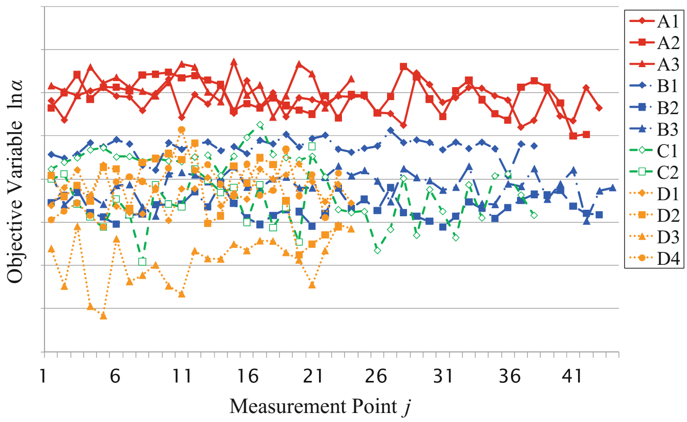


Fig. 5 Objective variable ln α

$$y(t_{ij}) = y(t_{i\cdot j-1}) + \alpha_{ij} \times \left(t_{ij} - t_{i\cdot j-1}\right), i = A1, A2, \ldots, D4, \quad j = 1, 2, \ldots, n_i \tag{4}$$

by adding an increment to the previous point $j-1$. The α_{ij} is the degree of damage per unit mileage at each measurement point. Because the degree of damage depends on the mileage, Eq. 4 isolates the effect of mileage by introducing α_{ij}, which is formulated using a linear regression model:

$$\begin{gathered} \ln \boldsymbol{\alpha}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_{ij}, \boldsymbol{\varepsilon}_{ij} \sim N\left(\mathbf{0}, \boldsymbol{\sigma}_e^2\mathbf{I}\right), \\ \mathbf{X}_{ij} = \left[\mathbf{1}, \mathbf{x}_{1ij}, \ldots, \mathbf{x}_{pij}\right], \boldsymbol{\beta}_i = \left[\beta_{0i}, \beta_{1i}, \ldots, \beta_{pi}\right]', \end{gathered} \tag{5}$$

where p is the number of covariates. In contrast, Figs. 4 and 5 shows the trend in ln α. Equation 5, which is a linear model, treats these data as objective variables.

4.2 Selection of Covariates

The selection of covariates $x = [x_1, x_2, \ldots, x_p]$ is important in formulating a linear model. There are several statistical approaches to covariate selection, such as stepwise selection. However, correlation is complicated in the analysis of online monitoring data. Therefore, creating a meaningful model using only these selection methods is difficult. We thus identified five covariates that could be used instead as shown in Table 1.

Table 1 Selected covariates and their meanings

Covariate	Meanings
Avg. GVW	Average of distribution of gross vehicle weight (GVW)
Gradient	Percentage of running time when road gradient is 3% or more
Avg. Speed	Average of distribution of vehicle speed
Avg. Accelr Pos.	Average of distribution of accelerator position angle
Start Freq.	Number of start times per km

4.2.1 Selection Process

The selection process was as follows. First, the load applied to the vehicle's driveline is considered to be the "running resistance" input from the tire. Running resistance is divided into "rolling resistance," "grade resistance," and "air resistance," which are, respectively, proportional to vehicle weight, road gradient, and vehicle speed. For a driver, the direct input device for the driveline is the accelerator pedal. Since engine torque changes with the operation of the pedal, pedal operation contributes to the load on the driveline. Also considered is the higher instantaneous load that occurs when the vehicle is started or the transmission is shifted (this load is called "shock torque").

4.2.2 Correlation of Covariates

A scatter plot matrix based on the data for sample *A*1 is shown in Fig. 6. The correlations between the covariates, which are shown in the upper right part of the figure, are reasonably small. Table 2 shows the tolerance and the variance inflation factor (VIF) calculated from the same data. These values are often used as an index of multicollinearity. Since the tolerance values are 0.1 or more and the VIF values are less than 10, there is no multicollinearity between the covariates.

4.3 Prediction Results for Degree of Damage

The data set used contained 15 measurement points for each sample. The measurements were made every week for 15 weeks (sample *C*2 was an exception with only 13 points). The results of estimation for sample *A*1 are shown in Fig. 7 and Table 3. The adjusted R-squared value (0.97) is high enough for practical purposes, and the predicted values were in good agreement with the actual values (Fig. 7).

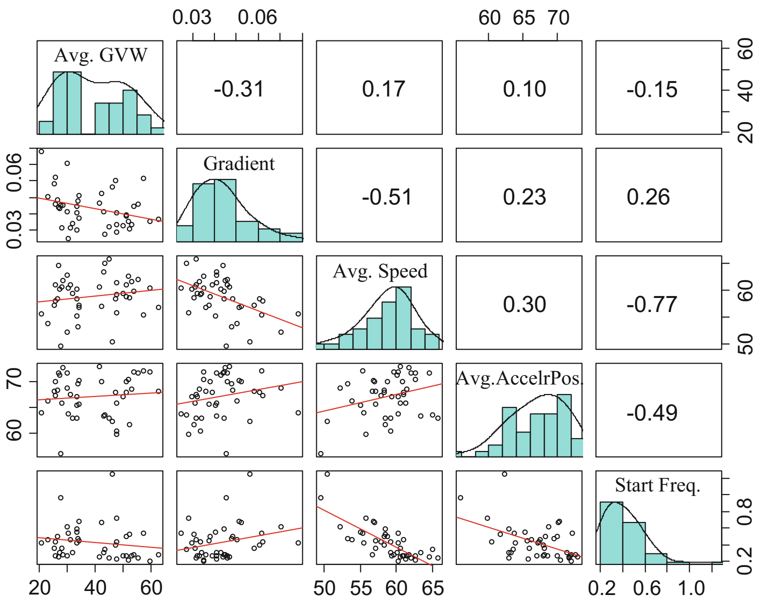


Fig. 6 Scatter plots and correlations (sample $i = A1$, $n = 40$)

Table 2 Multicollinearity (sample $i = A1$)

	Tolerance	VIF
Avg. GVW	0.207	4.839
Gradient	0.436	2.294
Avg. Speed	0.184	5.436
Avg. Accelr Pos.	0.306	3.267
Start Freq.	0.421	2.376

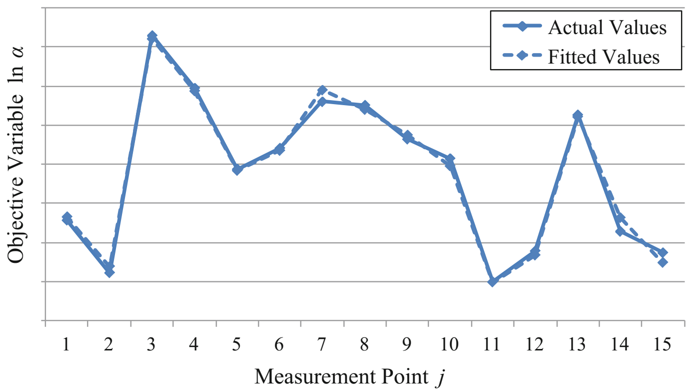


Fig. 7 Prediction results using objective variable (sample $i = A1$)

Table 3 Estimation results using Eq. 4 (sample $i = A1$)

	Coef.	*p*-value	Residual standard Err.	Adjusted R-squared
Intercept	−13.985	<0.000	0.032	0.970
Avg. GVW	0.005	0.075		
Gradient	−2.001	0.081		
Avg. Speed	−0.013	0.019		
Avg. Accelr Pos.	0.043	<0.000		
Start Freq.	0.326	0.002		

4.4 Predicted Lifetime Results

The relative error

$$\frac{|L_a - L_b|}{L_a} \times 100\,(\%) \tag{6}$$

is calculated from lifetime L_a obtained directly from the load frequency (solid line in Fig. 7) and lifetime L_b obtained from the linear model (dashed line in Fig. 7). We used the relative error as an index of prediction accuracy.

As shown in Fig. 8, the relative error was less than 1% except for sample $D3$. This means that the lifetime can be predicted from easily collectable usage data with about the same accuracy as with using the load frequency. Furthermore, the simple linear model used can be easily applied to actual product design and development.

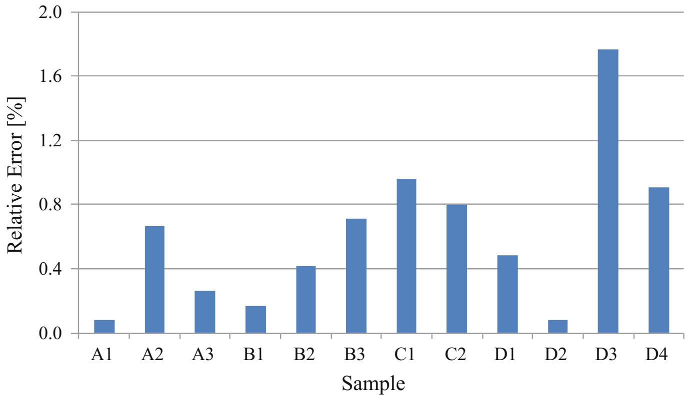


Fig. 8 Relative error in predicted lifetime (%) (Derived from Eq. 5)

5 Model Extension

5.1 Generalization of Prediction Procedure

The previous section presented the results for one sample. For the model to be practical, it should be able to handle a new sample as well. We thus applied the model to all 12 samples. Table 4 and Fig. 9 show the results of estimation using

$$\ln \boldsymbol{\alpha}_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{ij}, \quad \boldsymbol{\beta} = \left[\beta_0, \beta_1, \ldots, \beta_p\right]' \tag{7}$$

in which the effect of each covariate is estimated as a fixed effect for any sample. As shown in Fig. 9, the errors in the predicted lifetimes were much higher (that for sample $D4$ exceeded 35%). This is because covariate effects differ among subjects (drivers), operators (companies), usage conditions, and environmental conditions.

