Lecture Notes in Mechanical Engineering

Ming J. Zuo
Lin Ma
Joseph Mathew
Hong-Zhong Huang *Editors*

Engineering Asset Management 2016

Proceedings of the 11th World Congress on Engineering Asset Management

Springer

Lecture Notes in Mechanical Engineering

About this Series

Lecture Notes in Mechanical Engineering (LNME) publishes the latest developments in Mechanical Engineering—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNME. Also considered for publication are monographs, contributed volumes and lecture notes of exceptionally high quality and interest. Volumes published in LNME embrace all aspects, subfields and new challenges of mechanical engineering. Topics in the series include:

- Engineering Design
- Machinery and Machine Elements
- Mechanical Structures and Stress Analysis
- Automotive Engineering
- Engine Technology
- Aerospace Technology and Astronautics
- Nanotechnology and Microengineering
- Control, Robotics, Mechatronics
- MEMS
- Theoretical and Applied Mechanics
- Dynamical Systems, Control
- Fluid Mechanics
- Engineering Thermodynamics, Heat and Mass Transfer
- Manufacturing
- Precision Engineering, Instrumentation, Measurement
- Materials Engineering
- Tribology and Surface Technology

More information about this series at http://www.springer.com/series/11236

Ming J. Zuo • Lin Ma • Joseph Mathew •
Hong-Zhong Huang
Editors

Engineering Asset Management 2016

Proceedings of the 11th World Congress on Engineering Asset Management

Springer

Editors
Ming J. Zuo
Department of Mechanical Engineering
University of Alberta
Edmonton, Alberta
Canada

Lin Ma
Brisbane, Queensland
Australia

Joseph Mathew
Asset Institute
Brisbane, Queensland
Australia

Hong-Zhong Huang
Institute of Reliability Engineering
UESTC
Sichuan, China

ISSN 2195-4356 ISSN 2195-4364 (electronic)
Lecture Notes in Mechanical Engineering
ISBN 978-3-319-62273-6 ISBN 978-3-319-62274-3 (eBook)
DOI 10.1007/978-3-319-62274-3

Library of Congress Control Number: 2017953872

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

11th WCEAM—25 to 28 July 2016 Jiuzhaigou, China

Quality, reliability, risk, maintenance, safety, and engineering asset management are becoming increasingly important areas of endeavor for industries, governments, and academia. Today, more than ever these areas of endeavor must be addressed for all engineering systems such as space shuttles, civil aircrafts, nuclear systems, etc.

The 2016 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE 2016) was held in conjunction with the 11th World Congress on Engineering Asset Management (WCEAM 2016) in Jiuzhaigou, Sichuan, China, July 25–28, 2016, hosted by the International Society of Engineering Asset Management (ISEAM). The QR2MSE is an annual international conference series which brings together leading academics, industry practitioners, and research scientists from around the world to advance the body of knowledge in quality, reliability, maintenance, and safety of engineering systems; to establish and strengthen the link between academia and industry; to promote applications of research results in practice; and to showcase the state of the art of industrial technologies. The QR2MSE international conference was founded in 2011 and has grown admirably through the support of many academic organizations and colleagues and has become a premier conference in this field in Asia. WCEAM has been held annually since 2006 commencing with the inaugural event on the Gold Coast in Queensland, Australia, and has objectives that are well aligned with the objectives of QR2MSE but addressing the overall multidisciplinary field of Engineering Asset Management. The joint event attracted 288 delegates from 21 countries.

The joint congress hosted an excellent technical program and several opportunities for networking through social events including the congress dinner. The conference program comprised nine keynote speeches, one workshop, and 20 regular technical paper oral sessions.

The keynotes were:

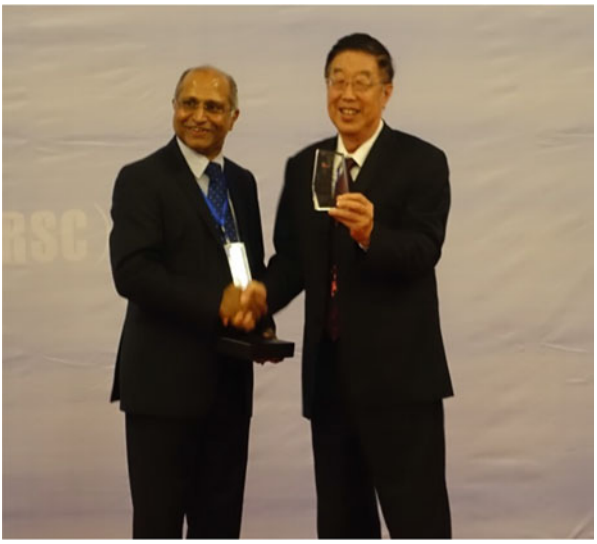
1. Professor Narayanaswamy Balakrishnan, McMaster University, Ontario, Canada, on "Parametric and semiparametric inference for one-shot device testing."
2. Adjunct Professor Joseph Mathew, Asset Institute, Brisbane, Australia, on "Innovations in engineering asset management."
3. Professor Jin Wang, Liverpool John Moors University, Liverpool, UK, on "Risk-based decision making in design and operation of large engineering systems under uncertainties."
4. Professor Dong Ho Park, Hallym University, Chuncheon, Korea, on "Two-dimensional system maintenance strategy and its effects."
5. Professor Gao Jinji, "Process machinery diagnosis and maintenance based on industry internet and big data analysis."
6. Professor Guoliang Huang, University of Missouri, Columbia, USA, on "Modelling, experimental investigation and reliability of elastic metamaterials."
7. Professor Suk Joo Bae, Hangyan University, Seoul, Korea, on "Reliability issues in fuel cell technology."
8. Professor Lin Ma, Queensland University of Technology, Brisbane, Australia, on "Reliability modelling of electricity transmission networks using maintenance records."
9. Professor Xiayue Wu, National University of Defense Technology, Changsha, China, on "Mission reliability evaluation of spaceflight Telemetry, Tracking and Control (TT&C) systems."

For QR2MSE/WCEAM 2016, each submitted paper was reviewed by two invited reviewers. In total, 190 papers were selected from all submissions and included in the conference proceedings. Thirty-three papers were subjected to additional reviews by two Fellows of the International Society of Engineering Asset Management (ISEAM) per paper, with at least one acceptance before the paper was included in the Springer eBook proceedings.

ISEAM Fellows at the Congress

We would like to acknowledge the efforts of all of the Technical Program Committee members who reviewed the submitted papers. The members of the Technical Program Committee are as follows: Suk Joo Bae, Hanyang University, Korea; In Hong Chang, Chosun University, Korea; David W. Coit, Rutgers University, USA; Yi Ding, Zhejiang University, China; Serkan Eryilmaz, Atilim University, Turkey; ILia B. Frenkel, Sami Shamoon College of Engineering, Israel; Olivier Gaudoin, Ensimag Laboratoire Jean Kuntzmann, France; Antoine Grall, Troyes University of Technology, France; Bo Guo, National University of Defense Technology, China; Abdelmagid S. Hammuda, Qatar University, Qatar; Yu Hayakawa, Waseda University, Japan; Cheng-Fu Huang, Feng Chia University, Taiwan; Shinji Inoue, Tottori University, Japan; Chao Jiang, Hunan University, China; Renyan Jiang, Changsha University of Science and Technology, China; M. Rezaul Karim, University of Rajshahi, Bangladesh; Dae Kyung Kim, Chonbuk National University, Korea; Jongwoon Kim, Korea Railroad Research Institute Testing and Certification Center, Korea; Grzegorz Koszalka, Lublin University of Technology, Poland; Hongkun Li, Dalian University of Technology, China; Yan-Feng Li, University of Electronic Science and Technology of China, China; Zhaojun Li, Western New England University, USA; and Ming Liang, University of Ottawa, Canada.

This year's ISEAM Lifetime Achievement award for exceptional dedication and contribution to advancing the field of Engineering Asset Management was presented to Professor Gao Jinji of Beijing University of Chemical Technology. ISEAM's Lifetime Achievement Award recognizes and promotes individuals who have made a significant contribution to research, application, and practice of a discipline in engineering asset management over a continued period of time. The award was presented at the Gala Dinner of the event.



Prof Gao receiving the ISEAM Lifetime Achievement Award from the ISEAM Chair, Joe Mathew

We would like to express our gratitude to the sponsors for providing their generous support. The sponsors were Sichuan Provincial Key Laboratory of Reliability Engineering, University of Electronic Science and Technology of China (UESTC), National Collaborative Innovation Center of Major Machine Manufacturing in Liaoning, Dalian University of Technology (DUT), Armed Force Engineering Institute, Institute of Special Vehicle, National University of Defense Technology, Northwestern Polytechnical University, The National Taiwan University of Science and Technology, Sichuan Go Yo Intelligence Technology Co., Ltd., Reliability committee of Sichuan Provincial Mechanical Engineering Society, International Society of Engineering Asset Management (ISEAM), European Safety and Reliability Association (ESRA), European Federation of National Maintenance Societies (EFNMS), The Korean Reliability Society (KORAS), Reliability Engineering Association of Japan (REAJ), Polish Safety and Reliability Association (PSRA), Equipment Support Commission of China Ordnance Society, Reliability Committee of Chinese Operations Research Society, IEEE Chengdu Section, National Natural Science Foundation of China (NSFC), and Center for System Reliability and Safety, UESTC.

We are grateful for the voluntary assistance provided by members of the International Advisory Committee, the Program Committee, the Organization Committee, and the Conference Secretariat.

Last but not the least, we would like to appreciate the Congress General Chair and Co-Chairs: Professor Hong-Zhong Huang, UESTC, Adjunct Prof. Joseph Mathew, Asset Institute, Australia, Prof. Carlos Guedes Soares, University of Lisbon, Portugal, and Prof. Dong Ho Park, Hallym University, Korea; the Program

Committee Chair and Co-Chairs: Prof. Liudong Xing, University of Massachusetts, USA, Prof. John Andrews, University of Nottingham, UK, Prof. Lirong Cui, Beijing Institute of Technology, China, and Prof. Wei Sun, Dalian University of Technology, China; and the Organizing Committee Chair and Co-Chairs: Prof. Lin Ma, Queensland University of Technology, Australia, Prof. Yi-Kuei Lin, National Taiwan University of Science & Technology, Taiwan, Prof. Byeng Dong Youn, Seoul National University, Korea, and Prof. Chengming He, Armed Force Engineering Institute, China.

Alberta, Canada — Ming J. Zuo
Brisbane, QLD, Australia — Joseph Mathew
Brisbane, QLD, Australia — Lin Ma
Sichuan, China — Hong-Zhong Huang

Contents

A Model for Increasing Effectiveness and Profitability of Maintenance Performance: A Case Study

Basim Al-Najjar and Hatem Algabroun

Abstract In today's market, companies strive to achieve the competitive advantages. Failing in achieving these goals could threaten the companies' existence. Failures in the operative level impact negatively on achieving these goals. In order to record these failures for better actions planning, special systems are often used for counting the number of failures, the duration of machines downtime and uptime for assessing the total downtime and classifying problems/failures in categories that are decided in advance. In this study, we develop a model to break down the contents of a company failure databases, prioritize failures, assess economic losses due to failure impact on the competitive advantages and suggest a method of how maintenance actions should be rank-ordered cost-effectively. The model is tested using real data. The major results showed that losses are mainly due to two categories i.e. "Bad quality" and "Less profit margin", where failures of "Gear", "Bearing" and "Raw materials quality" cause most of the losses. It is concluded that this model will enable the user to quickly identify and prioritize maintenance and improvement efforts cost-effectively.

Keywords Failures impact • Failure classification • Maintenance impact on competitive advantages • Maintenance decision making • Maintenance performance enhancement

1 Introduction

A good position in the market nowadays requires companies to manage their resources, production process, different systems and sub-systems in a cost-effective manner. Different subsystems/working areas, e.g. maintenance, quality control, production logistics, should be integrated and synchronized to fulfill company goals. Changing in any of these working areas could influence other related areas, due to their internal interactions [1]. Maintenance activities have a very high influence on company performance, due to its importance and impact on different

B. Al-Najjar (✉) • H. Algabroun
Linnaeus University, P G Vejdes väg, 351 95 Växjö, Sweden
e-mail: basim.al-najjar@lnu.se; hatem.algabroun@lnu.se

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_1

working areas, such as quality, production, safety, production cost, working environment, delivery on time, etc. Therefore, a reliable and efficient maintenance not only increases the profitability, but it also improves the overall performance of the company [2]. Maintenance is responsible of reducing the probability of failure and unplanned stoppages in order to minimize the impact of failure consequences on company competitive advantages and goals through maintaining the continuity of a production process, production cost and product quality at a predetermined rate. Companies' production systems are exposed to different types of failures. These failures could be due to technical issues or other elements involved in the production process, e.g. human resources, operating conditions, systems' quality, facilities reliability, etc., which should not be neglected when dealing with maintenance [2–4]. Different types of failures have usually different consequences on competitive advantages. For example a paper/pulp machine can be exposed to many different failures, e.g. electrical, electronic, hydraulic, pneumatic and mechanical, which could have different impact on the machine performance and the company competitive advantages, such as production delay, lower quality, higher production costs, etc. In order to achieve proper utilization of the available resources, failures with significant impacts should be identified and prioritized in order to effectively reduce/eliminate their consequences. This study presents a model that aims to identify and prioritize failures according to their significance and impact on the company's competitive advantages in order to be able to plan and conduct cost-effective maintenance. In order to investigate whether the problem addressed in this study has been treated before, a survey was conducted and three papers [5–7] were found relevant to the problem in this paper.

Al-Najjar [6] presented Maintenance Function Deployment (MFD) model, which aims to pinpoint, analyze, and prioritize causes behind losses in the working areas belonging to the competitive advantages, such as losses due to bad quality and delayed deliveries. MFD breaks down these losses in a backwards method to approach the root-causes behind the losses which are usually spread in the operative level of different disciplines in a production process. MFD also provides business based production analysis. This is done through quantifying losses (in production time and economy), in accordance with the strategic goals of the company, and identifying causes behind them. Then it breaks down the causes and its costs into their root causes. The author used an example to test the model, and the results showed that the model could be used to identify, analyze and quantify losses in order to make cost effective improving decisions. In Al-Najjar and Jacobsson [7], a model that demonstrates the interactions between man-machine-maintenance-economy (MMME) was developed, in order to support cost-effective decisions. The model systematically gathers, categorizes, estimates and quantifies the losses in the production time due to failures in order to identify and prioritize the problem areas in the production process. A software program then was built based on the model and tested in a case study at the automaker FIAT, Italy. The results of the case study showed a possibility to identify the problematic areas. Also, as the model compares the time loss in different time periods, it captures the deviations over time for different categories. Ahamed Mohideen et al. [5] presented a method that reduces the breakdown costs and recovering time in the construction plant systems. It starts

by categorizing and analyzing the breakdown records in order to identify the main breakdowns and the sub-breakdowns using cause effect analysis, and then ranking them using Pareto analysis. The model was tested by a case study on four types of machines in a construction company using four years breakdown records. Major contributing failures and their causes were identified and a strategical plan is proposed accordingly.

The survey showed that the impact of failures -in particular- on the companies' competitive advantages is not payed the proper attention. The model developed in this paper considers in particular the impact of failures on the competitive advantages that companies strive to achieve, maintain and improve continuously, in order to prioritize, analyze and support cost effective decisions for failures that have the most significant impact. This is why the problem addressed in this study is: How to increase maintenance effectiveness, i.e. maintenance capability of achieving a desired result, and profitability?

The next section will discuss the impact of failures on the competitive advantages. Then we will present the model development and model test using real data. It will be followed then by results discussion and conclusion.