5.2 Stratified Analysis

A simple and reasonable method to use when the covariate effects differ among certain groups is stratified analysis. We performed stratified analysis using two stratification factors: operator and usage.

5.2.1 Stratification by Operator

As mentioned above, the vehicles were operated by four companies. There were various differences among the operators, including the type of cargo normally carried and management style. These differences may be reflected in the degree of gear damage. In fact, the degree of damage (see Fig. 4) tended to be similar among the vehicles of each operator. Therefore, we estimated a linear model by stratification into four groups (operator A, B, C, D).

The confidence intervals of the regression coefficients estimated for each operator's data for each covariate are shown in Fig. 10. The covariates that did not

Table 4 Estimation results using Eq. 7

	Coef.	*p*-value	Residual Standard Err.	Adjusted R-squared
Intercept	−15.453	<0.000	0.227	0.870
Avg. GVW	0.011	0.004		
Gradient	2.866	0.091		
Avg. Speed	−0.017	0.035		
Avg. Accelr Pos.	0.062	<0.000		
Start Freq.	0.119	0.352		

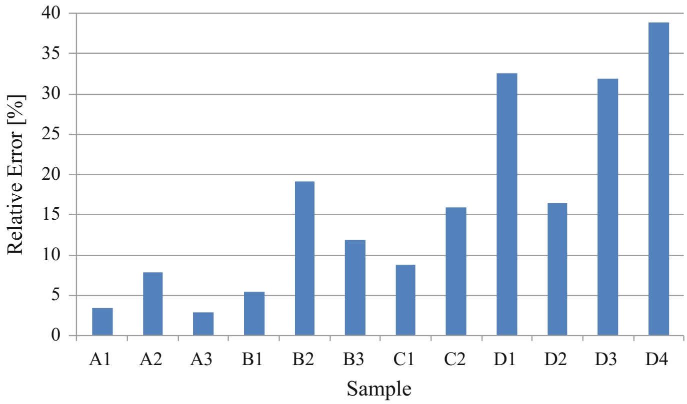


Fig. 9 Relative error in predicted lifetime (%) (Derived from Eq. 7)

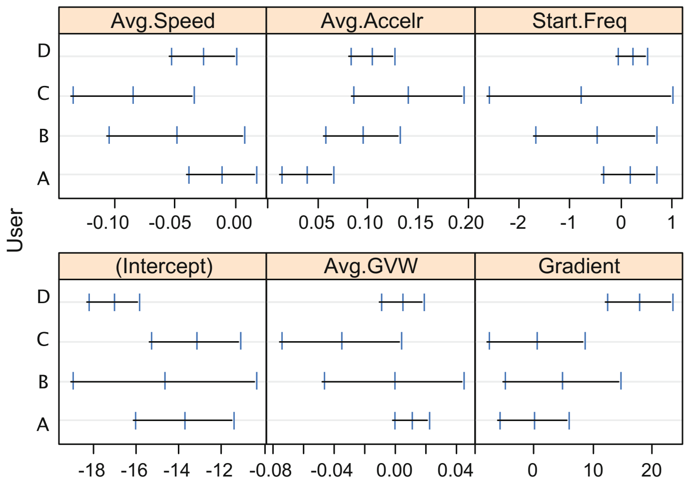


Fig. 10 Confidence interval of regression coefficients by operator

overlap the confidence interval for each group were “Intercept,” “Gradient,” and “Avg. Accelr Pos.” These results show that the covariate effects differed among operators.

5.2.2 Stratification by Usage

Stratification by operator is a simpler approach, but the model is based on only four operators and is thus not yet practical. Another possible approach is stratification based on differences in usage. Table 5 shows the results of principal component analysis on the covariates.

This analysis used ten variables that were chosen by eliminating the variables having high mutual correlation. The data used were the same as in the previous analysis, so each sample had 15 measurement points (except for one). Looking at the cumulative proportion, we see that the first to third principal components describes 80% of the data. The principal component scores and factor loadings are shown in Fig. 11. The chart on the left is for the first and second PCs, the one on the right is for the third and second PCs.

In the figure, the plot points are the principal component scores for the data points. The representations for the 12 types of samples are shown in the figure legend.

Table 5 Results of principal component analysis (PC: principal component)

No		PC 1	PC 2	PC 3	PC 4	PC 5
	Eigenvalue	4.005	3.043	1.211	0.854	0.423
	Proportion	0.400	0.304	0.121	0.085	0.042
	Cumulative proportion	**0.400**	**0.705**	**0.826**	**0.911**	**0.954**
1	Avg. GVW	0.816	−0.274	−0.117	−0.128	0.456
2	Gradient	0.028	0.149	0.855	0.476	0.047
3	Avg. Speed	0.607	0.765	−0.073	−0.052	−0.115
4	Std. dev. Speed	0.785	−0.315	0.109	−0.295	−0.369
5	Avg. Engine RPM	0.908	0.139	−0.030	0.280	−0.153
6	Std. dev. Engine RPM	0.656	−0.548	-0.259	0.364	−0.089
7	Avg. Accelr Pos.	0.914	0.197	0.154	−0.213	0.148
8	9th Gear Freq.	−0.223	0.867	0.132	−0.336	−0.051
9	Shift Freq.	0.170	−0.659	0.582	−0.375	−0.004
10	Start Freq.	−0.427	−0.845	−0.039	−0.104	−0.088

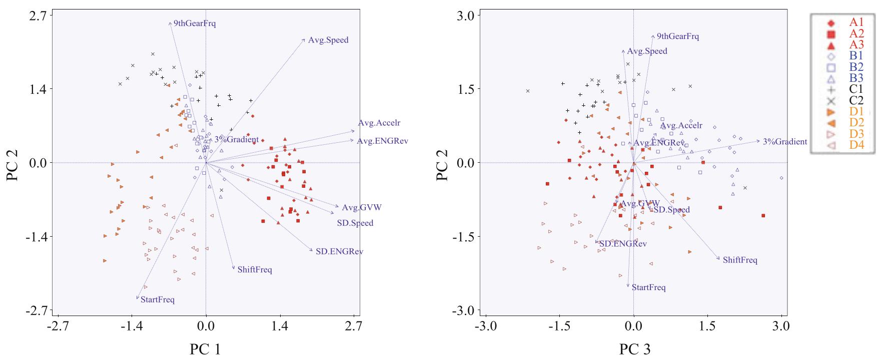


Fig. 11 Principal component scores (*plot points*) and factor loadings (*arrows*)

Focusing on the factor loadings (indicated by the arrows), we see that GVW, accelerator position angle, and engine RPM contribute to the first PC, so they reflect the load capacity. The ninth gear frequency, vehicle speed, shift frequency, and start frequency contribute to the second PC, so they reflect the characteristics of the route (e.g., highway driving vs. local road driving). Finally, the gradient contributes to the third PC, so it reflects the road gradient of the route. Looking at the principal component scores, we see that each sample's plot includes a variation of measurement point.

The PCA results indicated differences in-vehicle usage. Those of operator *A* tended to carry a heavier load. Those of operator *D*, except for vehicle *D*2, tended to travel flat local roads. Among the vehicles of operator *B*, only *B*1 had slightly different characteristics.

We used the Ward method and the principal component score to perform cluster analysis. The dendrogram shown in Fig. 12 was obtained from hierarchical cluster analysis. The vertical axis represents the dissimilarity, and the horizontal axis represents each data point (labels omitted due to large number of data points).

Neighboring data points were clustered into groups. The tree was cut at the level of the red dashed line, resulting in five clusters (I, II, III, IV, and V). Table 6 shows the results of cluster assignment.

The clusters assigned on the basis of highest frequency are shown in Table 7.

Each sample was stratified on the basis of usage in accordance with the clustering results. The clusters are the stratification factors. The same as Figs. 12 and 13 show the confidence intervals of the regression coefficients estimated for each cluster. All the coefficient confidence intervals (except "Start Frequency") did not overlap for each group, indicating that the effect of the covariates differed among clusters.

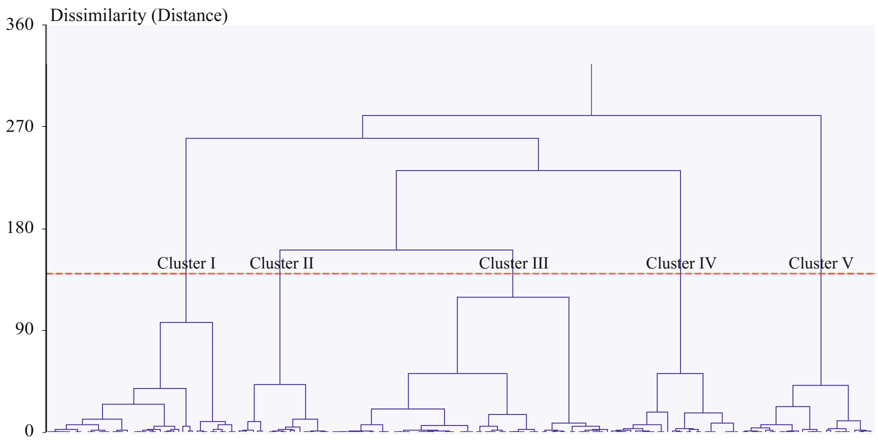


Fig. 12 Results of cluster analysis (*Dendrogram*)

Table 6 Results of cluster assignment

Sample	Size	Cluster				
		I	II	III	IV	V
*A*1	15	14	1	0	0	0
*A*2	15	10	5	0	0	0
*A*3	15	15	0	0	0	0
*B*1	15	0	13	1	1	0
*B*2	15	1	0	14	0	0
*B*3	15	0	0	15	0	0
*C*1	15	1	0	0	14	0
*C*2	13	0	1	0	12	0
*D*1	15	0	0	15	0	0
*D*2	15	0	0	15	0	0
*D*3	15	0	0	0	0	15
*D*4	15	0	0	0	1	14

Table 7 Results of cluster assignment based on highest frequency

Sample	*A*1	*A*2	*A*3	*B*1	*B*2	*B*3	*C*1	*C*2	*D*1	*D*2	*D*3	*D*4
Cluster	I	I	I	II	III	III	IV	IV	III	III	V	V

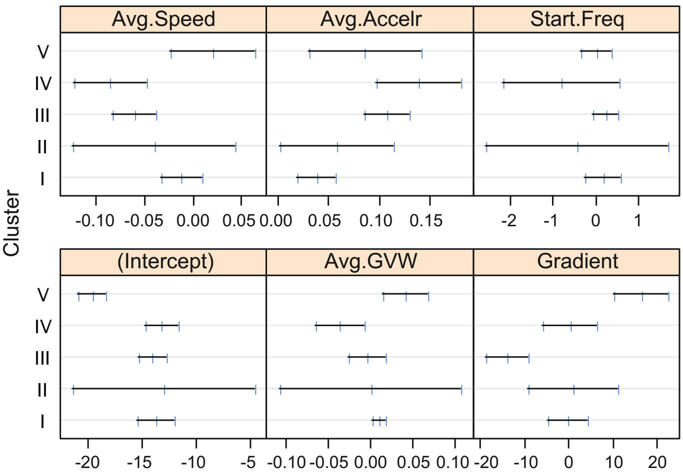


Fig. 13 Confidence interval of regression coefficients of each cluster

5.2.3 Prediction Results for Stratification

Th lifetimes predicted on the basis of the stratified analysis are shown in Fig. 14.

In the figure, "Not stratified" means the results of using the generalization prediction procedure (Fig. 9). "Stratified 1" means the results of stratification by

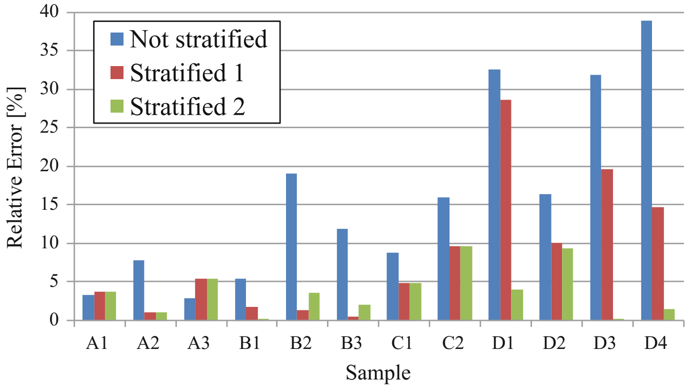


Fig. 14 Relative error of prediction of lifetime (%) (derived from stratified analysis)

operator (Sect. 5.2.1). It reduced the overall prediction error. "Stratified 2" means the results of stratification by usage (Sect. 5.2.2). It greatly reduced the overall prediction error. In particular, the most effective action is to stratify the vehicles of operator *D* into two groups. In short, stratification based on usage improves prediction accuracy.

The proposed method performs clustering and principal component analysis using covariates and predicts the degree of damage and lifetime from covariates. Application of this method, even to a new sample (even if only covariates), would enable groups to be determined and a prediction model to be fitted.

6 Summary

Our proposed method for predicting vehicle component lifetimes using an extensive amount of data collected during vehicle operation takes into account variations in usage conditions and models data as covariates. The generalized version is effective even for a new data sample. Our stratified analysis produced fairly accurate results and demonstrated that using online monitoring data is useful for predicting the lifetimes of vehicle components. The purpose of this system is to perform a reliable market evaluation for setting an appropriate target value at the design and development stage. To do this, the lifetime data will be required more and more. Therefore, more practical and effective operation of this system is required. Since the flow of application of this system could be shown, we will be able to develop to the actual business as the next step.

The validity of our sampling is problematic because only 12 vehicles were included in the sample. Sampling based on an experimental design is necessary. As mentioned in this study, market research using principal component analysis can take advantage about it. To address the sampling size problem, we will also consider the production history of the target component (as represented in Fig. 1).

References

JSME S 002 (1994): *Standard Method of Statistical Fatigue Testing (Revised Edition)*, The Japan Society of Mechanical Engineers (in Japanese).

Koide, T. (2005): "Estimation of life and strength by fatigue test," *Hagurumasouchi Ni Kansuru Jitsuyogijyutukakushin No Tameno Chosakenkyu Bunkakai (RC205 Report of Research Subcommittee for Practical Innovation on Gearing)*, The Japan Society of Mechanical Engineers, pp. 21–29 (in Japanese).

Kumazaki, C., M. Yokoyama, W. Yamamoto, and K. Suzuki (2012): "Optimal sampling plan and lifetime prediction based on online information," *Shinraisei To Shisutemu Anzengaku (Proceeding of 16th Study of System Safety and Reliability)*, Graduate School of Information Systems, University of Electro-Communications, pp. 46–51 (in Japanese).

Meeker, W. Q. and Y. Hong (2014): "Reliability Meets Big Data: Opportunities and Challenges," *Quality Engineering*, Vol. 26, No. 1, pp. 102–116.

Suzuki, K. and H. Tsubaki (2013): "Use of information of online monitoring and the entire scheme to prevent trouble," *Shinraisei To Shisutemu Anzengaku (Proceeding of 17th Study of System Safety and Reliability)*, Graduate School of Information Systems, University of Electro-Communications, pp. 38–45 (in Japanese).

Estimation of Lifetime Distribution with Covariates Using Online Monitoring

Masahiro Yokoyama, Watalu Yamamoto and Kazuyuki Suzuki

Abstract Online monitoring becomes a common tool for keeping high reliability of products and systems in these days. The information to maintain the target product includes usage history, system conditions, and environmental conditions. Nowadays they are monitored and reported in real time and stored as big data. On modeling the failure mechanisms statistically, these variables are primary candidates for covariates which affect the failure mechanism. There is literature on modeling the lifetime of products with the accelerated lifetime model on an accumulation of the covariate effects, which is a nonlinear transformation of the observed lifetime. The existing literature requires a parametric form of the transformed lifetime distribution in advance. However, such knowledge may be difficult to acquire in advance in some cases. When we assume an incorrect distribution, it is called misspecification. This paper proposes a strategy to use the log-normal likelihood for the estimation of the parameters of the covariate effects when the underlying distribution is either the Weibull or log-normal distribution. It is derived that the score function of the log-normal likelihood is identified as an approximation for the Weibull cases. The simulation study shows the relationship among the sample distribution of parameter estimates and underlying distributions. For the Weibull cases, if the shape parameter is large, the bias of the resulting estimates is small.

Keywords Misspecification · Weibull distribution · Log-normal distribution · Accelerated lifetime model · Covariate

M. Yokoyama (✉)
JSQC, Polytechnic University, Kodaira, Tokyo, Japan
e-mail: m-yokoyama@uitec.ac.jp

W. Yamamoto · K. Suzuki
JSQC, University of Electro-Communications, Chofu, Tokyo, Japan

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_24

1 Introduction

Online monitoring in real time using the Internet and Global Positioning System (GPS) has become a common tool for keeping products and systems operating reliably. The information gathered includes usage history, system conditions, and environmental conditions and is stored as big data. In the statistical modeling of failure mechanisms, variables contained in this information are primary candidates for covariates that affect failure mechanisms.

Information on the covariates can be used to improve product reliability and safety. For example, in the analysis of the lifetime distribution of a truck engine, the lifetime is usually estimated on the basis of a distance traveled (mileage) before catastrophic failure. However, other variables, such as an average slope of routes traveled and an average weight of loads carried affects the load on the engine. Therefore, even if the same amount of mileage, the lifetime is different. Thus, the mileage is not always sufficient information for optimizing inspection and maintenance times.

2 Previous Studies and Proposal

Since covariates cannot be treated as a timescale directly, they need to be incorporated into a model. Hong and Meeker (2010, 2013) provided a general method that utilizes dynamic covariate information to predict field-failure. In this study, the cumulative exposure model was used to describe the effect of a dynamic covariate on the failure-time distribution. Nelson (2001) describe the cumulative exposure model as represented in Eq. 1 on the basis of the accelerated lifetime model (Nelson 1972).

$$T^*(\boldsymbol{\beta}) = \int_0^T \exp[\boldsymbol{\beta}\mathbf{z}(s)]ds, \tag{1}$$

where $\boldsymbol{\beta} = [\beta_1 \quad \beta_2 \quad \cdots \quad \beta_Q]$ is a vector of each covariate parameter ($\boldsymbol{\beta}^*$ is the true value of $\boldsymbol{\beta}$). This equation represents a nonlinear transformation of the lifetime T to the quantity $T^*(\boldsymbol{\beta})$, which incorporates covariate information. The values of the covariates are either discrete or continuous. Meeker and Escobar (1998) called this equation the “proportional quantities (PQ) model” or the “scale accelerated failure-time (SAFT) model.”

Let $\mathbf{z}(t)$ be the covariate vector acquired at continuous time point t and Q be the total number of the covariates.

$$\mathbf{z}(t) = \begin{bmatrix} z_1(t) \\ z_2(t) \\ \vdots \\ z_Q(t) \end{bmatrix}$$

In estimating a distribution of the lifetime T, it is assumed that T has a parametric distribution. In the approach used by Hong and Meeker (2010, 2013), the parametric form of the lifetime distribution is needed in advance. However, such information may be difficult to acquire in advance. The assumption of an incorrect distribution is called "misspecification."