2 Impacts of Failures on Competitive Advantages

Failures in production process could occur due to causes arise from different reasons, for example machines, operators, maintenance staff, management and working environment. This is why other elements than the machine should be considered in the model development. The spectrum and contents of maintenance in this article is referring to a broader definition used by Total Quality Maintenance (TQMain); TQMain aims to maintain the quality of the elements (working areas) constituting the production process (and not only the machines) cost effectively [3, 4].

In manufacturing companies, recorded breakdown data is very important to enhance the production process. Special management systems and techniques are used to classify failures and losses into predetermined categories [7–10]. The categorization helps engineers and management to determine the breakdown that occurs most often in each category in order to improve the related procedures. This may help eliminating the breakdown of multiple machines in the related category [10]. Generally, failures could be defined according to different factors, such as failure impact on: production cost, product quality, operation safety or reason of correction action [10]. However, these failure classifications are not sufficient to determine the failure impact on the competitive advantages (CA) of companies. Therefore the model developed aims to determine the impact of failures on a company CA using the failure database available in the company. The common strategic goals of companies in a wide range of industries is to deliver on time a high product quality at a competitive price with low violation of environment, accepted by the society and keeping the manufacturing machine in a reliable condition [6]. The CA that will be considered in this paper are: high product quality, on time delivery, competitive price, low violation of environment, society

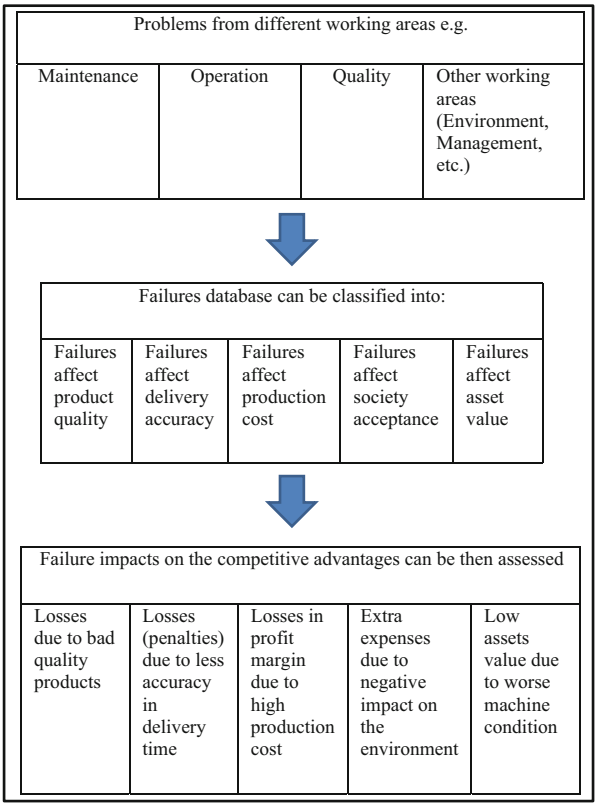


Fig. 1 Failures classification with respect to their impact on the competitive price

acceptance and maintain good assets condition, i.e. asset value, see Fig. 1. Companies should always utilize their resources efficiently and effectively in order to reach the CA with a reasonable profit margin. Failing in achieving this will threaten the company existence. In general, failures in the operative level impact negatively on achieving these goals [3]. In order to assess the significance of failures' consequences according to their impact on the CA of a company, we choose first to classify failures (or sort out failures registered in the company databases) according to their working areas. This step is necessary to:

1. Analyse failures, i.e. to identify root cause, damage initiation and development, imminent failures and failure modes.
2. Specify how failure' consequences converted from technical to economic.
3. Quantify the impact of each failure with respect to predetermined competitive advantages or strategic goal defined by the company, see Fig. 1.

In this paper, failures are classified according to the competitive advantages as follow:

1. Product quality failures: Failing in maintaining the condition of the cutting tools and machines could negatively impact the quality of the final product. This may cause rejection of processed items or final products, and therefore will cause extra costs due to the need of the re-production of the rejected items.
2. Delivery on time failures: Long downtime of producing machines due to failures will shorten the production time, which at the end, affects the production plan and delivery schedules. Therefore, late deliveries and additional production expenses, e.g. due to penalties, should be expected.
3. Competitive price failures: Failures, e.g. in a bearing of a rotating machine, in serially connected production process machines will lead to downtime and production stoppage of the whole production line. Such failures increase the non-added values time, which in turn increase the production cost. The latter will be reflected at the end in the final product price. This will affect a company achieving the production cost at a particular level in order to be able to offer their customers a competitive price with a reasonable profit margin.
4. Society acceptance failures: Improper maintenance of, e.g. a diesel engine could in some cases increase the pollution ejected to the environment, which will lead to extra costs to adapt working conditions of the engine to the national or international regulations.
5. Machine condition (Asset value) failures: Improper operation and maintenance of machines will lead to increased deterioration and reduced machines reliability, and consequently, continuously reduced value of the company's assets.

3 Model Development

Companies use, in general, specific databases to chronologically register information regarding, e.g. type of machine, failure causes, failure mode, failure stoppage times, used spare parts, etc. Failures analysis, classification and impact have been considered by many authors [5, 7, 9]. In literature, failures have been classified with respect to different attributes, and their impact has been assessed using, for example downtime cost, i.e. the losses in production time [3] and the references cited there. But, there is lack of models assessing the real total economic losses due to failures especially when it is associated with company competitive advantages [6]. In this paper, a model is developed to describe the probable economic impact of every failure on CA (CA-Failures), see Fig. 2. Failure economic consequences/impacts are identified and assessed through breaking down the contents of the failure database to determine the effect of failures on the CA, see also Fig. 1. CA-Failures model can be used for the failures that are already saved in machine or company databases and those that may occur in future. Occurred failures can already be classified into for example; mechanical, electrical, electronic, hydraulic, pneumatic failures and others, or they may be mixed in only one class of failures, such as Miscellaneous/Others. For simplicity we choose the classification that is mostly applied in industry as described above. In order to identify and assess the

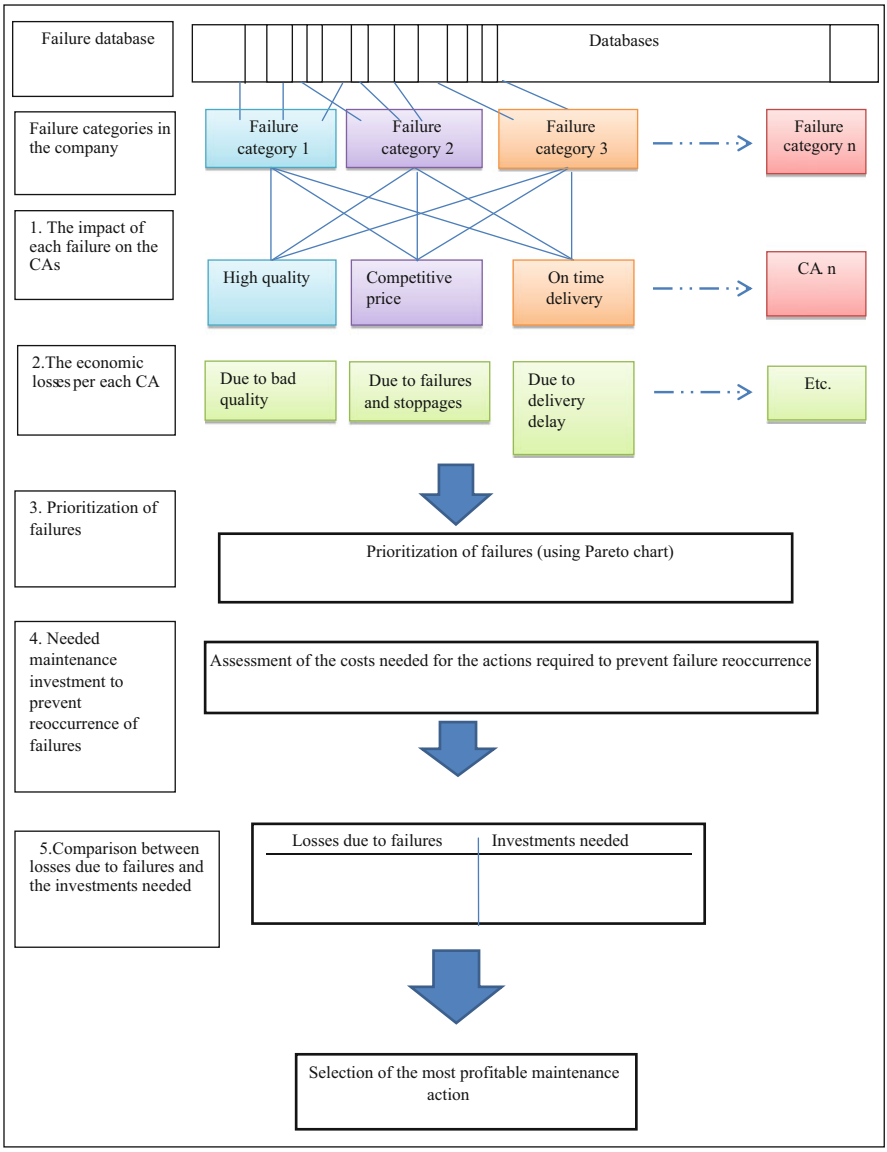


Fig. 2 Model (CA-Failures) operative flow

impact of every failure, it is crucial to analyze it with respect to the CA of a company. When a failure is analyzed, it will be possible to realize its impact on the CA, for example, failure of a gearbox due to oil leakage to the ground, may causes several problems related to different CAs. First of all, it results in stoppage of the production of that machine. Breakdown of a producing machine in a serially connected production process leads to production stoppage and losses of production

time, income and profit margin in the whole production line. Also, this will cost additional expenses to remove the oil to avoid harming soil. It may also cause accidents and fires which expose staff and properties safety to high risk and catastrophic consequences. These failures may lead to delay in the delivery of production according to the delivery time schedule which in turn results to pay penalties, losses of customer if it is repeated and consequently market shares. Therefore, prioritizing failures due to their significance can be done through summing the impacts of every failure on the CAs and applying Pareto diagram.

The highest prioritized failure is that which cause more damage and leads to highest economic losses. Significant failures can be prioritized then and suggest the most profitable maintenance actions through comparing the economic losses with the maintenance expenses. In this way, it will be possible to conduct cost-effective maintenance through prioritizing the most profitable maintenance actions, see also Fig. 2.

4 Model (CA-Failures) Test

In order to examine the applicability of CA-Failures, real data from Auto CNC-Bearbetning i Emmaboda AB was tested on the model. The company is located in the Southern part of Sweden and specialized in producing small series of mechanical parts for water pumps and other industrial products. For simplicity, one product, "sleeve", which is a component for water pumps, and one production line were selected. The product is one of many other products and it was selected basing on the criteria of being a relatively expensive product and it is produced in a relatively large quantity due to its high sale ratio. The data set contains failures occurring in three months, which was collected from the production line under study. It is consisted of 19 failures that are distributed among four categories namely: Hydraulic, Electrical, Mechanical, and Others. Note that, in some companies the latter category could possibly contain a considerable amount of registered failures due to easiness; therefore attention should be paid on this category [7].

In order to calculate the impact of each failure on the CAs, the failure sample was then broken down into single failures. Assessment of the impact of each failure on the CAs was not available in the case company. Therefore the total losses in each CA were assessed by considering the ideal production situation for the production line under study and subtract it from the actual situation. For example in Table 1, "less profit margin" was assessed as follow: the ideal production rate was 74 items/shift. This amount was then subtracted from the average of the actual production rate which was about 56 items/shift and therefore the losses were about 71.3%. Table 1 shows the losses distribution. Other losses in the CAs (e.g. losses due to delivery delays, environment violations, etc.) were insignificant according to the company's experience, so they have been neglected by the company and hence no information was provided.

Table 1 Losses according to the company CAs

No.	Loss categories according to the CA	Losses in units/shift	Share of losses (%)	Comments
1	Bad quality	7.3	28.7	Losses due to internal causes, e.g. scrap, reworking, and external causes (compensations for customers, warranties, etc.)
2	Losses in profit margin, which influences product price and possibility of offering customers a competitive price	18	71.3	Unnecessary production costs due to failures, short stoppages and disturbances
Total		25.3	100	

Table 2 Failures impact on company's CAs

Number of failures occurred	Reason/ Component behind machine failure	Losses per failure due to production of defective items in percentage	Losses per failure due to increased production cost in percentage	Total failure impact in percentage
1	Shaft	1	2	3
2	Cooling system	1	4	5
1	User error	1	2	3
5	Cutting tools	4.7	3	7.7
1	Pump	1	2.3	3.3
7	Bearing	2	23	25
1	Raw material quality	17	9	26
1	Gear	1	26	27
Total 19		28.7%	71,3%	100%

In order to find the impact of each failure on each CA, reasonable estimation of each failure impact on the losses per each CA were made utilizing the authors' experience in consultation with the case company, see Table 2. Then a summation of each failure impact on each CA was performed (see Table 2). There were only two categories of losses connected to two CA i.e. "Bad quality" and "Losses in profit margin".

A Pareto diagram is then applied to determine the failures of the most economic significance, i.e. those failures whose losses are profitable to avoid. Figure 3 shows

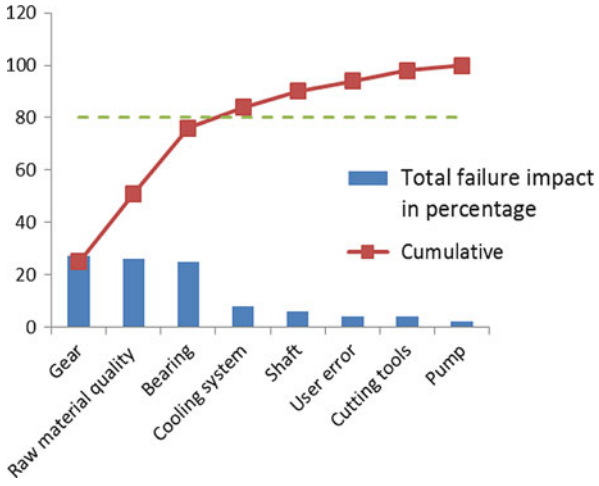


Fig. 3 Pareto diagram for failure impact on CA

the Pareto diagram application. Pareto analysis shows that only failures of "Raw material quality", "Bearing" and "Gear" cause about 78% of the total loss on the CAs. Whereas adding the failure of "Cooling system" will increase the loss by 8% only, therefore it was decided to prioritize only the first three mentioned failures.

The prioritized failures are then analysed to improve the related process or technique in order to prevent the failures' reoccurrence. Then, the needed investment is assessed. However, the needed investment should be compared with the losses in order to make a cost effective decision, whether to neglect, postpone or solve the problem. A target value should be estimated, before being able to make a cost effective decision. The target value refers to the expected losses after applying a particular solution—as a percentage of the total failure losses. Also, determining a target value will be useful in order to enhance the continuous improvement and activities follow-up.

When failures are prioritized according to their significance, the investments needed are identified and target values are set for each CA. A cost-effective decision can then be achieved, see Table 3.

In the next section, a discussion of the results and how CA-Failures support cost effective decisions will be presented.