We propose using the log-normal likelihood to estimate the parameters of the covariate effects when the underlying distribution is either the Weibull or log-normal distribution. With this approach, a model can be built that incorporates the covariates even if the underlying distribution of $T^*(\boldsymbol{\beta}^*)$ which is the Weibull distribution is misspecified as the log-normal one. This enables the use of the predicted lifetimes of individual user's products using the observed covariates.

In Sect. 3, the derived score function of the log-normal distribution is shown to approximate that of the Weibull distribution. The simulation study described in Sect. 4 clarified the relationships among the sample distribution of parameter estimates and the underlying distributions. For the case of the Weibull distributions, if the shape parameter is large, the bias of the resulting estimates is small.

3 Analytical Study Using Likelihood Function

3.1 Covariate

Here we deal with the case in which the target covariates are observed at time $t_j^O (j = 1, \ldots, J)$ for a discrete interval (Fig. 1). The covariate value is assumed to be constant between each pair of t_j^O. This represents the case in which the truck carries a particular weight during each interval.

The symbols here used are defined as follows.

The value of the covariate can be observed at the failure point, so $T = t_J^O$.

The nonlinear transformation of lifetime T to quantity $T^*(\boldsymbol{\beta})$ on the basis (1) is illustrated in Fig. 2.

Let i be the product number of each user, T_i be the lifetime of each product, $\mathbf{z}(t_{ij}^O)$ be a vector of the value of the observed covariates at t_{ij}^O for each product, and $T_i^*(\boldsymbol{\beta})$ be the amount of the transformed lifetime for each product.

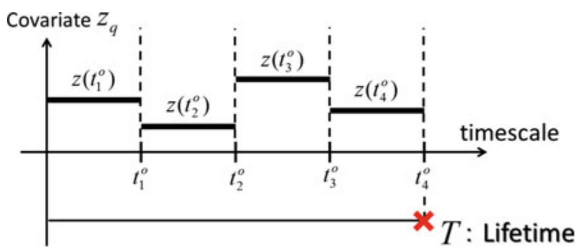


Fig. 1 Covariate z_q observed at discrete intervals ($J = 4$)

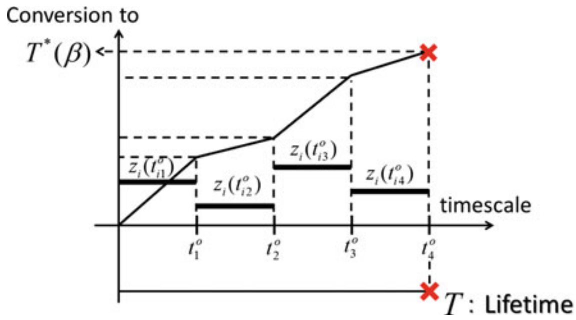


Fig. 2 Transformation to $T^*(\boldsymbol{\beta})$

3.2 Likelihood Function and Score Function

We define μ as the average and σ as the standard deviation of the log-transformed log-normal distribution.

The likelihood function for the log-normal distribution is given by

$$L_{\text{LN}}(\boldsymbol{\beta},\mu,\sigma)=\prod_{i=1}^{n}\left\{\exp[\boldsymbol{\beta}\mathbf{z}(t^o_{i,B_i})]\cdot\frac{1}{\sqrt{2\pi}\sigma T_i^*(\boldsymbol{\beta})}\cdot\exp\left[-\frac{(\log T_i^*(\boldsymbol{\beta})-\mu)^2}{2\sigma^2}\right]\right\}. \quad (2)$$

The score function obtained by differentiating this function with respect to $\boldsymbol{\beta}$ is given by

$$g_{\text{LN}}(\boldsymbol{\beta})=\sum_{i=1}^{n}(\mathbf{z}(t^o_{i,B_i}))-\sum_{i=1}^{n}\frac{T_i^{*\prime}(\boldsymbol{\beta})}{T_i^*(\boldsymbol{\beta})}-\frac{1}{\hat{\sigma}}\cdot\sum_{i=1}^{n}\left\{\left(\frac{\log(T_i^*(\boldsymbol{\beta}))-\hat{\mu}}{\hat{\sigma}}\right)\cdot\frac{T_i^{*\prime}(\boldsymbol{\beta})}{T_i^*(\boldsymbol{\beta})}\right\}. \quad (3)$$

In the maximum likelihood estimation (MLE) for $T^*(\boldsymbol{\beta}^*)$ assuming the log-normal distribution, the MLE is given by the estimation of $\boldsymbol{\beta}$ in which the score function is 0.

The log-likelihood function for the Weibull distribution (log-location-scale) is given by

$$\begin{aligned}\log L_{\text{WEI}}(\boldsymbol{\beta})=&\sum_{i=1}^{n}\left(\boldsymbol{\beta}\mathbf{z}(t^o_{i,B_i})\right)-\sum_{i=1}^{n}\log(\hat{\sigma})-\sum_{i=1}^{n}\log(T_i^*(\boldsymbol{\beta}))\\&+\sum_{i=1}^{n}\left\{\frac{\log(T_i^*(\boldsymbol{\beta}))-\hat{\mu}}{\hat{\sigma}}-\exp\left(\frac{\log(T_i^*(\boldsymbol{\beta}))-\hat{\mu}}{\hat{\sigma}}\right)\right\}.\end{aligned} \quad (4)$$

The score function obtained by differentiating this function with respect to $\boldsymbol{\beta}$ is given by

$$g_{\text{WEI}}(\boldsymbol{\beta})=\sum_{i=1}^{n}(\mathbf{z}(t^o_{i,B_i}))-\sum_{i=1}^{n}\frac{T_i^{*\prime}(\boldsymbol{\beta})}{T_i^*(\boldsymbol{\beta})}-\frac{1}{\hat{\sigma}}\cdot\sum_{i=1}^{n}\left\{\left(\exp(\frac{\log(T_i^*(\boldsymbol{\beta}))-\hat{\mu}}{\hat{\sigma}})-1\right)\cdot\frac{T_i^{*\prime}(\boldsymbol{\beta})}{T_i^*(\boldsymbol{\beta})}\right\}. \quad (5)$$

Let $\exp\left(\frac{\log(T_i^*(\boldsymbol{\beta}))-\hat{\mu}}{\hat{\sigma}}\right)$ be $\exp(A)$. By Taylor expansion around $A = 0$, we get $\exp(A) \approx \exp(0) + \exp(0) \cdot A = 1 + A$. Therefore, the first-order approximation of (5) is given by

$$g_{\mathrm{WEI}}(\boldsymbol{\beta}) \approx \sum_{i=1}^{n} (\mathbf{z}(t_{i,B_i}^{o})) - \sum_{i=1}^{n} \frac{T_i^{*\prime}(\boldsymbol{\beta})}{T_i^*(\boldsymbol{\beta})} - \frac{1}{\hat{\sigma}} \cdot \sum_{i=1}^{n} \left\{ \left(\frac{\log(T_i^*(\boldsymbol{\beta})) - \hat{\mu}}{\hat{\sigma}} \right) \cdot \frac{T_i^{*\prime}(\boldsymbol{\beta})}{T_i^*(\boldsymbol{\beta})} \right\}$$
$$= g_{\mathrm{LN}}(\boldsymbol{\beta}). \tag{6}$$

By using this approximate relationship, we can approximately estimate $\boldsymbol{\beta}$ using MLE assuming the log-normal distribution, even if $T^*(\boldsymbol{\beta}^*)$ follows the Weibull distribution.

4 Simulation

4.1 Method

In the simulation, $\boldsymbol{\beta}$ was estimated using MLE with $T_i^*(\boldsymbol{\beta}^*)$ assumed to follow the log-normal distribution. The covariate value $\mathbf{z}(t_{ij}^{O})$ for each time point and the lifetime T_i were used to estimate $\boldsymbol{\beta}$. The simulation data were generated as explained below.

1. $T_i^*(\boldsymbol{\beta}^*)$

The values of $T_i^*(\boldsymbol{\beta}^*)$ were generated in accordance with the Weibull distribution defined in Sect. 4.2.2.

2. $\mathbf{z}(t_{ij}^{O})$

In this simulation, the generated values of $\mathbf{z}(t_{ij}^{O})$, which take discrete value, are remained constant during each observation. We assumed two kinds of covariates: $z_1(t_{ij}^{O})$ were generated using a normal distribution with a mean of 1.0 and a standard deviation of 0.1; $z_2(t_{ij}^{O})$ were generated independently for each product for each observation time. The true values of parameter vector $\boldsymbol{\beta}$ representing the magnitude of the effects of the covariates were defined as 1.0 for both types; i.e., $\beta_1^* = \beta_2^* = 1.00$.

3. T_i

The values of T_i were determined from $T_i^*(\boldsymbol{\beta}^*)$ and $\mathbf{z}(t_{ij}^{O})$ generated as described above. The interval between each pair of covariate observations was determined as follows.

$t_{i1}^{O} = (t_{i2}^{O} - t_{i1}^{O}) = \cdots = (t_{i,J_i-1}^{O} - t_{i,J_i-2}^{O}) = 1$ (Interval is constant.)

$(t^O_{i,J_i} - t^O_{i,J_i-1}) \leq 1$ (Interval is less than 1 only for interval before last observation.)

Thus,

$$T^*_i(\boldsymbol{\beta}^*) = \sum_{j=1}^{J_i-1} \exp[\beta^*_1 \cdot z_1(t^O_{ij}) + \beta^*_2 \cdot z_2(t^O_{ij})] + \exp[\beta^*_1 \cdot z_1(t^O_{i,J_i}) + \beta^*_2 \cdot z_2(t^O_{i,J_i})] \cdot (t^O_{i,J_i} - t^O_{i,J_i-1}).$$

Therefore, the lifetime T_i for each product was obtained using t^O_{i,J_i}.