5 Discussions

After analysing the failure data using the model CA-Failures, it is shown that the losses are mainly due to two categories i.e. "Bad quality" and "Losses in profit margin" and the losses were 28.7% and 71.3% of the total failure respectively (see Table 1). Failures of "Gear", "Raw materials quality" and "Bearing" cause most of the failures losses, i.e. 78% in total (see Fig. 3 and Table 3). To prevent these losses,

Table 3 Analysis of the prioritized failures (Prioritized failures are "Gear", "Raw material quality" and "Bearing")

Failures	Loss per prioritized failure as a percentage of the total losses	Suggested solutions for process improvement or failure prevention (Suggested in consultation with by the company)	Investment needed for each solution (as a percentage of the total losses)	Target value i.e. the expected losses after the investment (as a percentage of the total losses)
Gear	27	Apply CBM using vibration	15	2
		Maintenance staff training	5	
Raw material quality	26	To find and establish an analysis technique for quality control of the raw materials	17	Unknown
Bearing	25	Apply CBM using vibration	13	1
		Training of operator to increase awareness about the bearing condition	2	
		Maintenance staff training	5	
Shaft	3	–	–	–
Cooling system	5	–	–	–
User error	3	–	–	–
Cutting tools	7.7	–	–	–
Pump	3.3	–	–	–
Total	100			

investments are needed; by comparing the needed investments and the target value a cost-effective decision could be made. In the given case study, failures of "Gear" need two actions to prevent the failure reoccurrence, i.e. applying CBM using vibration monitoring techniques and training of maintenance staff. The total investment needed for that is estimated to be equivalent to 20% of the total losses (i.e.100%). After applying the suggested solution, the "Gear" failure losses are expected to be reduced from 27% to only 2% (i.e. 25% of the total losses will be saved). By this investment, the company will gain about 5% in the first year, see Table 3.

In order to prevent "Bearing" failures, we reutilize CBM using vibration monitoring technique and trained maintenance staff and operators. This requires additional 20% of the total loss to be invested. This solution is expected to reduce the

"Bearing" failure losses from 25% to only 1% (i.e. 24% of the total losses will be saved). By this investment, the company will gain about 4% in the first year, see Table 3. However on the long run, additional gain is also expected.

Failures due to "Raw material quality" demand additional investment of around 17% of the total losses. Since the target value was unknown, therefore a decision concerning the investment cost-effectiveness could not be made and therefore this issue has been left to company until the estimation of the target value is done.

6 Conclusions

The model structurally analyses the failure database to identify and prioritize the failures that affect the company the most. It supports the decision maker to customize a maintenance plan that suits the company's special situation. There was no problem found in applying the failure data to the model, thus for future work using the model in more case studies could be interesting.

It is important to note that in some cases, one solution could help in solving multiple problems in different categories, for example, training of the operators (see Table 3) could help to prevent failures that are not only related to bearing but also to prevent other failures that are related to, for example, misuse. Using a method, e.g. Maintenance Function Deployment in Al-Najjar [6], to determine and identify the interaction and the values shared among different failure areas could be interesting.

References

1. Al-Najjar B (2003) Modelling the design of the integration of quality, maintenance and production. In: Proceeding of the 16th international congress COMADEM 2003, Växjö, Sweden, pp 23–31
2. Waeyenbergh G, Pintelon L (2002) A framework for maintenance concept development. Int J Prod Econ 77(3):299–313
3. Al-Najjar B (2007) The lack of maintenance and not maintenance which costs: a model to describe and quantify the impact of vibration-based maintenance on company's business. Int J Prod Econ 107(1):260–273
4. Sherwin DJ, Al-Najjar B (1999) Practical models for condition monitoring inspection intervals David. J Qual Maint Eng 5(3):203–221
5. Ahamed Mohideen PB, Ramachandran M, Narasimmalu RR (2011) Construction plant breakdown criticality analysis – part 1:UAE perspective. BIJ 18(4):472–289
6. Al-Najjar B (2011) Maintenance impact on company competitiveness & profit. In: Asset management: the state of the art in Europe from a life cycle perspective. Springer Science Business Media B.V., Dordrecht, pp 115–141
7. Al-Najjar B, Jacobsson M (2013) A computerised model to enhance the cost-effectiveness of production and maintenance dynamic decisions. J Qual Maint Eng 19(2):114–127
8. Hoon Lee J, Chan S, Soon Jang J (2010) Process-oriented development of failure reporting, analysis, and corrective action system. Int J Qual Stat Reliab 2010:1–8

9. Villacourt M, Drive M (1992) Failure reporting, analysis and corrective action system in the us semiconductor manufacturing equipment industry: a continuous improvement process. In: Thirteenth IEEE/CHMT international electronics manufacturing technology symposium. IEEE, pp 111–115
10. Villacourt M, Govil P (1993) Failure reporting, analysis, and corrective action system. Marcel Dekker, New York

Maintenance Process Improvement Model by Integrating LSS and TPM for Service Organisations

Barrak Alsubaie and Qingping Yang

Abstract This paper presents an integrated model to provide guidance and support for those organisations who aim to reach world-class standards in maintenance processes through continual improvement. A strategic model has been developed through conceptual integration of three popular process improvement strategies, which are six sigma, total productive maintenance (TPM) and lean. Lean Six Sigma can operate in parallel with the TPM strategy and will make it easier to understand by shop floor operators. The application of the model has been demonstrated using a case study in maintenance of a fleet of military vehicles. The proposed model is very generic in nature and can be applied to any service organisations with maintenance functions to achieve high process performance and overall equipment effectiveness.

Keywords Lean Six Sigma • TPM Strategy • Integrating Lean Six Sigma and TPM • Maintenance Process Improvement

List of Abbreviations

A	Availability
CTQ	Critical to quality characteristics
DMAIC	Define, measure, analyse, improve and control
DPMO	Defect per million opportunities
LSS	Lean Six Sigma
MTBF	Mean time between failures
OEE	Overall equipment effectiveness
PE	Performance rate
PM	Preventive maintenance
Q	Quality rate
SIOPC	Supplier-input-process-output-customer
SMED	Single minute exchange of dies

B. Alsubaie • Q. Yang (✉)
Brunel University London, Uxbridge, UK
e-mail: barrak888@hotmail.com; QingPing.Yang@brunel.ac.uk

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_2

TPM Total productive maintenance
VOC voice of customer

1 Introduction

Maintenance management refers to the process of scheduling and allocating resources to the maintenance activities (repair, replacement and preventive maintenance) [1]. The leading objective of the maintenance function in any organization is to maximize asset performance and optimize the use of maintenance resources. The implementation of current maintenance management systems has not reached the expected level of success (e.g. maintenance schedules are not implemented on time, and priorities are difficult to identify) [2]. The underlying reason is the lack of maintenance management skills and practical experience, which leads to poor impacts and negative effects on performance [2]. Unnecessary repair or inspection will increase maintenance budget commitments and may decrease quality performance, as described by [3] concerning the wastes in the maintenance area. These issues indicate that maintenance processes have nonvalue-adding steps that need continual improvement.

The challenge of "designing" the ideal model to drive maintenance activities according to [4] has become a research topic and a major question for attaining effectiveness and efficiency in maintenance management and achieving enterprise objectives. This study has been carried out based on a maintenance division which is responsible for maintenance of a fleet of military vehicles. The maintenance division has been facing ever-increasing military expenses to maintain military readiness with aging vehicle fleet systems. Hence, the division is keenly interested in finding a suitable model with practical guidelines for the maintenance providers to improve the service processes. Whilst various authors have proposed what they consider as the best practices or models for maintenance management, this study emphasizes the integration of the state-of-the-art approaches in process improvement for the effective and efficient management of the vehicle fleet maintenance, as presented in this paper.

2 Model Strategies

2.1 Integration of Six Sigma and Lean

Lean Six Sigma combines lean methods and Six Sigma, using specific DMAIC processes to provide companies with better speed and lower variability to increase customer satisfaction [5]. The first phase in DMAIC process is to define project objectives and customer needs. The second phase is to measure the current process performance as well as quantifying the problems. The third phase is to analyse the

process and find the causes of problems, particularly the root causes. The fourth phase is to improve the process, i.e. correcting the causes of defects and reducing process variability. The final phase is to control the process and maintain the improved performance. These five phases can assist Lean Six Sigma teams to systematically and gradually develop process rationalisation, starting with defining the problem and then introducing solutions targeted to the fundamental causes, so constructing the optimal implementation method and ensuring the sustainability of solutions [6]. This approach has gained increasing recognition in process improvement practices.

2.2 TPM

Total Productive Maintenance (TPM) may be defined as an innovative approach to maintenance that improves equipment effectiveness, eliminates breakdowns, and supports autonomous maintenance by operators through day-to-day activities including the total workforce [7]. TPM is a maintenance management program with the objective of reducing equipment downtime and improving overall equipment effectiveness [8]. Nevertheless, TPM is not a maintenance specific policy; it is a culture, a philosophy and a new attitude for maintenance. The effective adoption and implementation of strategic TPM initiatives in the manufacturing organizations is a strategic approach to improve the performance of maintenance activities [9]. TPM brings maintenance into focus as a crucial and very important part of the business. TPM seeks to engage all levels and functions in an organization to maximize the overall effectiveness of production equipment.

2.3 Integration of TPM and Lean Six Sigma

This study has proposed an integrated approach of TPM with LSS to reach world class maintenance performance, with the core model shown in Fig. 1. Lean Six Sigma forms the basic foundation for the TPM strategy and makes it easier to understand by shop floor operators who are the most important enablers of successful TPM implementation. Within the five phases of DMAIC, various problems and sub-processes of the maintenance department are defined, the process performance is measured, the most important causes of the defects or non-conformities are identified and analyzed, improvement or corrective actions are then taken with the improvements sustained by standardisation and continuing process control. Moreover, the iterative process of DMAIC is used as the main operational approach for the implementation of this model in order to achieve continual improvement of maintenance activities and ultimately to reach world class performance in terms of both sigma level and overall equipment effectiveness. The implementation of this model or approach will be supported with a rich collection of tools from Six Sigma, lean, TPM, quality control and problem solving practices.

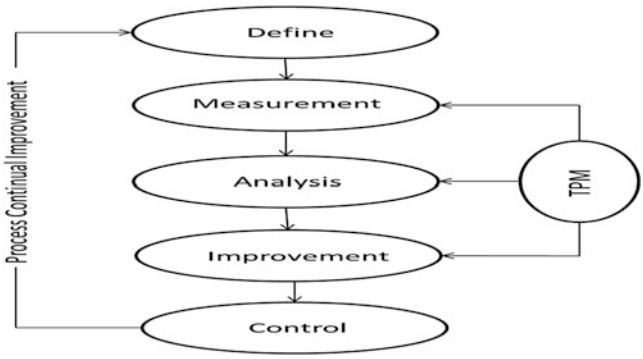


Fig. 1 Methodology to develop integrated model

Table 1 Key activities and tools of implementing the TPM oriented LSS maintenance model

Stage	Activities	Tools
1. Define	• Build process improvement team • Identify problems & weaknesses of the process • Select CTQ characteristics	• SIOPC • Brainstorming • VOC • Pareto analysis
2. Measure	• Select measuring system • Gather information about key maintenance processes • Calculate the current OEE	• Process map • TPM • OEE
3. Analyse	• Identify root causes of problems • Implement basic levels of TPM • Identify improvement opportunities	• Cause and effect diagram • TPM
4. Improve	• Propose solutions and implement changes for maintenance improvement • Evaluate the process performance • Calculate the new OEE	• Seven Wastes • SMED • Poka—yoke • 5S • TPM
5. Control	• Standardize the best practices • Integrate the changes to the organisation knowledge base • Continual improvement	• SPC • Performance management • Education and training

In any process improvement project, utilization of a well-defined improvement procedure is critically important. The procedure and key activities of the TPM oriented Lean Six Sigma can be summarized in Table 1, under the DMAIC phases, together with typical tools.

3 Case study

In order to test the proposed TPM oriented LSS model for the vehicle maintenance, a case study was performed for the engine maintenance process, since engine for a vehicle is as vital as the heart of a human being and its maintenance is essential. As demands for service quality and cost reduction in vehicle maintenance have both increased in recent years, the effectiveness of a maintenance system for engines has become an important issue. Engines are subject to deteriorations in relation to both usage and ageing, which leads to reduced product quality and increased maintenance costs. The maintenance division executes preventive maintenance (PM) on engines to prevent or slow down such deteriorations.

3.1 Define

Step_D1: The project started with Define phase that gives a clear problem definition using the supplier-input-process-output-customer (SIPOC) tool. This tool describes the step-by-step process for the engine maintenance as shown in Fig. 2. The first process is the engine service or maintenance. The input to this process includes the engine to be serviced, parts and preventive maintenance program and procedure, whilst the supplier is the maintenance crew. The output of this process is engine serviced, the customer is the field service unit. The second process is repair and replacement of engine. The inputs to this process are operation notification and work order, the supplier is the field service unit. The output of this process is engine repaired or replaced and the customer is the field service unit.

Step_D2: The engine preventive maintenance (PM) being analysed is verified to be significant by the field study. Engine PM cost represents a high percentage of vehicle PM cost. The team members participate in brainstorming sessions to identify critical to quality characteristics (CTQ) based on the voices of customer input. Also, the component(s) failure that results in high machine downtime or cost (due to machine breakdown) is classified as critical components. Critical engines failures have been reported for the engines in the field study which causes

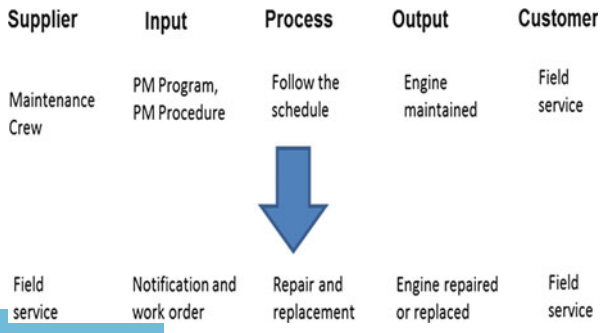


Fig. 2 Two example SIPOC processes

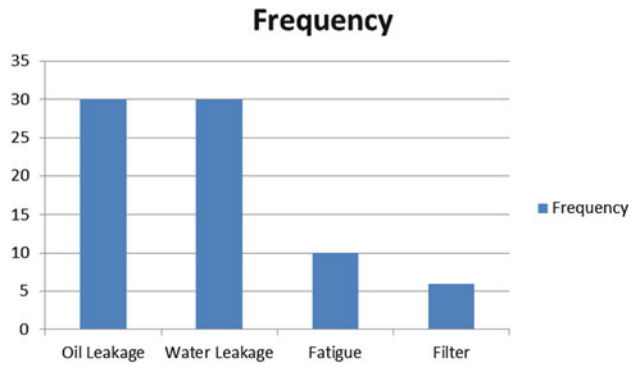


Fig. 3 Pareto analysis of engine failures

significant cost of the PM and also deviations from the customer satisfaction targets. The project was scoped down to oil and water leakage since they contribute to about 60% of the total failure cost as determined through the use of Pareto analysis, as shown in Fig. 3.

3.2 Measurement

Step_M1: To measure the factors that contribute to the process and failures on the subject engine, a number of tools from the Six Sigma toolbox are used such as process mapping and fishbone diagram. The process map (Fig. 4) provides a visual view of all maintenance and operation steps that take place from the time an engine failure is detected through putting it back to service all the way to operation and monitoring until it fails again.

Step_M2: Since the CTQ characteristics, i.e. oil and coolant leakage, are identified in the Define phase, a data collection plan needs to be developed. The measurement system should be examined prior to data collection. In this case, the existing service report is used to facilitate the collection of primary data. Monthly reporting is particularly useful in monitoring the maintenance tasks performed by the maintenance personnel and calculating the maintenance cost. Also, each vehicle has its own maintenance history book to record the repairs/replacement done to it. From these records, the data on the maintenance history of the engines can be extracted. To quantify the problem, data gathering was initiated on the failures costs of engines.

Step_M3: For a specific CTQ characteristic, the sigma level can be calculated from DPMO (defect per million opportunities) as:

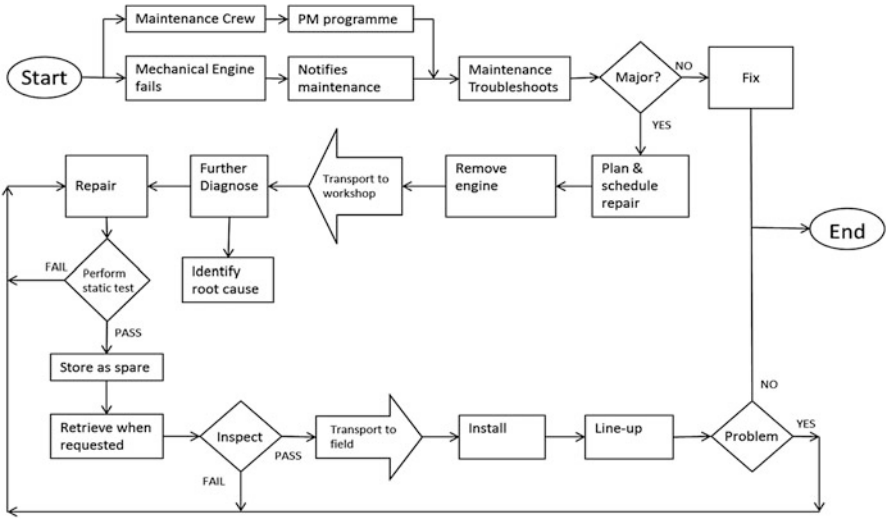


Fig. 4 Process map of engine maintenance

Table 2 Initial process capability

CTQ	No. of units	No. of opportunities	No. of defects	DPMO	Sigma level	C_{pk}
Oil leakage	1000	7	30	4285	2.45	1.4
Coolant leakage	1000	3	30	10,000	2.3	1.2

$$DPMO = \frac{total\ number\ of\ defects}{number\ of\ units\ x\ number\ of\ opportunities}$$

The process capability indices C_{pk} and the corresponding sigma levels are summarised in Table 2. The sigma level of a process can be used to express its capability as to how well it performs with respect to specifications.

3.3 Analysis

Step_A1: To ascertain the root cause(s) of key engine failures, an analysis using the cause-and-effect diagram is therefore carried out during a brainstorming session of the LSS team. Figure 5 shows the root causes of the engine failure problems.

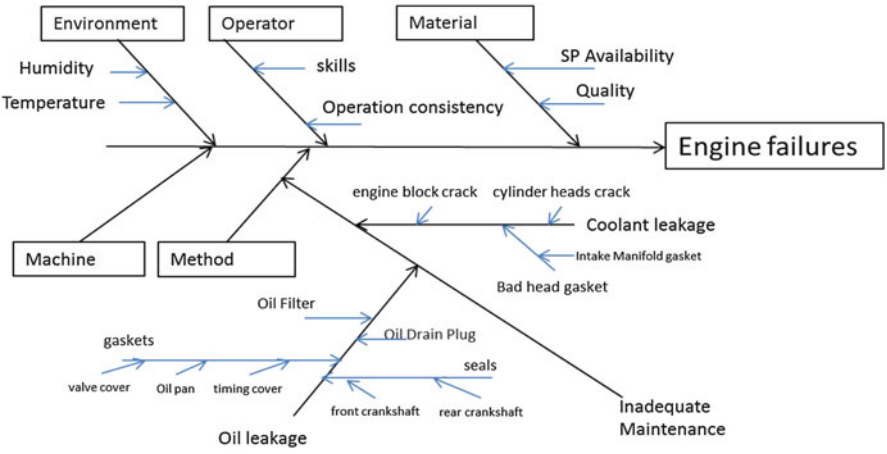


Fig. 5 Fishbone diagram for engine failures

Table 3 Initial OEE

Process	*A%*	*PE%*	*Q%*	*OEE%*
Engine repair	86	80	97	66
world-class performance	90	95	99	85

Step_A2: According to [8], OEE measurement is an effective way of analysing the efficiency of a single machine or an integrated system. It is a function of availability, performance rate and quality rate, and can be expressed as follows:

$$\text{OEE} = \text{Availability (A)} \times \text{Performance rate (PE)} \times \text{Quality rate (Q)}$$

On average the engine maintenance workshop can complete 20 engines monthly. The records have shown the number of defective engines for both causes (oil and coolant leakages) was 7 annually. Hence, the quality rate (Q) which is the percentage of the working engines out of the total produced can be calculated as 97%. The maintenance workshop normally runs for 30 days with 4 days break scheduled, so the Planned Maintenance Time is 26 days. On average, 4 days will be lost in maintenance each month due to unavailable parts or equipment, and the Operating Time is thus 22 days per month, with an availability (A) of 22/26≈85%. The standard cycle time for the engine maintenance is 25 units/month or 0.88 days/unit. As the workshop can actually complete 240 units during the year or 20 units per month, which gives the actual cycle time of 22 days/20 unit = 1.1 days/unit. The performance rate (PE) is thus 0.88/1.1 = 80%. The initial OEE is therefore about 66%, well below the word class performance (Table 3).

3.4 Improvement

Step_I1: Four levels of maintenance have been implemented in the maintenance division. Level 1 is carried out by the autonomous maintenance teams (drivers or operators). These teams apply basic maintenance, including regular daily cleaning regimes, as well as sensory maintenance tasks (smell, sound, sight, touch, etc.). Level 2 typically involves simple repairs or replacement of components. Level 3 involves more difficult repairs and maintenance, including the repair and testing of components that have failed at the Level 2, and Level 3 maintenance is carried out by the maintenance department, as it is beyond the capabilities of the lower levels, usually requiring major overhaul or rebuilding of end-items, subassemblies, and parts. Level 4 involves the engineering department, becoming more proactive in the development of PM practices, including machine modification and enhancement strategies that allow easier maintenance, among others. Level 4 tasks also entail monitoring maintenance activities and are directed primarily at approaches to increase the MTBF to achieve a higher degree of machine availability. The aim here is to extend the MTBF so that the machinery can remain productive longer, thus providing a greater return on machine performance.

Step_I2: This step is concerned with the implementation of TPM at field study organization. Various pillars of TPM i.e. 5S, Jishu Hozen, Kobetsu Kaizen, Planned Maintenance and OEE have been implemented, as shown in Fig. 6.

(a) 5S: Making problems visible is the first step of improvement. 5S are defined as Sort, Set in Order, Shine, Standardize and Sustain. Table 4 shows some applications of this tool in maintenance process.
(b) Jishu Hozen: it is also called autonomous maintenance. The operators are responsible for keeping their equipment to prevent it from deteriorating.
(c) Kobetsu Kaizen: Kaizen involves small improvements and is carried out on a regular basis, involving people of all levels in the organization. A detailed and thorough procedure is followed to eliminate losses systematically using various Kaizen tools as follows:

- Poka Yoke devices: It is Japanese term in English which means mistake proofing or error prevention. Poka Yoke devices have been developed and used in-house.
- Leakage problem: To identify the reasons for a leakage, a fishbone diagram is prepared, as shown in Fig. 7.

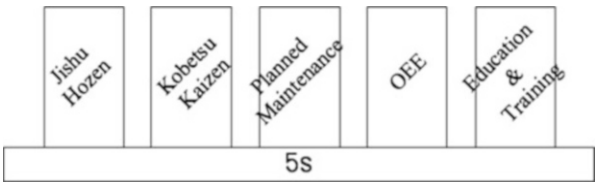


Fig. 6 Pillars of TPM

Table 4 Implementation of 5S

5S	Before	After
Sort	• Rejected parts are kept inside the workshop.	• The parts are removed and the space is freed.
Set in Order	• Earlier patches on the floor disturb material movement using trolley. • Tools are placed randomly in racks and no labelling is done.	• Patches are filled with cement thus helping smooth material flow. • Tools are stored in their respective places identified with labelling.
Shine	• Work place not very tidy and clean.	• Clean and tidy work place.
Standardize	• No operator report is kept. • Operator details are not displayed on the notice board.	• Writing hourly report is compulsory. • Operator details are displayed on the notice board.
Sustain		• Organisation mission and vision statements are displayed in Arabic as well as English. • Suggestion scheme stating that whoever gives the best suggestion will receive a prize.

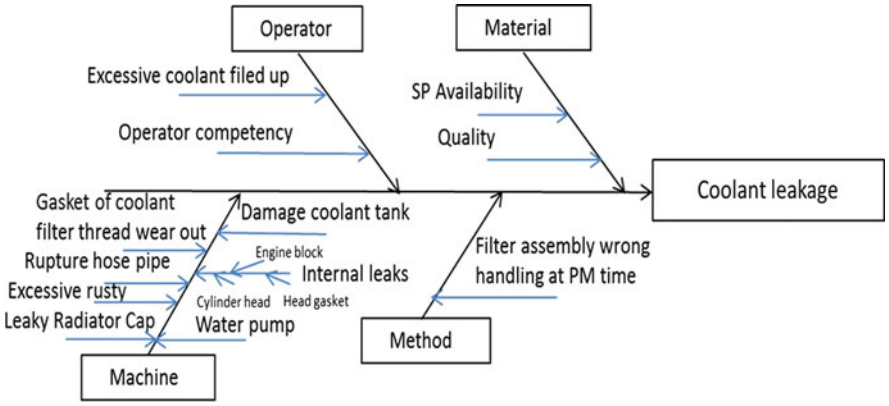


Fig. 7 Fishbone diagram for coolant leakage

- New Layout: A new layout is proposed as shown in Fig. 8. The proposed layout is designed to minimize the handling of parts.

(d) Education and training: TPM education and training programs have been prepared to achieve three objectives:

- Managers will learn to plan for higher equipment effectiveness and implement improvements intended at achieving zero breakdowns and zero defects.
- Maintenance staff will study the basic principles and techniques of maintenance and develop specialized maintenance skills.

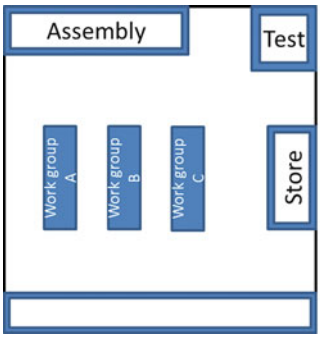


Fig. 8 Layout of engines workshops

Table 5 OEE improvement of engine repair process

	A%	*PE%*	*Q%*	*OEE%*
Initial OEE	86	80	97	66
Improved OEE	92	87	98.5	79
world-class performance	90	95	99	85

- Drivers and maintenance staff will learn how to identify abnormalities as such during their daily and periodic inspection activities.

(e) Planned Maintenance: It is aimed to have trouble free vehicles without any breakdown and ensure components at good quality level giving total customer satisfaction.

(f) OEE is calculated after the implementation. Based on the initial assessment, the availability has increased to 92%, the performance rate 87% and the quality rate 98.5%, with an overall OEE of 79%. Whilst this is still below the world class 85% performance, it has significantly improved the initial OEE of 66% (Table 5). Continual improvement is required to reach the world class performance.

3.5 Control

The Control phase includes the following activities:

- Management of processes of change;
- Documentation and standardization of the improved maintenance process;
- Monitoring of the maintenance process through control charts;
- Identifying opportunities for further improvement of the maintenance process.

4 Discussions and Conclusions

A new model based on TPM and Lean Six Sigma has been presented to provide guidance and support for service organisations who aim to reach world-class standards in maintenance processes through continual improvement. The application of the model has been illustrated using a case study in maintenance of a fleet of military vehicles. The above discussions are largely based on the case study, particularly the engine repair process due to the water and oil leakage problems. However, this approach is very generic in nature and can be applied to any other maintenance or repair process (e.g. engine repair due to heavy friction), and to any service organisations with maintenance functions. Of course, the complexity of the proposed model will depend on the application since the nature and the number of CTQs are application specific. The proposed model as the framework together with the use of common tools emphasize the process approach and will therefore be generally applicable in such service organisations. Use of this model will likely help to achieve high process performance and overall equipment effectiveness. The model also provides a good framework and methodology to continually improve the maintenance performance.

References

1. Cassady CR, Murdock WP, Nachlas J, Pohl E (1998) Comprehensive fleet maintenance management. In: IEEE international conference on systems, man, and cybernetics, vol 5. IEEE, New York, pp 4665–4669
2. Aldairi J, Khan M, Munive-Hernandez J (2015) A conceptual model for a hybrid knowledge-based Lean Six Sigma maintenance system for sustainable buildings. In: Proceedings of the World Congress on Engineering, vol 2
3. Milana M, Khan MK, Munive JE (2014) A framework of knowledge based system for integrated maintenance strategy and operation. In: Applied mechanics and materials. Trans Tech Publications, Zurich, pp 619–624
4. Uday K et al (2009) The maintenance management framework: a practical view to maintenance management. J Qual Maint Eng 15(2):167–178
5. Wang H, Pham H (2006) Reliability and optimal maintenance. Springer, Berlin
6. Cheng C, Chang P (2012) Implementation of the Lean Six Sigma framework in non-profit organisations: a case study. Total Qual Manag Bus Excell 23(3–4):431–447
7. Nakajima S (1989) TPM development program: implementing total productive maintenance. Productivity Press, Cambridge
8. Chaneski W (2002) Total productive maintenance–an effective technique. Mod Mach Shop 75 (2):46–48
9. Ahuja IPS, Khamba JS (2008) Total productive maintenance: literature review and directions. Int J Qual Reliab Manage 25(7):709–756

Group Replacement Policies for Repairable *N*-Component Parallel Systems

Chuan-Wen Chiu, Wen-Liang Chang, and Ruey-Huei Yeh

Abstract For a multiple component parallel system, the operation of each component is independent, that is, any failure of component will not affect the normal operation of the system. Due to the inevitable deterioration of the component, it may fail more frequently when its age increases. In order to reduce the number of failures, the component should be replaced at a certain age. When a replacement action of the component is performed, it will incur a replacement cost and a setup cost. Under this situation, grouping replacement for all components is an important policy in order to reduce setup cost. This paper derives a formula of the grouping replacement for a *n*-component parallel system. Further, an optimal grouping replacement method is offered, and the optimal replacement time of each group is obtained.

Keywords Grouping replacement policy • Group replacement time • Parallel system • Setup cost

1 Introduction

With advances in technology, customer demand for products have also increased, making the production system relatively complex. In order to maintain the stable operation of the system, maintenance policies should be planned and performed, especially when a system consists of multiple components. In general, multiple components in a system may be connected in series or parallel. When multiple components are connected in series, any failure of either one component will cause system downtime. However, when multiple components are connected in parallel, the failure of any component will not affect the operation of the other component.

In practice, when a component fails during the operation period, the failed component may be rectified by minimal repair, imperfect repair, perfect repair, or

C.-W. Chiu (✉) • R.-H. Yeh
National Taiwan University of Science and Technology, Taipei City, Taiwan
e-mail: M10401015@mail.ntust.edu.tw; rhyeh@mail.ntust.edu.tw

W.-L. Chang
Cardinal Junior Tien College of Healthcare and Management, New Taipei City, Taiwan
e-mail: chang@ctcn.edu.tw

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_3

replacement. Due to the inevitable deterioration of the component, it may fail more frequently when its age increases. In order to reduce the number of failures, the component should be replaced by a new component at a certain age. In this paper, two maintenance policies for components are considered: (1) When the component fails during the operating period, the failed component is corrected by minimal repair and (2) The component is replaced at a pre-specified time under normal operation.