4.1.1 Comparison of Estimations with Coefficient of Variation

Kordonsky and Gertsbakh (1993) proposed using the time scale that minimizes the coefficient of variation (c.v.). The method using this approach does not depend on the type of distribution.

$$\text{c.v.}(T^*(\boldsymbol{\beta})) = \frac{\sqrt{\text{var}(T^*(\boldsymbol{\beta}))}}{\text{mean}(T^*(\boldsymbol{\beta}))},$$

where $\text{var}(T^*(\boldsymbol{\beta}))$ is the sample variance of $T^*(\boldsymbol{\beta})$ and $\text{mean}(T^*(\boldsymbol{\beta}))$ is the sample mean of $T^*(\boldsymbol{\beta})$. We compared this method with our method.

4.2 Simulation Results

4.2.1 Maximum Likelihood Estimate of $T^*_i(\boldsymbol{\beta}^*)$ that Follows a Log-Normal Distribution

First we show that $T^*_i(\boldsymbol{\beta}^*)$ can be correctly estimated using MLE assuming the log-normal distribution for $T^*_i(\boldsymbol{\beta}^*)$ generated in accordance with the log-normal distribution. Table 1 shows the estimations for $\beta^*_1 = \beta^*_2 = 1.00$. The parameter μ represents the set value of the average of the log-transformed log-normal distribution, and σ represents the set value of the standard deviation of the log-transformed log-normal distribution. The estimation of $\boldsymbol{\beta}$ using data from 1,000 products was repeated 500 times. The results shown in the upper rows are the average estimated values, and those shown in the lower rows are the standard deviations.

Table 1 Estimations obtained for $T_i^*(\boldsymbol{\beta}^*)$ when $T_i^*(\boldsymbol{\beta}^*)$ was generated using the log-normal distribution ($\beta_1^* = \beta_2^* = 1.00$, $\mu = 2.0$, $n = 1,000$, no. of repetitions = 500)

Set value		Proposed method		Minimizing coefficient of variation	
μ	σ	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
2.0	1	0.991	0.997	0.165	0.119
		0.301	0.286	1.305	1.489
2.0	0.7	0.994	0.991	0.619	0.605
		0.209	0.222	0.489	0.493
2.0	0.5	1.001	0.998	0.801	0.795
		0.168	0.163	0.244	0.053
2.0	0.1	1.001	0.999	0.979	0.977
		0.032	0.035	0.032	0.035
2.0	0.01	1.000	1.000	1.000	1.000
		0.003	0.004	0.003	0.004

Upper rows contain estimate means, and lower rows contain standard deviations

These results show that the value of $\boldsymbol{\beta}$ was correctly estimated by the MLE.

4.2.2 Maximum Likelihood Estimate of $T_i^*(\boldsymbol{\beta}^*)$ that Follows a Weibull Distribution

Here we show that $\boldsymbol{\beta}$ can be approximately estimated using MLE assuming the log-normal distribution for $T_i^*(\boldsymbol{\beta}^*)$ generated in accordance with the Weibull distribution. That is, $\boldsymbol{\beta}$ is estimated when the Weibull distribution is misspecified as the log-normal distribution.

The true value of $\boldsymbol{\beta}$ was set to 1.00, i.e., $\beta_1^* = \beta_2^* = 1.00$. The estimation of $\boldsymbol{\beta}$ was repeated 1,000 times. Table 2 shows the average values and the standard deviations. The parameter n represents the number of samples, m represents the shape parameter, and η represents the scale parameter of the Weibull distribution. Figure 3 graphically represents the estimates of β_1 for $\eta = 100$ in Table 2.

The results shown in Table 2 and Fig. 3 for the misspecified distribution assumptions show that parameter $\boldsymbol{\beta}$ can be approximately obtained. When the value of the shape parameter m of the Weibull distribution was large, the bias of the resulting estimates of $\boldsymbol{\beta}$ was small.

Next, we investigated the variation in the estimated results. Figure 4 shows the box plot (for $\eta = 100$, $n = 10{,}000$) of estimation result β_1 in Table 2. The greater the value of the shape parameter m, the smaller the variation in the estimate.

Table 2 Estimations obtained for $\boldsymbol{\beta}$ when $T_i^*(\boldsymbol{\beta}^*)$ was generated using the Weibull distribution ($\beta_1^* = \beta_2^* = 1.00$, no. of repetitions = 1,000)

$m = 1$	$\eta = 50$			$\eta = 100$		
$\hat{\beta}_1$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.974	0.960	0.987	0.957	0.958	0.955
SD	0.322	0.138	0.085	0.320	0.140	0.099
$\hat{\beta}_2$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.945	0.957	0.987	0.947	0.963	0.953
SD	0.326	0.138	0.092	0.310	0.143	0.099
$m = 2$	$\eta = 50$			$\eta = 100$		
$\hat{\beta}_1$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.951	0.961	0.966	0.983	0.974	0.975
SD	0.282	0.132	0.085	0.301	0.133	0.093
$\hat{\beta}_2$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.958	0.957	0.965	0.972	0.979	0.969
SD	0.275	0.127	0.091	0.301	0.135	0.094
$m = 3$	$\eta = 50$			$\eta = 100$		
$\hat{\beta}_1$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.978	0.980	0.984	0.987	0.983	0.987
SD	0.245	0.108	0.078	0.285	0.125	0.092
$\hat{\beta}_2$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.977	0.983	0.987	1.001	0.987	0.988
SD	0.249	0.111	0.080	0.275	0.122	0.091
$m = 4$	$\eta = 50$			$\eta = 100$		
$\hat{\beta}_1$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.992	0.987	0.991	0.997	0.995	0.994
SD	0.221	0.098	0.069	0.258	0.113	0.081
$\hat{\beta}_2$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	1.011	0.989	0.993	1.001	0.995	0.993
SD	0.215	0.096	0.066	0.256	0.115	0.079
$m = 5$	$\eta = 50$			$\eta = 100$		
$\hat{\beta}_1$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5000	n = 10,000
mean	0.996	0.995	0.994	0.997	0.996	0.996
SD	0.188	0.083	0.059	0.232	0.102	0.073
$\hat{\beta}_2$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.998	0.998	0.997	0.997	0.995	0.995
SD	0.185	0.081	0.062	0.228	0.104	0.071
$m = 6$	$\eta = 50$			$\eta = 100$		
$\hat{\beta}_1$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.995	1.000	1.000	0.999	0.999	0.998
SD	0.159	0.074	0.050	0.202	0.093	0.063
$\hat{\beta}_2$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.997	0.998	1.000	1.004	0.994	1.000
SD	0.157	0.074	0.052	0.208	0.089	0.066

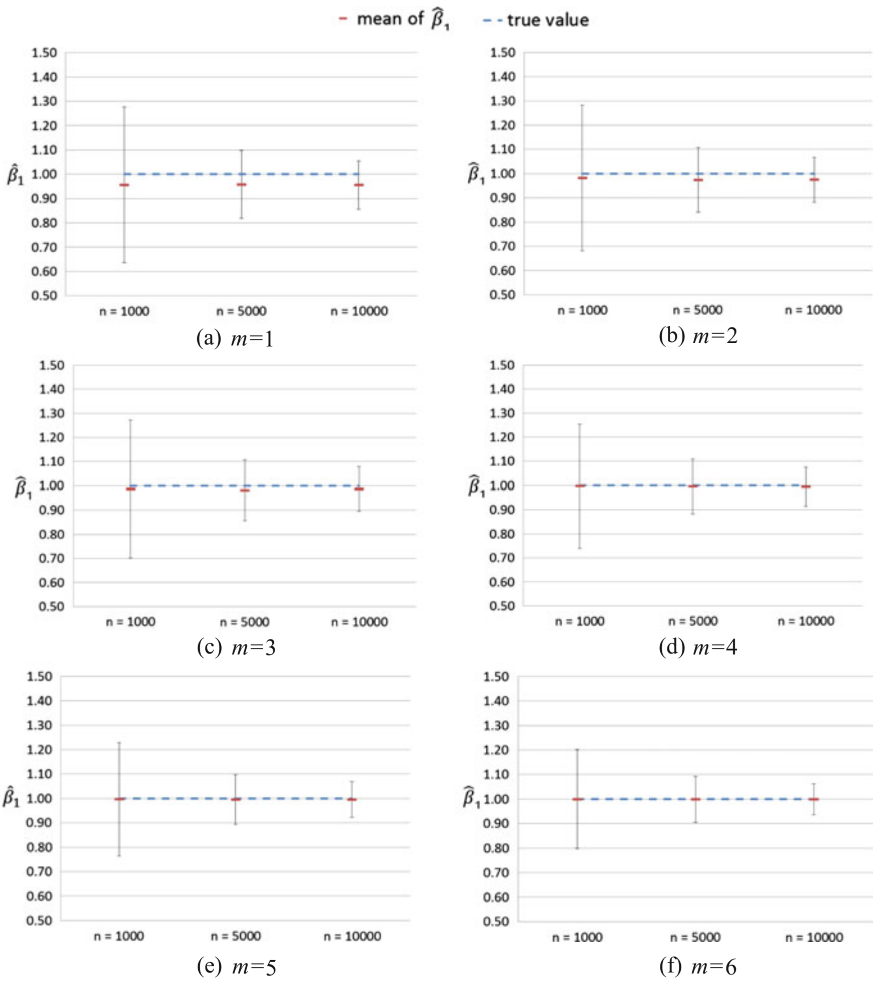


Fig. 3 Estimation results for β_1 for ($\eta = 100$, $n = 10{,}000$) case in Table 2 ($\beta_1^* = \beta_2^* = 1.00$, no. of repetitions = 1,000). (Center of plot represents mean value of 1,000 repetitions; *upper* and *lower* lines represent mean value ± standard deviation)

Table 3 shows the estimate results for the coefficient of variation minimization method. Estimation accuracy was good when the value of the shape parameter m of the Weibull distribution was large. However, the results of this method were less accurate than that of the MLE method assuming the log-normal distribution i.e., the misspecified distribution.

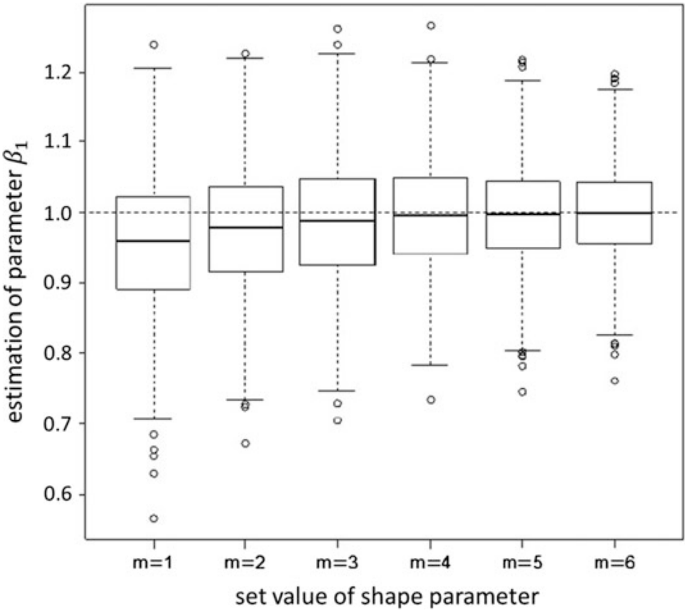


Fig. 4 Box plot of estimation result β_1 for ($\eta = 100$, $n = 10{,}000$) case in Table 2 ($\beta_1^* = \beta_2^* = 1.00$, no. of repetitions = 1,000)

Table 3 Result of estimation of $\boldsymbol{\beta}$ by minimizing of coefficient of variation for $T_i^*(\boldsymbol{\beta}^*)$ that follows the Weibull distribution ($\beta_1^* = \beta_2^* = 1.00$, no. of repetitions = 1,000)

m = 1	η = 50			η = 100		
$\hat{\beta}_1$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.511	0.517	0.504	0.466	0.487	0.530
SD	1.180	0.506	0.347	1.631	0.730	0.530
$\hat{\beta}_2$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.450	0.550	0.523	0.500	0.502	0.492
SD	1.129	0.524	0.361	1.624	0.732	0.522
m = 2	η = 50			η = 100		
$\hat{\beta}_1$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.856	0.858	0.864	0.876	0.874	0.868
SD	0.401	0.186	0.131	0.570	0.260	0.181
$\hat{\beta}_2$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.884	0.857	0.865	0.852	0.872	0.862
SD	0.422	0.183	0.000	0.579	0.267	0.183
m = 3	η = 50			η = 100		
$\hat{\beta}_1$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000

(continued)

Table 3 (continued)

m = 1	η = 50			η = 100		
mean	0.923	0.931	0.931	0.928	0.931	0.934
SD	0.274	0.122	0.083	0.390	0.166	0.124
$\hat{\beta}_2$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.933	0.933	0.934	0.958	0.937	0.934
SD	0.275	0.124	0.089	0.362	0.170	0.124
m = 4	η = 50			η = 100		
$\hat{\beta}_1$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.954	0.958	0.957	0.965	0.966	0.965
SD	0.213	0.094	0.067	0.288	0.130	0.095
$\hat{\beta}_2$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.972	0.961	0.960	0.968	0.962	0.955
SD	0.204	0.091	0.066	0.305	0.132	0.091
m = 5	η = 50			η = 100		
$\hat{\beta}_1$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.970	0.974	0.972	0.990	0.974	0.977
SD	0.171	0.074	0.054	0.242	0.105	0.080
$\hat{\beta}_2$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.974	0.976	0.973	0.984	0.970	0.976
SD	0.173	0.075	0.057	0.241	0.108	0.076
m = 6	η = 50			η = 100		
$\hat{\beta}_1$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.982	0.982	0.983	0.980	0.980	0.981
SD	0.142	0.066	0.045	0.199	0.090	0.062
$\hat{\beta}_2$	n = 1,000	n = 5,000	n = 10,000	n = 1,000	n = 5,000	n = 10,000
mean	0.979	0.980	0.984	0.990	0.974	0.982
SD	0.137	0.066	0.046	0.196	0.089	0.065

5 Conclusion and Future Study

This study has shown that the parameter $\boldsymbol{\beta}$ can be approximately estimated by the MLE assuming the log-normal distribution for the lifetime $T_i^*(\boldsymbol{\beta}^*)$ even if $T_i^*(\boldsymbol{\beta}^*)$ follows the Weibull distribution. This means that a model incorporating covariates observed online can be built regardless of whether the distribution of $T_i^*(\boldsymbol{\beta}^*)$ is actually the Weibull or log-normal distribution.

Future work includes finding the ways to improve reliability by using online condition monitoring data and the proposed method. We will also analytically investigate the relationship between the accuracy of the approximation and the shape parameter m in the Weibull distribution. Similar model building for other distributions would enable online condition monitoring data to be more widely applied.

References

Hong, Y. and Meeker, W. Q. (2010): “A model for field failure prediction using dynamic environmental data,” Mathematical and Statistical Models and Methods in Reliability Statistics for Industry and Technology, Springer, pp. 223–233.

Hong, Y. and Meeker, W. Q. (2013): “Field-failure predictions based on failure-time data with dynamic covariate information,” Technometrics, Vol. 55, pp. 135–149.

Kordonsky, K. B. and Gertsbakh, I. B. (1993): “Choice of the best time scale for system reliability analysis,” European Journal of Operational Research, Vol. 65, pp. 235–246.

Meeker, W. Q. and Escobar, L. A. (1998): Statistical Methods for Reliability Data, Wiley.

Nelson, W. (1972) “Theory and applications of hazard plotting for censored failure data,” Technometrics, Vol. 14, pp. 945–966.

Nelson, W. (2001): “Prediction of field reliability of units, each under differing dynamic stresses, from accelerated test data,” Handbook of Statistics, Elsevier, Vol. 20, pp. 611–621.

Analysis of Window-Censored Repair Logs and Its Application to Maintenance

Watalu Yamamoto, Lu Jin and Kazuyuki Suzuki

Abstract When tailor-made systems are introduced into fields, markets, or factories, the replacement parts or subcomponents for maintenance are produced and purchased simultaneously so as to keep the quality and reliability of those parts homogeneous against the parts installed into systems at the beginning. Those parts are then kept in stock at yards and will be kept waiting for their turns. Furthermore, the maintenance records of such systems are often window censored, under which the records are both left and right censored. In this talk, we model such situations as a renewal process with incomplete repairs, in which a failed subcomponent is replaced with an unused-but-not-new subcomponent. The statistical inference of the failure time distribution and the effect of in-stock term are implemented using stochastic EM algorithm, which is needed to solve the maximization problem of the incomplete data likelihood under window censoring. A real case is analyzed using the proposed model and the effect of in-stock-at-yards term on the total lifetime is assessed by assuming Weibull distribution. We also examine the sampling properties of the estimators for Weibull distributions.

Keywords EM algorithm · Left and right censoring · Recurrent events · Weibull distribution · Window censoring

W. Yamamoto (✉) · L. Jin · K. Suzuki
Department of Informatics, University of Electro-Communications,
Chofu, Tokyo 182-8585, Japan
e-mail: watalu@inf.uec.ac.jp

L. Jin
e-mail: jinlu@inf.uec.ac.jp

K. Suzuki
e-mail: suzuki@inf.uec.ac.jp

C.M. Tan and T.N. Goh (eds.), *Theory and Practice of Quality and Reliability Engineering in Asia Industry*,
DOI 10.1007/978-981-10-3290-5_25

1 Introduction

When tailor-made systems are introduced into fields, markets, or factories, the replacement parts or subcomponents for maintenance are produced and purchased simultaneously so as to keep the quality and reliability of those parts homogeneous against the parts installed into systems at the beginning. Those parts are then kept in stock at yards and will be kept waiting for their turns. Furthermore, the maintenance records of such systems are often window censored, under which the records are both left and right censored. In this talk, we model such situations as a renewal process with incomplete repairs, in which a failed subcomponent is replaced with an unused-but-not-new subcomponent. The statistical inference of the failure time distribution and the effect of in-stock term are implemented using stochastic EM algorithm, which is needed to solve the maximization problem of the incomplete data likelihood under window censoring.

First, we formulate the model and the estimation procedure in Sects. 2 and 3. Since E-step is complicated, we propose to use an approximation in Sect. 4. Section 5 is dedicated to the analysis of a real case by the proposed method and the effect of in-stock-at-yards term on the total lifetime is assessed by assuming Weibull distribution. Section 6 shows a result of numerical experiments to see the property of the proposed procedure under common window and Weibull interevent distribution.

2 Recurrent Events Formulation

Let $t_{ij}, j = 1, \ldots, q_i$, be the times at which an event occurred, for $i = 1, \ldots, n$. The event time t_{ij} follows some probability distribution

$$f(t_{ij}) = f(t_{ij}|H(t_{ij}))$$

conditional on the history $H(t_{ij})$ up to the occurrence time t_{ij}.

Let $H_n = \{t_{ij};\ i = 1,\ \ldots,\ n,\ j = 1,\ \ldots,\ q_i\}$ be the event history on n repairable systems within operation periods $\{[b_i, c_i]; i = 1, \ldots, n\}$. For each system, $H_i([b_i, t)) = \{t_{ij}; b_i \le t_{ij} < t\}$ is the history of events occurred within the interval $[b_i, t)$. Sometimes we drop b_i and simply write the history up to time t as $H_i(t) = \{t_{ij}; b_i \le t_{ij} < t\}$.