For a system that consists of multiple components in parallel, all components or some components may be treated as one group and an optimal replacement policy for the group can be derived. Under this situation, it is called group replacement policy. On the other hand, we may focus on each component and find the optimal replacement policy for each component. It is called individual replacement policy. However, either way may not be optimal when the maintenance actions of the components are cost-dependent. For example, minimal repair may be carried out easily, but replacements should be performed by a professional technician team. When a replacement action is performed, there is a fix setup cost, including the downtime cost and the cost of hiring a professional team to install the components. Furthermore, the setup cost may be the same for replacing a single component or a group of components. In this paper, we focus on investigating the influences of setup cost on the grouping replacement policy for a parallel system, which consists of multiple components in parallel. We derive the grouping models of multiple components and analyze the influences of setup cost for the number of grouping of multiple components and the replacement time of each group through numerical examples.

In the past, many research works have been published on the repair and replacement policies of a single component or system. In general, the maintenance policies for repairable products can be divided into two main categories: repair and replacement. About repair policy, Barlow and Hunter [1] proposed the minimal repair concept and applied it in the reliability area. Minimal repair refers to a failed device is restored to its normal operation, and its failure rate remains the same as that just prior to failure. Nakagawa and Kowada [2] showed that the failure process of a system during the operating period follows a Non-homogeneous Poisson Process (NHPP) when the minimal repair is carried out. Chien et al. [3] considered a preventive replacement model with minimal repair and a cumulative repair-cost limit by introducing the random lead time for replacement delivery.

About replacement policy, Berg and Epstein [4] gave a rule for choosing the least costly of the above three policies (age replacement policy, block replacement policy and failure replacement policy) under conditions specified. Beichelt [5] introduced a generalized block replacement policy and derived the long-run cost rate and integral equations of the renewal type for the basic reliability expressions. Finally, a numerical example is illustrated. Chen and Feldman [6] studied a modified minimal repair/replacement problem that is formulated as a Markov decision process and assumed that the operating cost is an increasing function of the age of the system. Further, a computational algorithm for the optimal policy is suggested based on the total expected discounted cost. Sheu and Griffith [7] considered a generalized age-replacement policy with age-dependent minimal repair and random leadtime and developed a model for the average cost per unit time. Further, determination of the minimum-cost policy time is described and illustrated with a

numerical example. Chien and Sheu [8] supposed a system can have two types of failure: type I failure (minor) or type II failure (catastrophic) and proposed a generalization of the age replacement policy for such a system. Finally, some numerical examples are illustrated and analyzed.

About grouping replacement policy, Shafiee and Finkelstein [9] investigated an optimal age-based group maintenance policy for a multi-unit series system where the components are subjected to gradual degradation. Zhao et al. [10] proposed several approximate models for optimal replacement, maintenance, and inspection policies. However, most of the literature have not considered the case when the components are cost-dependent in performing the maintenance actions. This paper is organized as follows. The grouping models of multiple components are constructed and the optimal replacement time of each group is obtained in Sect. 2. The Weibull case is illustrated in Sect. 3. In Sect. 4, the influences of setup cost on the number of grouping of multiple components are illustrated through numerical examples. Finally, some conclusions are drawn in Sect. 5.

2 Mathematical Models

Consider a parallel system comprises n components (P_i, $i = 1, 2,\ldots, n$) as in Fig. 1. Due to the inevitable deterioration of the component, it may fail more frequently when its age increases. In order to reduce the number of failures, the components should be replaced at a certain age. For the replacement of n components P_i, $i = 1, 2,\ldots, n$, all components can be divided into k ($1 \le k \le n$) groups replacement. For example, if $n = 4$, the maximal number of group is four (i.e., four components are divided into $k = 4$ groups), the minimal number of group is one (i.e., four components are divided into $k = 1$ group). In Table 1, all possible results for the grouping method of four components and number of components of each group are shown. When $k = 2$ (i.e., four components are divided into 2 groups), the number of components of each group is 1 and 3 or 2 and 2. For four components within 2 groups, all possible combination results are seven as Table 2.

To construct a cost rate model of a n-component parallel system for grouping replacement of the components, the mathematical notations used in this paper are summarized as follows.

$f_i(t)$ lifetime distribution of the ith component, for $i = 1, 2,\ldots, n$.
$h_i(t)$ failure rate function of the ith component, for $i = 1, 2,\ldots, n$.
$H_i(t)$ cumulative failure rate function of the ith component; $H_i(t) = \int_0^t h_i(u)\ du$, for $i = 1, 2,\ldots, n$.
C_{mi} minimal repair cost of ith component, for $i = 1, 2,\ldots, n$.
C_{ri} replacement cost of ith component, for $i = 1, 2,\ldots, n$.
C_s setup cost for performing a replacement
T_{gj} group replacement time of group j, for $j = 1, 2,\ldots, k$ and $1 \le k \le n$.

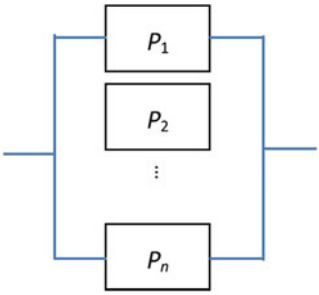


Fig. 1 n-component parallel system

Table 1 The grouping methods for four components

Groups	Number of components in each group
$k = 4$ groups (g_1, g_2, g_3, g_4)	(1, 1, 1, 1)
$k = 3$ groups (g_1, g_2, g_3)	(1, 1, 2)
$k = 2$ groups (g_1, g_2)	(1, 3) or (2, 2)
$k = 1$ group (g_1)	(4)

Table 2 The combinations for 2 groups ((1, 3) or (2, 2))

	Number of groups ($k = 2$ groups (g_1, g_2))	
Number of components	(1, 3)	(2, 2)
Number of combinations	$\frac{n!}{Y_1!Y_2!} = \frac{4!}{1!3!} = 4$	$\frac{4!}{2!2!} = 6 \div 2 = 3$
Combination of components	($\{P_1\}$,$\{P_2$,P_3,$P_4\}$)	($\{P_1$,$P_2\}$,$\{P_3$,$P_4\}$)
	($\{P_2\}$, $\{P_1$, P_3, $P_4\}$)	($\{P_1$,$P_3\}$,$\{P_2$, $P_4\}$)
	($\{P_3\}$, $\{P_1$,P_2, $P_4\}$)	($\{P_1$,$P_4\}$,$\{P_2$, $P_3\}$)
	($\{P_4\}$, $\{P_1$, P_2, $P_3\}$)	

2.1 Cost Rate Models

For a n-component parallel system, all components are divided into k ($1 \le k \le n$) groups replacement and the replacement time of the jth group is T_{gj} for $j = 1, 2, \ldots, k$ and $1 \le k \le n$. When a replacement action of the jth group is performed, it will incur a replacement cost $\sum_{i \in g_j} C_{ri}$ and a setup cost C_s, which includes the downtime cost and the cost of hiring a professional team to install the components. Therefore, the average replacement cost rate of system is $\sum_{j=1}^{k} \left[\sum_{i \in g_j} C_{ri} / T_{g_j} + C_s / T_{g_j} \right]$, for $1 \le k \le n$.

Since n components are connected in parallel, any failure of component will not affect the normal operation of the system. Within the group replacement time T_{gj} for $j = 1, 2, \ldots, k$ and $1 \le k \le n$, any failure of the component is rectified by minimal repairs and incurs a fixed repair cost C_{mi}, for $i = 1, 2, \ldots, n$. Let the failure rate function and cumulative failure rate function of the component P_i are $h_i(t)$ and $H_i(t)$, for $i = 1, 2, \ldots, n$, respectively. Due to the inevitable deterioration of the component,

the failure rate function $h_i(t)$ is an increasing function of t. Since a failed component is rectified by minimal repairs, the failure process of the component within the operating period is a nonhomogeneous Poisson process with the failure rate function $h_i(t)$. Under this repair policy, the expected repair cost rate of system within group replacement time T_{g_j} is $\sum_{j=1}^{k} \sum_{i \in g_j} C_{mi} H_i\left(T_{g_j}\right) / T_{g_j}$, for $1 \le k \le n$.

For example, suppose that $n = 4$ components, $k = 2$ groups, and the number of components of each group is 2. The components of the 1st and the 2nd groups are $\{P_1, P_3\}$ and $\{P_2, P_4\}$, respectively (i.e., $g_1 = \{P_1, P_3\}$ and $g_2 = \{P_2, P_4\}$). Under this grouping method for 4 components, the average replacement cost rate of system is $(C_{r1}+C_{r3}+C_s)/T_{g1} + (C_{r2}+C_{r4}+C_s)/T_{g2}$. The expected repair cost rate of system is $[C_{r1}H_1(T_{g1}) + C_{r3}H_3(T_{g1})]/T_{g1} + [C_{r2}H_2(T_{g2}) + C_{r4}H_4(T_{g2})]/T_{g2}$. Under above description of replacement and repair policies, the excepted total average cost rate of system is

$$E\left[TC\left(T_{g_1}, \ T_{g_2}, \ \ldots, T_{g_k}\right)\right] = \sum_{j=1}^{k} \sum_{i \in g_j} A_i\left(T_{g_j}\right) + \sum_{j=1}^{k} \frac{C_s}{T_{g_j}} \tag{1}$$

where $A_i\left(T_{g_j}\right) = \frac{C_{mi} H_i\left(T_{g_j}\right) + C_{ri}}{T_j}$ if component $i \in g_j$; otherwise, $A_i\left(T_{g_j}\right) = 0$.

The objective is to find the optimal group replacement time T_{gj} in Eq. (1). The optimal group replacement time is derive in Sect. 2.2.

2.2 *Optimal Group Replacement Time*

From Eq. (1), the excepted average cost rate of the jth group for $j = 1, 2, \ldots, k$ and $1 \le k \le n$ group can be obtain as follows.

$$E\left[TC\left(T_{g_j}\right)\right] = \frac{\sum_{i \in g_j} \left[C_{mi} H_i\left(T_{g_j}\right) + C_{ri}\right] + C_s}{T_{g_j}} \tag{2}$$

To find the optimal replacement time T_{g_j}, we can take the first derivative of Eq. (2) with respect to T_{g_j} and then the result is obtained as follows.

$$\frac{dE\left[TC\left(T_{g_j}\right)\right]}{dT_{g_j}} = \frac{\left(\sum_{i \in g_j} \left\{C_{mi}\left[T_{g_j} h_i\left(T_{g_j}\right) - H_i\left(T_{g_j}\right)\right] - C_{ri}\right\}\right) - C_s}{T_{g_j}^2} \tag{3}$$

Setting Eq. (3) equals to zero, the optimal replacement time $T_{g_j}^*$ of the jth group is obtained by solving the equation

$$\sum_{i\in g_j}\left\{C_{mi}\left[T_{g_j}h_i\left(T_{g_j}\right)-H_i\left(T_{g_j}\right)\right]-C_{ri}\right\}=C_s \tag{4}$$

Substituting $T^*_{g_j}$ into Eq. (4), Eq. (4) can be expressed as

$$\sum_{i\in g_j}\left[C_{mi}H_i\left(T^*_{g_j}\right)+C_{ri}\right]+C_s=\sum_{i\in g_j}C_{mi}T^*_{g_j}h_i\left(T^*_{g_j}\right) \tag{5}$$

Substituting Eq. (5) into Eq. (2), the expected average cost rate of the jth group can be rewritten as

$$E\left[TC\left(T^*_{g_j}\right)\right]=\frac{\sum_{i\in g_j}C_{mi}T^*_{g_j}h_i\left(T^*_{g_j}\right)}{T^*_{g_j}} \tag{6}$$

From Eq. (6), the expected average cost rate of system can be obtained as follows.

$$E\left[TC\left(T^*_{g_1},\ T^*_{g_2},\ \ldots,\ T^*_{g_k}\right)\right]=\sum_{j=1}^{k}\left(\frac{\sum_{i\in g_j}C_{mi}T^*_{g_j}h_i\left(T^*_{g_j}\right)}{T^*_{g_j}}\right) \tag{7}$$

The Weibull case is illustrated for Eq. (5) in the next section.

3 Weibull Case

Consider that the lifetime distribution of the components P_i follows a Weibull distribution with a scale parameter α_i and a shape parameter β_i, for $i = 1, 2, 3, \ldots, n$. It is well-known that the probability density function of a Weibull distribution is $f_i(t)=\alpha_i\beta_i(\alpha_i t)^{\beta_i-1}e^{-(\alpha_i t)^{\beta_i}},\ t\geq 0$. By definition the failure rate function of a Weibull distribution is $h_i(t)=\alpha_i\beta_i(\alpha_i t)^{\beta_i-1}$ and the cumulative failure rate function is $H_i(t)=(\alpha_i t)^{\beta_i}$. Note that $h_i(t)$ is an increasing function of t if $\beta_i > 1$. For the Weibull case, substituting $h_i(t)$ and $H_i(t)$ into Eq. (5), the result is

$$\sum_{i\in g_j}\left[C_{mi}(\beta_i-1)\left(\alpha_i T^*_{g_j}\right)^{\beta_i}-C_{ri}\right]=C_s \tag{8}$$

Observing Eq. (8), there is no closed-form solution for $T^*_{g_j}$, unless $\beta_i = \beta$ for $i = 1, 2, \ldots, n$. When $\beta_i = \beta$ for $i = 1, 2, \ldots, n$., the optimal group replacement time $T^*_{g_j}$ the jth group for $j = 1, 2, \ldots, k$ and $1 \leq k \leq n$ can be obtained as follows.

$$T^*_{g_j} = \left[\frac{\left(\sum_{i \in g_j} C_{ri} \right) + C_s}{(\beta - 1) \sum_{i \in g_j} C_{mi} \alpha_i^{\beta}} \right]^{\frac{1}{\beta}} \tag{9}$$

Based on Eq. (9), when n components are divided into 1 group (i.e., $k = 1$), the optimal group replacement time $T^*_{g_1}$ is

$$T^*_{g_1} = \left[\frac{\left(\sum_{i=1}^{n} C_{ri} \right) + C_s}{(\beta - 1) \sum_{i=1}^{n} C_{mi} \alpha_i^{\beta}} \right]^{\frac{1}{\beta}} \tag{10}$$

When n components are divided into n groups (i.e., $k = n$), the optimal group replacement time $T^*_{g_j}$ for $j = 1, 2, \ldots, n$ is

$$T^*_{g_j} = \left[\frac{C_{rj} + C_s}{(\beta - 1) C_{mj} \alpha_j^{\beta}} \right]^{\frac{1}{\beta}} \tag{11}$$

The performance of the optimal replacement policy is evaluated and the properties of grouping replacement policy are illustrated through numerical examples in chapter 4 "Low Speed Bearing Condition Monitoring: A Case Study".

4 Numerical Examples

In this section, the performances of the optimal group replacement policies are evaluated through numerical examples of Weibull cases. Suppose that 5 components are divided into k ($1 \leq k \leq 5$) groups. The parameter values are considered for 5 components and components are arranged according to mean lifetime $\mu_i = (1/\alpha_i)\Gamma(1 + 1/\beta_i)$ in Table 3.