If the durations between successive event occurrences

$$\Delta t_{ij} = t_{ij} - t_{i,j-1}$$

follow a common probability distribution, $f(\Delta t_{ij}; \theta)$, then the probability distributions are given by

$$f\left(t_{ij}|H(t_{ij});\theta\right) = f(t_{ij} - t_{i,j-1};\theta)$$

for $j = 1, \ldots, n_i$, and the underlying stochastic process is called as a renewal process with inter-renewal distribution $f(\Delta t;\theta)$. When failed components are replaced with brand new replacements, and any failure is not affected by surrounding system, the model of this type is used as the first working model.

For large systems, such as facilities of social infrastructure, the replacements for component failures are supplied together with the systems, when the systems are installed. When the components are not general and are custom-made, they are manufactured and delivered so as to let them have same quality and reliability as those installed initially to the operated systems. The replacements are kept in stock before they are required to work. In this case, the distribution of the event time, t_{ij}, depends on the length of in-stock period $t_{i,j-1} - b_i$.

If the differences among lifetime distribution for in-stock replacements and that for in-operation components are well described by proportional hazard models, then the cumulative hazard function for the observed failure at t_{ij} is best formulated as

$$\Lambda\left(t_{ij}|H\left(t_{ij}\right);\theta\right) = \int_{b_i}^{t_{i,j-1}} \alpha\lambda(u - b_i;\theta)du + \int_{t_{i,j-1}}^{t_{i,j}} \lambda(u - b_i;\theta)du,$$

where $\lambda(t;\theta)$ is the hazard function for the component in operation, and α is the proportionality constant for the difference of hazards between operations and stocks.

If we hire an accelerated lifetime model, the time before operation is accelerated or decelerated by the factor α and the time in operation is the standard factor 1, and the resulting density function is formulated as

$$f(t - b_i;\theta) = \begin{cases} f(\alpha(t - b_i);\theta) \\ f\left(t - t_{i,j-1} + \alpha\left(t_{i,j-1} - b_i\right);\theta\right) \end{cases}$$

where $t_{i,j-1}$ is the event time previous to t_{ij}.

When the underlying distribution is Weibull, these two models coincide with each other.

3 Window Censoring and EM Algorithm

If the observation periods are $[b_i, c_i]$, $i = 1, \ldots, n$, the failure logs are right censored, we can estimate the parameters of underlying model by maximizing the likelihood function

$$L(\theta) = \prod_{i=1}^{n} \left\{ \prod_{j=1}^{q_i} f\left(t_{ij} - t_{i,j-1}; \theta\right) \right\} \left\{1 - F\left(c_i - t_{i,q_i}; \theta\right)\right\}$$

where $F(t;\theta)$ is the cumulative distribution function.

When the failure records are censored in both sides, it is called window censored, in that the data is observed through windows. It is assumed in this article that each target has only one window. Any event is observed when it is occurred in the window, $[o_i, c_i]$, for $i = 1, \ldots, n$. It is also assumed that the installation dates, b_i, $i = 1, \ldots, n$, are available for all targets.

Let u_i be the number of failures left censored in $[b_i, o_i)$, r_i be the number of observed failures occurred within the time interval, $[o_i, c_i]$, and $q_i = u_i + r_i$ be the total number failures occurred in $[b_i, c_i]$. The number u_i is missing and failure records $t_{i,1}, \ldots, t_{i,u_i}$ are also missing. We would like to apply EM algorithm framework to handle this incompleteness.

Let $\mathcal{H}_{\mathrm{i}}(b_i, o_i)$ be the history of events before the window opened, $[b_i, o_i)$, $\mathcal{H}_{\mathrm{i}}(o_i, c_i)$ be the history of events within the window $[o_i, c_i)$, and $\mathcal{H}_{\mathrm{i}}[b_i, c_i)$ be the whole history. We would like to compute the conditional expectation of the log-likelihood function

$$l_i(\theta; \mathcal{H}_i(b_i, c_i)) = \sum_{j=1}^{q_i} \log f\left(t_{ij} - t_{i,j-1}; \theta\right) + \log\{1 - F(c_i - t_{i,q_i}; \theta)\}$$

with respect to $\mathcal{H}_{\mathrm{i}}(b_i, o_i)$ given the history within the window $\mathcal{H}_{\mathrm{i}}(o_i, c_i)$, as

$$Q_i(\theta) = E[l_i(\theta; \mathcal{H}_i(b_i, o_i)) | \mathcal{H}_i(o_i, c_i)] + l_i(\theta; \mathcal{H}_i(o_i, c_i)).$$

EM algorithm maximizes the sum of Q_i's to obtain the maximum likelihood estimate of the unknown parameters θ. If this conditional expectation is easily handled, the implementation of EM algorithm is rather simple. However, it is not the case for this situation. The likelihood contribution of the missing part and observed part have expressions as

$$l_i(\theta; \mathcal{H}_i(b_i, o_i)) = \sum_{j=1}^{u_i} \log f\left(t_{ij} - t_{i,j-1}; \theta\right) + \log\{1 - F(o_i - t_{i,u_i}; \theta)\}$$

and

$$\begin{aligned} l_i(\theta; \mathcal{H}_i(o_i, c_i)) = \log\left\{1 - F\left(t_{i,u_i+1} - o_i\right)\right\} + \sum_{j=u_i+1}^{u_i+q_i} \log f\left(t_{ij} - t_{i,j-1}; \theta\right) \\ + \log\{1 - F(c_i - t_{i,u_i+r_i}; \theta)\}. \end{aligned}$$

They require us to compute the conditional expectation with respect to the unobserved events and the count, $t_{i,1}, \ldots, t_{i,u_i}, u_i$, conditional on the observed events and the count $t_{i,u_i+1}, \ldots, t_{i,u_i+r_i}, r_i$, and other information such as the birth date b_i and the window o_i and c_i.

Once the conditional expectations of the log-likelihood function given the current estimates $\hat{\theta}$ in E-step, the remaining task is to maximize it with respect to θ again to update the estimates. This is M-step. For the details of EM algorithm, it is recommended to refer McLachlan and Krishnan (1997).

4 Imputing the Missing Logs by Approximated Monte Carlo Procedure

As mentioned in the previous section, the application of MCEM algorithm, EM algorithm with Monte Carlo approximation to E-step, proposed by Wei and Tanner (1990) is not handy for the situation we are investigating. Rather than to implement the exact conditional distribution, we would like to avoid this complicated calculations by introducing an approximation. The conditional distributions of the unobserved are random samples from the unconditional distribution when both the windows and the renewal distributions are homogeneous.

1. Generate random samples $\tilde{t}_{i,j}$ from the probability distribution with density function $f\left(t_{ij}|H\left(t_{ij}\right);\theta\right)$.
2. Stop generations when the largest generated time steps into the window, i.e., $\tilde{t}_{i,\tilde{u}_i} > o_i$, where $\tilde{u}_i$ is the number of missing failures generated.

This procedure generates the conditional distribution only when the first event time within window coincides with that of observed, $\tilde{t}_{i,\tilde{u}_i+1} = t_{i,u_i+1}$. Everybody knows this event has probability zero. However, this is justified under some special situations, such as common widows and homogeneous renewal distributions.

5 Case Study

This case contains failure logs from systems of two types, A and B. Both types provide same functionality and use similar parts. When a system is failed, a component which trigger the failure will be identified and replaced with a stocked replacement. These replacement components are not brand new in that they have been purchased when the systems are purchased, because of their specially ordered designs. The installation periods are late 80s and early 90s. The window for this case was opened in 1998 and was closed in 2011. The data set is displayed in Figs. 1 and 2.

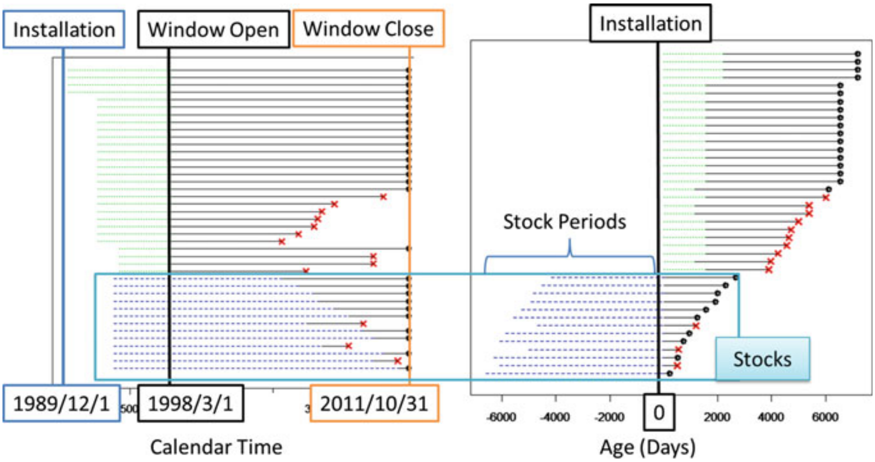


Fig. 1 The repair log for Type A

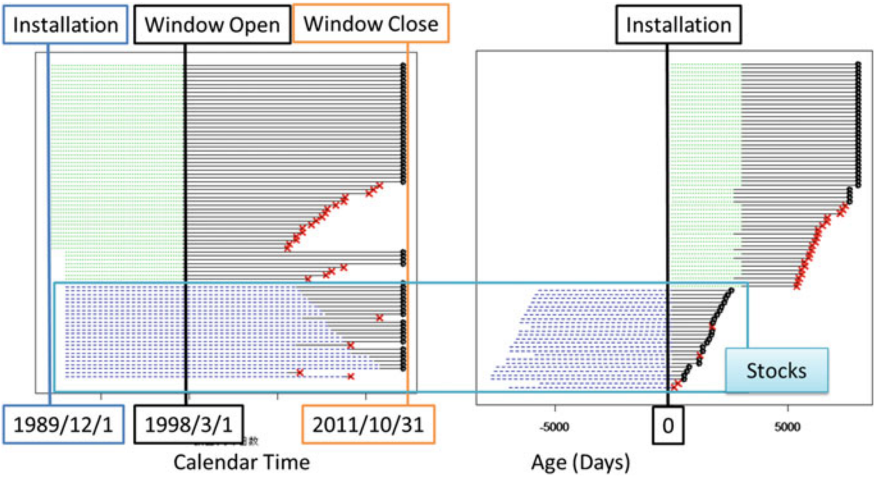


Fig. 2 The repair log for Type B

We would like to know whether degradation speeds in stock periods differ from those in operation, as well as the lifetime distributions vary among Types A and B.