For all possible grouping methods of 5 components are showed in Table 4. For example, 5 components are divided into $k = 2$ groups. The number of components of each group is 3 and 2 or 4 and 1. When the number of components of each group is 4 and 1, the optimal grouping method of components is $\{P_1, P_2, P_3, P_4\}$ and $\{P_5\}$. The optimal replacement time of $\{P_1, P_2, P_3, P_4\}$ and $\{P_5\}$ are 9.8 and 56.1, and the

Table 3 parameter values of the model for 5 components

	α_i	β_i	C_{ri}	C_{mi}	C_s
P_1	0.30	2.00	1160	330	800
P_2	0.26	2.00	850	150	
P_3	0.15	2.00	540	160	
P_4	0.10	2.00	1180	340	
P_5	0.06	2.00	1470	200	

Table 4 Optimal grouping policy of $n = 5$ components

	Components of each group j					Replacement time of each group T_j					
k	1	2	3	4	5	T_1	T_2	T_3	T_4	T_5	$E(TC)$
5	$\{P_1\}$	$\{P_2\}$	{P3}	{P4}	$\{P_5\}$	8.1	12.7	19.2	24.1	56.1	1125.1
4	$\{P_i\}_{i=1}^{2}$	{P3}	$\{P_4\}$	$\{P_5\}$		8.4	19.2	24.1	56.1		1053.0
3	$\{P_i\}_{i=1}^{3}$	$\{P_4\}$	{P5}			8.7	24.1	56.1			1007.9
3	$\{P_i\}_{i=1}^{2}$	$\{P_i\}_{i=3}^{4}$	{P5}			8.4	18.9	56.1			1015.6
2	$\{P_i\}_{i=1}^{3}$	$\{P_i\}_{i=4}^{5}$				8.7	28.9				1001.4
2	$\{P_i\}_{i=1}^{4}$	$\{P_5\}$				9.8	56.1				1002.1
1	$\{P_i\}_{i=1}^{5}$					11.2					1068.3

Table 5 The excepted average cost rate of $k = 1, 2$ groups

	$k = 2$ ($\{P_1, P_2, P_3\}$,$\{P_4\}$, $\{P_5\}$)		$k = 2$ ($\{P_1, P_2, P_3, P_4\}$, $\{P_5\}$)		$k = 1$ ($\{P_1, P_2, P_3, P_4, P_5\}$)	
C_s	(T_1,T_2)	$E(TC)$	(T_1,T_2)	$E(TC)$	T_1	$E(TC)$
800	(8.7,28.9)	1001.4	(9.8,56.1)	1002.1	11.2	1068.3
1100	(9.1,30.1)	1044.9	(10.1,59.4)	1037.3	11.5	1094.7
1400	(9.5,31.3)	1086.8	(10.4,63.1)	1071.3	11.7	1120.5
1700	(9.8,32.4)	1127.0	(10.7,66.3)	1104.1	12.0	1145.7
2000	(10.2,33.6)	1165.9	(11.0.69.4)	1136.1	12.3	1170.3
2300	(10.5,34.6)	1203.6	(11.3,72.3)	1167.1	12.5	1194.4
2600	(10.8,35.7)	1240.1	(11.6,75.1)	1197.3	12.8	1218.1
2900	(11.2,36.7)	1275.5	(11.9,77.9)	1226.7	13.0	1241.3
3200	(11.5,37.6)	1310.0	(12.1,80.5)	1255.4	13.2	1264.1
3500	(11.8,38.6)	1343.6	(12.4,83.0)	1283.5	13.5	1286.5
3800	(12.0,39.5)	1376.4	(12.6,85.5)	1310.9	13.7	1308.5

expected total average cost rate of the system is 1002.1. From Table 4, for all possible grouping methods of 5 components, the optimal grouping method is 2 groups. The number of components of each group is 3 and 2, and the components of each group is $\{P_1, P_2, P_3\}$ and $\{P_4, P_5\}$. The optimal replacement time of $\{P_1, P_2, P_3\}$ and $\{P_4, P_5\}$ are 8.7 and 28.9, respectively and the expected total average cost rate of the system is 1001.4.

Table 5 shows the optimal replacement time for 1 group and 2 groups under various setup cost C_s. Observing Table 5, when $800 \leq C_s \leq 1100$, the optimal

number of groupings is 2 groups, and the number of components of each group is $(\{P_1, P_2, P_3\},\{P_4, P_5\})$ or $(\{P_1, P_2, P_3, P_4\},\{P_5\})$. When $C_s = 825.5$, the 2 groups $(\{P_1, P_2, P_3\},\{P_4, P_5\})$ or $(\{P_1, P_2, P_3, P_4\},\{P_5\})$ is adopted, which has the same excepted average cost rate as 1005.1. The $(\{P_1, P_2, P_3\},\{P_4, P_5\})$ and $(\{P_1, P_2, P_3, P_4\},\{P_5\})$ are performed replacement at time epochs (8.8, 29.0) and (9.8, 56.4). In Table 5, the optimal replacement time for 1 group and 2 groups is showed under various setup cost C_s.

Observing Table 5, when $800 \leq C_s \leq 1100$, the optimal number of groupings is 2 groups, and the number of components of each group is $(\{P_1, P_2, P_3\},\{P_4, P_5\})$ or $(\{P_1, P_2, P_3, P_4\},\{P_5\})$. When $C_s = 825.5$, the 2 groups $(\{P_1, P_2, P_3\},\{P_4, P_5\})$ or $(\{P_1, P_2, P_3, P_4\},\{P_5\})$ is adopted, which has the same expected average cost rate as 1005.1. The optimal group replacement time of $(\{P_1, P_2, P_3\},\{P_4, P_5\})$ and $(\{P_1, P_2, P_3, P_4\},\{P_5\})$ is performed at time epochs (8.8, 29.0) and (9.8, 56.4). When $3500 \leq C_s \leq 3800$, the optimal number of groupings is 2 groups or 1 group. When $C_s = 3662$, the expected average cost rates of 2 groups $(\{P_1, P_2, P_3, P_4\},\{P_5\})$ and 1 group $(\{P_1, P_2, P_3, P_4, P_5\})$ are the same value as 1298.4. The optimal group replacement time of 2 and 1 groups are performed at time epochs (12.5, 84.4) and 13.6, respectively.

Table 6 shows the optimal replacement time of 2 and 3 groups under various setup cost C_s. Observing Table 6, when $500 \leq C_s \leq 600$, the optimal number of grouping is 2 groups $(\{P_1, P_2, P_3\},\{P_4, P_5\})$ or 3 groups $(\{P_1, P_2, P_3\},\{P_4\},\{P_5\})$. When $C_s = 549.5$, the optimal replacement time of 2 or 3 groups is performed at time epochs (8.4, 27.8) and (8.4, 22.5, 52.9), respectively, which has the same expected average cost rate as 963.5. When $300 \leq C_s \leq 400$, the optimal number of groupings is 3 groups and the number of components of each group is $(\{P_1, P_2, P_3\}, \{P_4\},\{P_5\})$ or $(\{P_1, P_2\},\{P_3, P_4\},\{P_5\})$. When $C_s = 355.5$, the optimal group replacement time of $(\{P_1, P_2, P_3\},\{P_4\},\{P_5\})$ and $(\{P_1, P_2\},\{P_3, P_4\},\{P_5\})$ are performed at time epochs (8.1, 21.2, 50.3) and (7.7, 17.2, 50.3), respectively, which has the same expected average cost rate of 927.5.

Tables 7 and 8 show the optimal replacement time for $k = 3$, 4 and $k = 4$, 5 groups. In Table 7, when $100 \leq C_s \leq 125$, the optimal number of groupings is 3 or 4 groups. When $C_s = 106$, the groups $(\{P_1, P_2\},\{P_3, P_4\},\{P_5\})$ and $(\{P_1, P_2\},\{P_3\}, \{P_4\},\{P_5\})$ are adopted, which has the same expected average cost rate as 874.1. The 3 and 4 groups are performed replacement at time epochs (7.2, 16.1, 46.79) and (10.4, 6.5, 19.4, 46.7), respectively. Observing Table 8, when $20 \leq C_s \leq 30$, the optimal number of groupings is 4 or 5 groups. When $C_s = 28.23$, the $(\{P_2, P_3\}, \{P_1\},\{P_4\},\{P_5\})$ or $(\{P_1\},\{P_2\},\{P_3\},\{P_4\},\{P_5\})$ is adopted, which has the same expected average cost rate as 874.18. The 4 and 5 groups are performed replacement at time epochs (10.1, 6.3, 18.8, 45.6) and (6.3, 9.3, 12.5, 18.8, 45.62), respectively.

Figure 2 shows that different optimal replacement policies should be adopted when the setup cost increases and the percentage of expected total cost rate reduction between individual replacement and grouping replacement is compared.

Table 6 The optimal replacement time of $k = 2, 3$ groups

	$k = 2\ g_1 = \{P_i\}_{i=1}^{3}$ $g_2 = \{P_4, P_5\}$		$k = 2\ g_1 = \{P_i\}_{i=1}^{4}$ $g_2 = \{P_5\}$		$k = 3\ g_1 = \{P_i\}_{i=1}^{3}$ $g_2 = \{P_4\}\ g_3 = \{P_5\}$		$k = 3\ g_1 = \{P_1, P_2\}$ $g_2 = \{P_3, P_4\}$ $g_3 = \{P_5\}$	
	T_1		T_1		T_1		T_1	
	T_2		T_2		T_2		T_2	
C_s	T_3	$E(TC)$	T_3	$E(TC)$	T_3	E(TC)	T_3	$E(TC)$
800	8.7	1001.4	9.8	1002.13	8.7	1007.9	8.4	1015.6
							18.9	
					24.1		56.1	
	28.9		56.1		56.1			
700	8.6	986.4	9.7	990.10	8.6	990.4	8.2	996.5
							18.5	
					23.5		54.9	
	28.5		54.9		54.9			
600	8.5	971.2	9.6	977.92	8.5	972.6	8.0	977.0
							18.2	
					22.8		53.6	
	28.0		53.6		53.6			
500	8.3	955.8	9.5	965.57	8.3	954.4	7.9	957.0
							17.8	
					22.2		52.3	
	27.6		52.3		52.3			
400	8.2	940.1	9.3	953.04	8.2	935.9	7.7	936.7
							17.4	
					21.5		50.9	
	27.2		50.9		50.9			
300	8.1	924.2	9.2	940.34	8.1	916.9	7.6	915.9
							16.9	
					20.8		49.5	
	26.7		49.5		49.5			
200	7.9	907.9	9.1	927.44	7.9	897.6	7.4	894.6
							16.5	
					20.1		48.1	
	26.3		48.1		48.1			
100	7.8	891.4	9.0	914.35	7.8	877.7	7.2	872.8
							16.1	
					19.4		46.7	
	25.8		46.7		46.7			

Table 7 The optimal replacement time of $k = 3, 4$ groups

	$k = 3\ g_1 = \{P_1, P_2, P_3\}\ g_2 = \{P_4\}\ g_3 = \{P_5\}$		$k = 3\ g_1 = \{P_1, P_2\}\ g_2 = \{P_3, P_4\}\ g_3 = \{P_5\}$		$k = 4\ g_1 = \{P_2,P_3\}\ g_2 = \{P_1\}\ g_3 = \{P_4\}\ g_4 = \{P_5\}$		$k = 4\ g_1 = \{P_1,P_2\}\ g_2 = \{P_3\}\ g_3 = \{P_4\}\ g_4 = \{P_5\}$	
C_s	T_1 T_2 T_3	$E(TC)$	T_1 T_2 T_3	$E(TC)$	T_1 T_2 T_3 T_4	$E(TC)$	T_1 T_2 T_3 T_4	$E(TC)$
300	8.1	916.9	7.6	915.9	11.0	934.5	7.6	929.9
	20.8		16.9		7.0		15.2	
	49.5		49.5		20.8		20.8	
					49.5		49.5	
275	8.0	912.1	7.5	910.6	11.0	926.9	7.5	923.3
			16.8		6.9		15.0	
	20.6		49.2		20.6		20.6	
	49.2				49.2		49.2	
250	8.0	907.3	7.5	905.3	10.9	919.3	7.5	916.6
	20.5		16.7		6.8		14.8	
	48.8		48.8		20.5		20.5	
					48.8		48.8	
225	7.9	902.4	7.4	900.0	10.8	911.6	7.4	909.8
	20.3		16.6		6.8		14.5	
	48.5		48.5		20.3		20.3	
					48.5		48.5	
200	7.9	897.6	7.4	894.6	10.7	903.9	7.4	903.0
	20.1		16.5		6.7		14.3	
	48.1		48.1		20.1		20.1	
					48.1		48.1	
175	7.9	892.6	7.4	889.2	10.6	896.1	7.4	896.1
	19.9		16.4		6.7		14.0	
	47.8		47.8		19.9		19.9	
					47.8		47.8	
150	7.8	887.7	7.3	883.8	10.5	888.2	7.3	889.1
	19.7		16.3		6.6		13.8	
	47.4		47.4		19.7		19.7	
					47.4		47.4	
125	7.8	882.7	7.3	878.3	10.5	880.2	7.3	882.1
	19.5		16.2		6.5		13.5	
	47.0		47.0		19.5		19.5	
					47.0		47.0	
100	7.8	877.7	7.2	872.8	10.4	872.2	7.2	875.0
	19.4		16.1		6.5		13.3	
	46.7		46.7		19.4		19.4	
					46.7		46.7	

Table 8 The optimal replacement time of $k = 4, 5$ groups

	$k = 4$ $g_1 = \{P_2, P_3\}$ $g_2 = \{P_1\}$ $g_3 = \{P_4\}$ $g_4 = \{P_5\}$		$k = 4$ $g_1 = \{P_1, P_2\}$ $g_2 = \{P_3\}$ $g_3 = \{P_4\}$ $g_4 = \{P_5\}$		$k = 5$ $g_1 = \{P_1\}$ $g_2 = \{P_2\}$ $g_3 = \{P_3\}$ $g_4 = \{P_4\}$ $g_5 = \{P_5\}$	
C_s	(T_1, T_2) (T_3, T_4)	$E(TC)$	(T_1, T_2) (T_3, T_4)	$E(TC)$	(T_1, T_2) (T_3, T_4) T_5	$E(TC)$
100	(10.4, 6.5) (19.4, 46.7)	872.2	(7.2, 13.3) (19.4, 46.7)	875.0	(6.5, 9.6) (13.3, 19.4) 46.7	878.3
90	(10.3, 6.4) (19.3, 46.5)	869.0	(7.2, 13.2) (19.3, 46.5)	872.1	(6.4, 9.6) (13.2, 19.3) 46.5	874.3
80	(10.3, 6.4) (19.2, 46.4)	865.7	(7.2, 13.1) (19.2, 46.4)	869.3	(6.4, 9.5) (13.1, 19.2) 46.4	870.2
70	(10.3, 6.4) (19.1, 46.2)	862.5	(7.2, 13.0) (19.1, 46.2)	866.4	(6.4, 9.5) (13.0, 19.1) 46.2	866.1
60	(10.2, 6.4) (19.1, 46.1)	859.2	(7.2, 12.9) (19.1, 46.1)	863.5	(6.4, 9.4) (12.9, 19.1) 46.1	862.0
50	(10.2, 6.3) (19.0, 45.9)	855.9	(7.1, 12.8) (19.0, 45.9)	860.6	(6.3, 9.4) (12.8, 19.0) 45.9	857.8
40	(10.2, 6.3) (18.9, 45.8)	852.6	(7.1, 12.6) (18.9, 45.8)	857.7	(6.3, 9.3) (12.6, 18.9) 45.8	853.7
30	(10.1, 6.3) (18.8, 45.6)	849.3	(7.1, 12.5) (18.8, 45.6)	854.7	(6.3, 9.3) (12.5, 18.8) 45.6	849.5
20	(10.1, 6.3) (18.7, 45.4)	846.0	(7.1, 12.4) (18.7, 45.4)	851.8	(6.3, 9.2) (12.4, 18.7) 45.4	845.3
10	(10.0, 6.2) (18.7, 45.3)	842.7	(7.1, 12.3) (18.7, 45.3)	848.8	(6.2, 9.2) (12.3, 18.7) 45.3	841.0
0	(10.0, 6.2) (18.6, 45.1)	839.3	(7.1, 12.2) (18.6, 45.1)	845.8	(6.2, 9.1) (12.2, 18.6) 45.1	836.8

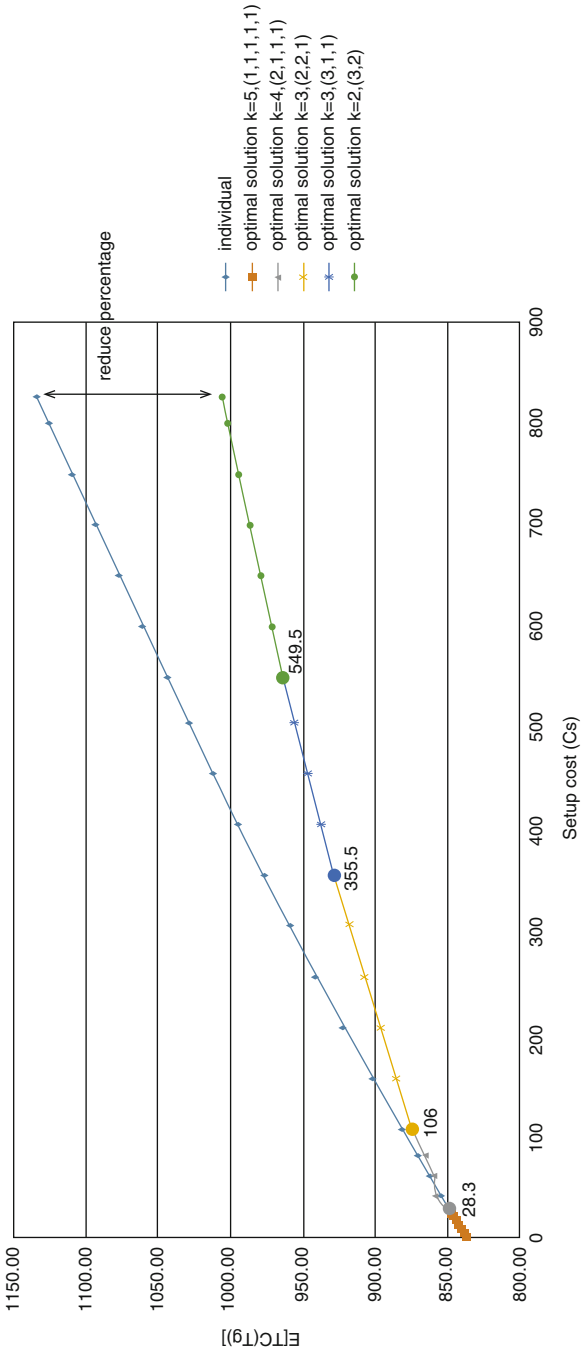


Fig. 2 The optimal grouping policy under various C_s

5 Conclusions

This paper investigates the group replacement policies for repairable n-component parallel systems and derives the grouping models when n components are divided into k ($1 \leq k \leq n$) groups. Further, the optimal group replacement time of each group is obtained. Especially, there are closed-form solutions when the lifetime distributions of components are Weibull. In addition, the influence of the setup cost on the optimal replacement policy is analyzed by numerical examples. It is observed that when the setup cost is relatively large, a lower number of groupings of components should be adopted and performed to minimize the expected total average cost rate of the system.

Acknowledgments This research is supported in part by grants (MOST-103-2221-E-011-061-MY3 and MOST 104-2221-E-469-001) from the Ministry of Science and Technology, Taiwan.

References

1. Barlow RE, Hunter LC (1960) Optimum preventive maintenance policies. Oper Res 8:90–100
2. Nakagawa T, Kowada M (1983) Analysis of a system with minimal repair and its application to replavcement policy. Eur J Oper Res 12:176–182
3. Chien YH, Chang CC, Sheu SH (2009) Optimal periodical time for preventive replacement based on a cumulative repair-cost limit and random lead time. Proc Inst Mech Eng O J Risk Reliab 223:333–345
4. Berg M, Epstein B (1978) Comparison of age, block and failure replacement policies. IEEE Trans Reliab 27:25–29
5. Beichelt F (1981) A generalized block-replacement policy. IEEE Trans Reliab 30:171–172
6. Chen M, Feldman RM (1997) Optimal replacement policies with minimal repair and age-dependent costs. Eur J Oper Res 98:75–84
7. Sheu SH, Griffith WS (2001) Optimal age-replacement policy with age-dependent minimal-repair and random-lead time. IEEE Trans Reliab 50:302–309
8. Chien YH, Sheu SH (2006) Extended optimal age-replacement policy with minimal repair of a system subject to shocks. Eur J Oper Res 174:169–181
9. Shafiee M, Finkelstein M (2015) An optimal age-based group maintenance policy for multi-unit degrading systems. Reliab Eng Syst Saf 59:374–382
10. Zhao X, Al-Khalifa NK, Nakagawa T (2015) Approximate methods for optimal replacement, maintenance, and inspection policies. Reliab Eng Syst Saf 144:68–73

Low Speed Bearing Condition Monitoring: A Case Study

Fang Duan, Ike Nze, and David Mba

Abstract The health condition of worm-wheel gearboxes is critical for the reliable and continuous operation of passenger escalators. Vibration sessors have been widely installed in such gearboxes and vibration levels are usually utilized as health indicators. However, the measurement of vibration levels is not robust in slow speed bearing condition monitoring. In this paper, the health condition of two slow speed bearings were evaluated using vibration data collected from sensors installed in the shaft of a worm wheel gearbox. It has been shown that the vibration level fails to indicate the bearing health condition. The assessment accuracy can be improved by combining several simple methods.

Keywords Low speed bearing • Worm-wheel gearbox • Vibration analysis • Health condition monitoring

1 Introduction

The London Underground (LU) is one of the world's oldest and busiest metro systems carrying in excess of one billion passengers each year. The system serves 270 stations and a total 11 train lines. In order to increase accessibility and cope with high passenger volumes, LU manages 430 escalators within its stations for expeditious commuter movement during peak periods of travel. These escalators cope with heavier loading, which is up to 13,000 passengers an hour and run more than 20 hours a day. Escalator availability is essential for the prompt transport of passengers to and from platforms. Therefore, it is tracked as a Key Performance Indicator (KPI) in LU performance reports released by the Mayor of London. Significant efforts have been made by LU to maintain the network escalator availability as high as above 95%.

F. Duan (✉) • D. Mba
School of Engineering, London South Bank University, 103 Borough Road, London SE1 0AA, UK
e-mail: duanf@lsbu.ac.uk

I. Nze
Capital Programmes Directorate, London Underground, London SW1E 5ND, UK

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_4

LU escalators are driven by electric motors. Power from the motor is transmitted to the drive gear, which drives the escalator steps along with the chain, predominantly via a worm-wheel gearbox. Worm-wheel gearbox generates high friction compared to other gear types due to the entirely sliding action (as opposed to rolling) between worm and gear. Gearbox wear occurs during normal operation. Manufacturing tolerances, inappropriate installation, poor maintenance and abnormal environmental conditions can accelerate gearbox wear [1, 2]. Hence, condition monitoring, regular inspection and maintenance are essential in order to obtain maximum usage and prevent from premature failure.

Preventive maintenance is the predominant mode of LU escalators service plan. Worm wheel replacement is carried out when wheel pitting damage is observed by maintenance crews. The assessment of the allowable pitting damage is dependent on crew experience and hence is subjective; without the establishment of consistent criteria for the amount of pitting that will render worm wheel change necessary. Premature replacements normally result in the wastage of otherwise useful worm wheel hours and additional labour costs. There has been a recent shift towards condition-based maintenance for greater effectiveness in manpower utilisation and machine-hour capitalization [3, 4]. Condition monitoring is the process of machine health assessment while it is in operation. It not only can prevent an unscheduled work stoppage and expensive repair in the event of catastrophic failures but can also optimise machine performance, provide effective plans for scheduled maintenance manpower, and suggest the advance procurement of machine spare parts that need to be replaced to bring the machine back to health.

Condition monitoring through the use of vibration analysis is an established and effective technique for detecting the loss of mechanical integrity of a wide range and classification of rotating machinery. Equipment rotating at low rotational speeds presents an increased difficulty to the diagnostician, since conventional vibration measuring equipment is not capable of measuring the fundamental frequency of operation. Also, component distress at low operational speeds does not necessarily show an obvious change in vibration signature. Further details can be found in [5–8]. Furthermore, in most of gears types, defects manifest as periodic impacts in the form of side-bands around the gear mesh frequencies [9]. However, such distinctive defect symptoms are not obvious for worm-wheel gearboxes due to continuous sliding interactions [10]. Therefore, diagnostics of worm-wheel gearbox defects with vibration analysis is challenging. As a case study, vibration data of bearings from two stations was utilized to assess vibration based condition monitoring method. The current condition monitoring system provides a vibration level threshold to trigger an alarm before failure. However, this vibration level based method is not a robust indicator of bearing heath condition. More advanced signal processing methods, such as spectral kurtosis kurtogram, envelope analysis and FM4* [11], or/and comprehensive assessments (e.g. acoustic emission [12]) are essential to improve accuracy.

2 Slow Speed Bearing Condition Monitoring

A main drive shaft of one of the LU escalators was selected to conduct condition monitoring. The shaft has two support bearings (SKF 23026 CCK/W33), spherical roller bearings, cylindrical and tapered bore, located at each end. The main drive shaft speed operates at 10 rpm. The calculated defect frequencies for the inner ring, outer ring and rolling element are 2.3, 1.87 and 1.58 Hz, respectively. The operating condition of the bearings was monitored by measuring vibration levels. Two vibration sensors (PRÜFTECHNIK VIB 6.127) were located at each end of the shaft, as shown in Fig. 1.

2.1 Vibration Data Analysis of Station A Top Shaft

The velocity overall trends of Station A top shaft left and right side are shown in Fig. 2a, b, respectively. A threshold value of 2 mm/s is noted on the figures. In Fig. 2a, the threshold was not exceeded at the shaft left side. However, on the right side the first observed data exceeding the threshold was noted in Oct. 2014. Thereafter several peaks above the threshold appeared in Jan. and Feb. 2015, as shown in Fig. 2b. After Feb. 2015, only a few peaks were observed before the bearing fault condition was identified in May 2015.

Similar patterns were shown in acceleration overall trends of Station A shaft left and right side in Fig. 3. It is known that damaged slow speed bearings will operate continuously irrespective of the level of damaged within the bearing. In some instances such damaged slow speed bearings have been known to completely grind away components within the bearing (rollers, cage) and for the shaft to eventually be supported by the bearing housing, leading to grinding of the shaft itself [13–15]. This implies that there are situations when the vibration of a damaged bearing will not increase with time due to increasing severity, on the contrary, the vibration may decrease. This is one of the reasons why other technologies have been assessed to give better indications of slow speed bearing condition [16, 17].

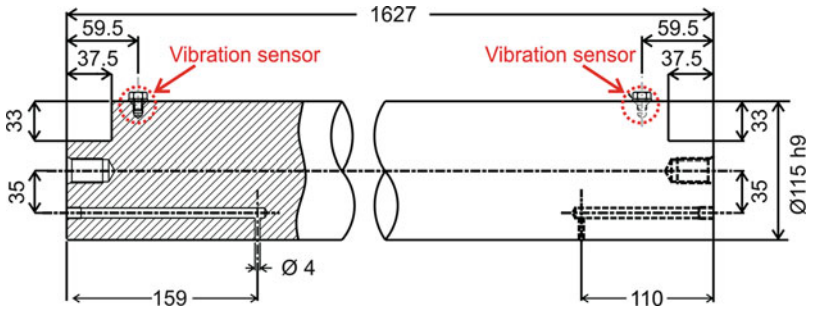


Fig. 1 Shaft and vibration sensors

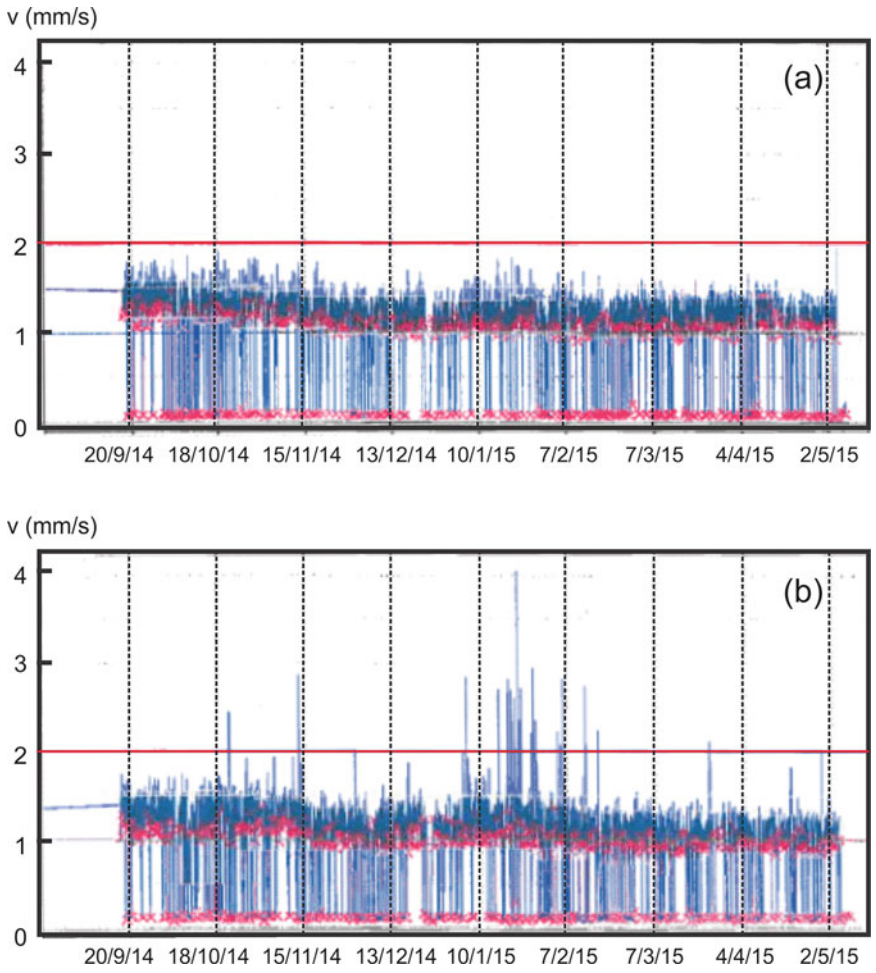


Fig. 2 Velocity overall trends of Station A top shaft. (**a**) Left side vibration sensor; (**b**) Right side vibration sensor

2.2 Vibration Data Analysis of Station B Top Shaft

Vibration waveform data, typically acquired prior to post processing, from a similar machine at Station B top shaft left and right side was provided by LU. Figure 4 shows the vibration signal of top-shaft left in Mar. 2013. A total of 307,200 points are plotted in Fig. 4a. Zoomed data plots of 100,000 to 110,000, and 100,000 to 101,000 data points are shown in Fig. 4b, c, respectively. Figure 4 highlights what can only be described as transient events superimposed on underling vibration data. However closer observation of the data reveals the transient impact type events are

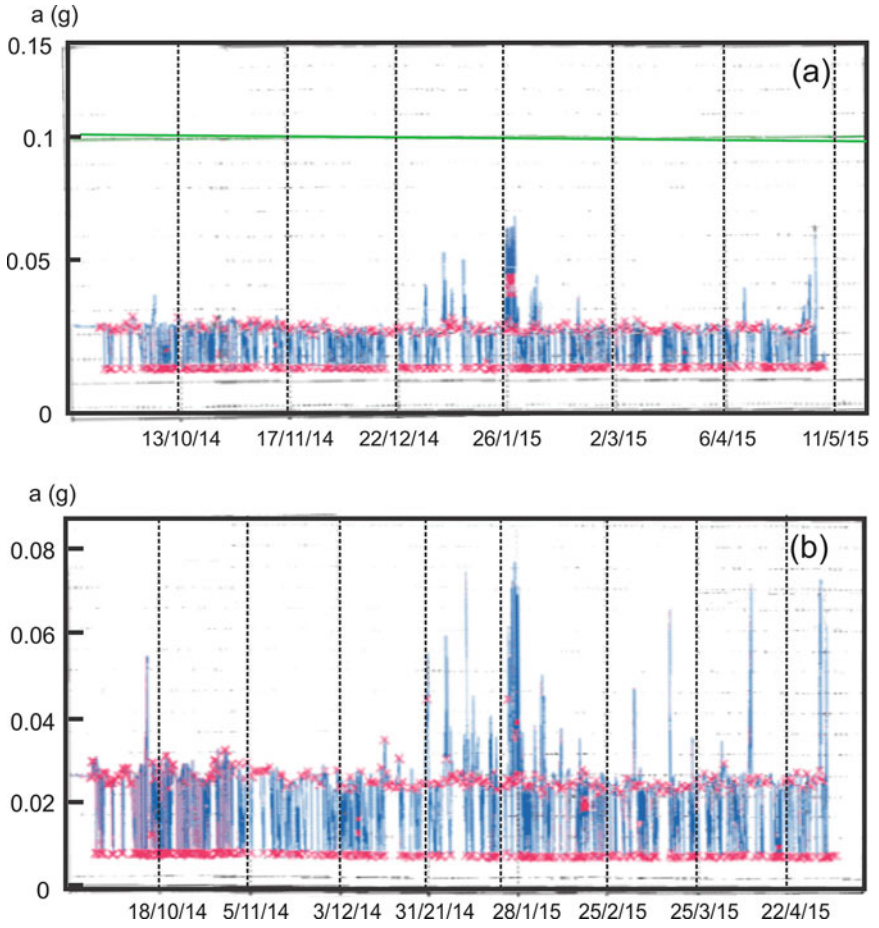


Fig. 3 Acceleration overall trends of Station A top shaft. (**a**) Left side vibration sensor; (**b**) Right side vibration sensor

not associated with electronic noise but are vibration type responses from the bearing. Such transient vibration events can be indicative of impending failure.