The lifetimes are assumed to follow Weibull distribution with density function

$$f(t) = \left(\frac{m}{\eta}\right) \cdot \left(\frac{t}{\eta}\right)^{m-1} \cdot \exp\left\{-\left(\frac{t}{\eta}\right)^{m}\right\},$$

and survival function

$$R(t) = \exp\left\{-\left(\frac{t}{\eta}\right)^m\right\}.$$

The parameters for Type A have suffix 1 and those for B do 2. Further we introduce the in-stock time factors, α_1 and α_2. The resulting likelihood function is as follows:

$$\begin{aligned} L = &\prod_{i_1=1}^{n_1}\prod_{j_1=1}^{d_{i1}} f\left(z_{i1,j1}\right) \cdot \{f(y_{i1})\}^{e_{i1}} \cdot \{R(y_{i1})\}^{1-e_{i1}} \\ &\cdot \prod_{l_1=1}^{N_1} \{f(t_{l1} + \alpha_1 \cdot T_{l1})\}^{e_{l1}} \cdot \{R(t_{l1} + \alpha_1 \cdot T_{l1})\}^{1-e_{l1}} \\ &\cdot \prod_{i_2=1}^{n_2}\prod_{j_2=1}^{d_{i2}} f\left(z_{i_2,j_2}\right) \cdot \{f(y_{i_2})\}^{e_{i_2}} \cdot \{R(y_{i_2})\}^{1-e_{i_2}} \\ &\cdot \prod_{l_2=1}^{N_2} \{f(t_{l_2} + \alpha_2 \cdot T_{l_2})\}^{e_{l_2}} \cdot \{R(t_{l_2} + \alpha_2 \cdot T_{l_2})\}^{1-e_{l_2}}, \end{aligned}$$

where we have changed the notations slightly. To compare these types, we set four candidate models.

Model 1: Separated models with parameters $m_1, m_2, \eta_1, \eta_2, \alpha_1, \alpha_2$

Model 2: Common shape model with parameters $m_1 = m_2 = m, \eta_1, \eta_2, \alpha_1, \alpha_2$

Model 3: Common in-stock time factor model with parameters $m_1, m_2, \eta_1, \eta_2, \alpha_1 = \alpha_2 = \alpha$

Model 4: Common parameters model with parameters $m_1 = m_2 = m, \eta_1, \eta_2, \alpha_1 = \alpha_2 = \alpha$

As is described in McLachlan (1997), the covariance matrix of estimators can be estimated using conditional expectation of the observed Fisher information matrix with components

$$I_{k,k'} = -\sum_{i=1}^{n} E\left[\frac{\partial^2 l_i(\theta; \mathcal{H}_i(b_i, o_i))}{\partial\theta_k \partial\theta_{k'}} | \mathcal{H}_i(o_i, c_i)\right] - \sum_{i=1}^{n} \frac{\partial^2 l_i(\theta; \mathcal{H}_i(o_i, c_i))}{\partial\theta_k \partial\theta_{k'}},$$

where expectations are calculated using the estimated parameters. We omit the detailed notations and simply write this operation as

$$I_{k,k'} = -\sum_{i=1}^{n} E\left[\frac{\partial^2 l_i(\theta; \mathcal{H}_i(b_i, c_i))}{\partial\theta_k \partial\theta_{k'}} | \mathcal{H}_i(o_i, c_i)\right] = -E\left[\frac{\partial^2 l}{\partial\theta_k \partial\theta_{k'}}\right]$$

For example, the Fisher information matrix for model 1 is given by

$$I = -E\begin{bmatrix} \frac{\partial^2 l}{\partial m_1^2} & \frac{\partial^2 l}{\partial m_1 \partial \eta_1} & \frac{\partial^2 l}{\partial m_1 \partial \alpha_1} \\ \frac{\partial^2 l}{\partial m_1 \partial \eta_1} & \frac{\partial^2 l}{\partial \eta_1^2} & \frac{\partial^2 l}{\partial \eta_1 \partial \alpha_1} \\ \frac{\partial^2 l}{\partial m_1 \partial \alpha_1} & \frac{\partial^2 l}{\partial \eta_1 \partial \alpha_1} & \frac{\partial^2 l}{\partial \alpha_1^2} \end{bmatrix}$$

and the covariance matrix of estimators are estimated by

$$I^{-1} = \begin{bmatrix} \mathrm{Var}[\hat{m}_1] & \mathrm{Cov}[\hat{m}_1, \hat{\eta}_1] & \mathrm{Cov}[\hat{m}_1, \hat{\alpha}_1] \\ \mathrm{Cov}[\hat{m}_1, \hat{\eta}_1] & \mathrm{Var}[\hat{\eta}_1] & \mathrm{Cov}[\hat{\eta}_1, \hat{\alpha}_1] \\ \mathrm{Cov}[\hat{m}_1, \hat{\alpha}_1] & \mathrm{Cov}[\hat{\eta}_1, \hat{\alpha}_1] & \mathrm{Var}[\hat{\alpha}_1] \end{bmatrix},$$

where I^{-1} is the inverse of matrix I. The marginal confidence intervals are estimated using normal approximations as

$$\hat{\theta}_i \pm u(a) \cdot \sqrt{\mathrm{Var}\left[\hat{\theta}_i\right]/n},$$

where $u(a)$ is $100 \times a$ percentile of the standard normal distribution.

Table 1 summarizes AIC for each model. AIC leads us to hire Model 4, the one with common shape parameter m and common in-stock time factor α.

The estimates and marginal confidence intervals for Model 4 is shown in Table 2.

Table 2 tells us that MTTF of the component of interest in Type A is shorter than that in Type B. The estimated in-stock time factor is 0.703. If we keep a replacement component in stock for 10 years, it degrades to the same level as that in operation for 7.03 years.

Table 1 AIC for models

	Model 1	Model 2	Model 3	Model 4
AIC	785.9	784.5	783.9	782.7

Table 2 Parameter estimates for Model 4

	95% Lower confidence limit	Point estimate	95% Upper confidence limit
m	4.26	4.33	4.39
η_1	7.71×10^3	7.75×10^3	7.80×10^3
η_2	9.35×10^3	9.42×10^3	9.49×10^3
α	0.692	0.703	0.715

6 Simulation Study

In this section, we would like to investigate whether our approximation in E-step is valid in common window cases. Let the intervals between successive events follow a Weibull distribution with shape parameter $m = 4$ and scale parameter $\eta = 600$. We set two cases for common windows, one with $[150, 1000]$ and another with $[350, 1000]$. The number of systems is set as $n = 1000$. The replications of simulation study are 100 (Table 3).

Before conducting simulations, we see that MTBF is $600 \cdot \Gamma(1 + 1/4) = 531$, and the cumulative probabilities for times 150 and 350 are $F(150) = 0.00389\%$ and $F(350) = 0.179\%$, respectively. Under Case 1, it is expected that there will be almost no failure for each before window opened. The systems in Case 2 will see a failure at most before window opened.

The initial estimates for EM iterations are set as $m^{(0)} = 3$ and $\eta^{(0)} = 550$. The results of 100 replications are summarized in Tables 4 and 5.

These tables allow us to conclude that our approximation to E-step of EM algorithm works well under common window situations.

Table 3 Windows for simulation study

	Open time	Close time
Case 1	150	1000
Case 2	350	1000

Table 4 Summary of sampling distributions of estimates (Case 1)

	Set	Average	Standard deviation
m	4	3.98	0.170
η	600	601.0	12.2

Table 5 Summary of sampling distributions of estimates (Case 2)

	Set	Average	Standard deviation
m	4	3.97	0.195
η	600	599.8	12.3

7 Remarks

Though our proposal works well, the precision of approximation and its effect to the precision of estimates needs to be assessed. We have an idea to use importance sampling to avoid approximations. However, the formulation of conditional expectations in recurrent events settings is not as simple as that for i.i.d. situations.

Acknowledgments The authors would like to thank Mr. Daisuke Tanada and Mr. Ryo Endo for supporting them in computation.

References

Wei, G. C. G. and Tanner, M. S. (1990): “A Monte Carlo Implementation of the EM Algorithm and the Poor Man’s Data Augmentation Algorithms”, *Journal of the American Statistical Association*, Vol. 85, pp. 699–704.

McLachlan, G. J., and Krishnan, T. (1997): “The EM Algorithm and Extensions”, John Wiley & Sons.

Authors Biography

Watalu Yamamoto is Assistant Professor in Department of Informatics at University of Electro-Communications. His major is industrial statistics.

Lu Jin is Assistant Professor in Department of Informatics at University of Electro-Communications. Her major is reliability engineering.

Kazuyuki Suzuki is Professor in Department of Informatics at University of Electro-Communications. He has served as the chairperson of Japanese Society for Quality Control and is currently serving as the chairperson of Reliability Engineering Association of Japan.