A time-frequency analysis and Fourier transform were undertaken on the vibration data. Figure 5 highlights that the transient vibration event contained frequencies of up to 100 Hz whilst Fig. 6 shows the spread of energy across 0–50 Hz, with the strongest energy concentration at 25 Hz. A FFT of the vibration data (data in Fig. 4a) is shown in Fig. 7. The shaft rotational speed is identified (0.17 Hz or 10 rpm) as well as several other peaks (0.43 Hz, 0.88 Hz and several multiples of 0.88 Hz). As the authors of this paper do not have information of the exact machine configuration at Station B, the sources of these vibration peaks cannot be identified. However, it does suggest that such low frequencies can be measured with the currently employed sensors.

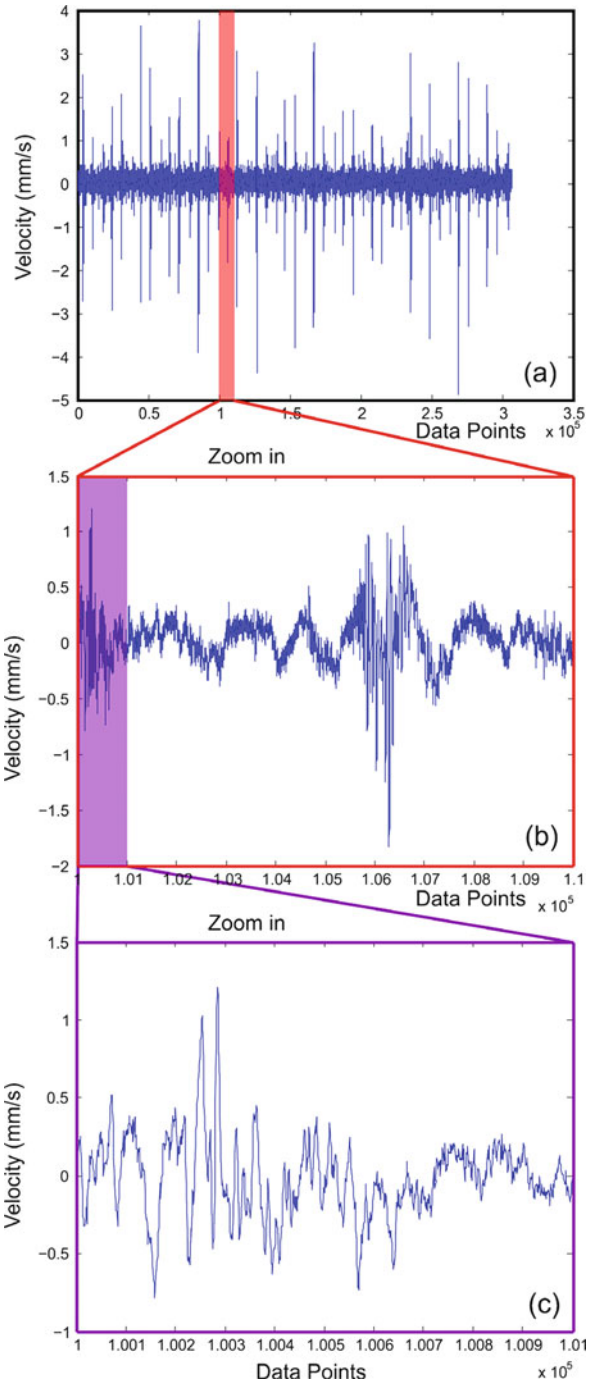


Fig. 4 Top shaft left vibration signal. (**a**) Total 307,200 points in March 2013; (**b**) Zoomed in data points from 100,000 to 110,000; (**c**) Zoomed in data points from 100,000 to 101,000

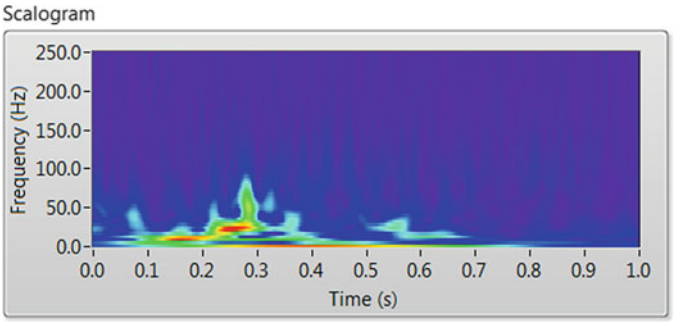


Fig. 5 Time-frequency plot of data displayed in Fig. 4c

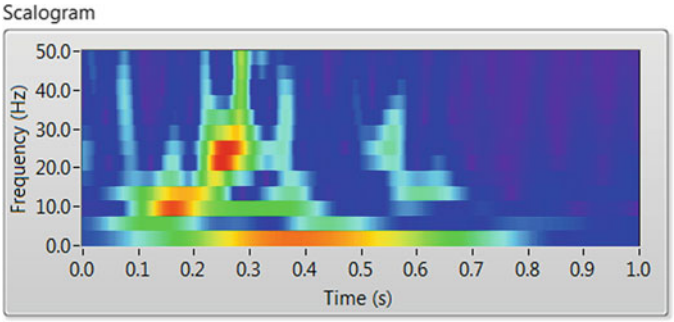


Fig. 6 Zoom of time-frequency plot of data displayed in Fig. 5

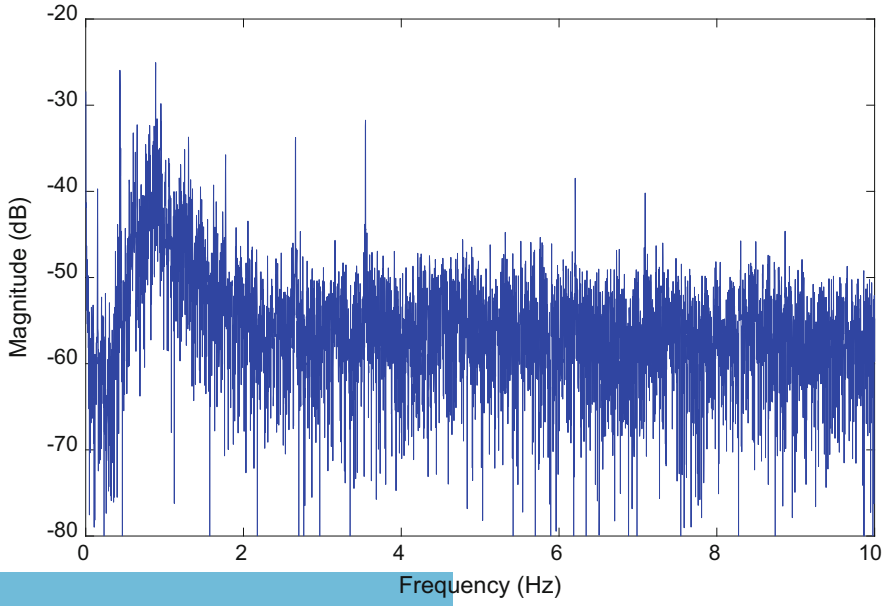


Fig. 7 Spectrum of vibration data in Fig. 4a

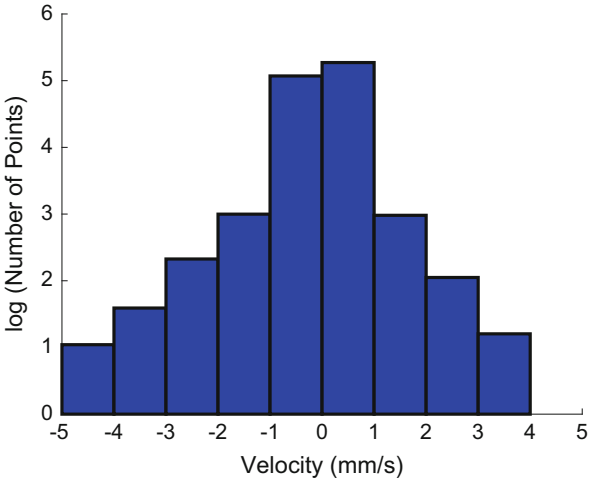


Fig. 8 Top shaft left vibration level distribution. Number of point in the logarithmic scale

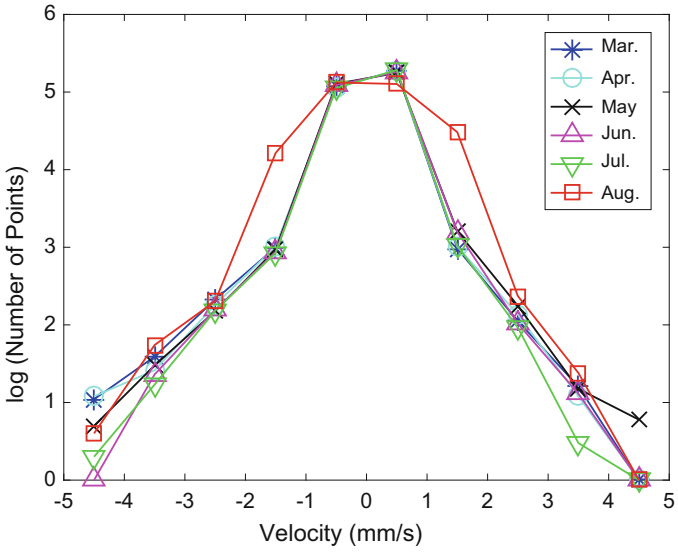


Fig. 9 Top shaft left vibration signal from March to August 2013. Vibration level distribution in the logarithmic scale

The distribution of vibration levels can also be utilized as a health indicator. The vibration level of data in Fig. 4a is in the range of −5 and 4 mm/s. The distribution of vibration levels of these data points are grouped in 10 integral areas, as shown in Fig. 8. For better visualization, the number of data points is plotted in the logarithmic scale. It can be seen that the vibration levels of most data points

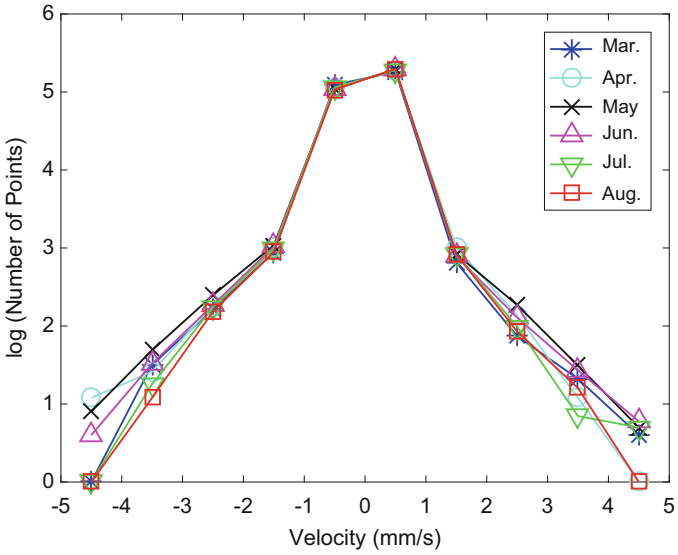


Fig. 10 Top shaft right vibration signal from March to August 2013. Vibration level distribution in the logarithmic scale

are within ± 1 mm/s. A normal distribution is observed on a logarithmic scale. The same method is utilised to process all data of both left and right vibration sensors from Mar. to Aug. 2013 from Station A. The vibration level distributions of top shaft left and right vibration signal are shown in Figs. 9 and 10, respectively. It can be observed that deviation of left side of the vibration signal in August has increased compared to the previous five months. On the contrary, there is no clear pattern in the escalator top shaft right in Fig. 10. This bearing was not damaged.

3 Conclusions

Using vibration signal as a health indicator for a worm-wheel gearbox bearing is discussed in the paper. The case study shows the single vibration method cannot provide sufficient information for a condition assessment and a combination of methods is recommended for meaningful analysis. An increasing or decreasing level of vibration can be attributed to slow speed bearing defects. This implies the interpretation of overall vibration data for slow speed bearings is not a robust indicator of condition. A better indicator may be trending vibration specific to a defined frequency or frequency bands.

References

1. Elforjani M, Mba D (2008) Monitoring the onset and propagation of natural degradation process in a slow speed rolling element bearing with acoustic Emissions. J Vib Acoust, Trans ASME 130:041013
2. Elforjani M, Charnley B, Mba D (2010) Observations of a naturally degrading slow speed shaft. Nondestruct Test Eval 25(4):267–278
3. Mba D, Jamaludin N (2002a) Monitoring extremely slow rolling element bearings: part-I. NDT & E Int 35(6):349–358
4. Mba D, Jamaludin N (2002b) Monitoring extremely slow rolling element bearings: part II. NDT & E Int 35(6):359–366
5. Berry JE (1992) Required vibration analysis techniques and instrumentation on low speed machines (particularly 30 to 300 rpm machinery). In: Advanced vibration diagnostic and reduction techniques. Technical Associates of Charlotte Inc, Charlotte
6. Canada RG, Robinson JC (1995) Vibration measurements on slow speed machinery. In: Proceedings of National Conference on Predictive Maintenance Technology (P/PM Technology), vol 8, no 6. Indianapolis, Indiana, pp 33–37
7. Kuboyama K (1997) Development of low speed bearing diagnosis technique. NKK/Fukuyama Works, Fukuyama
8. Robinson JC, Canada RG, Piety RG (1996) Vibration monitoring on slow speed machinery: new methodologies covering machinery from 0.5 to 600 rpm. In: Proceedings of the Fifth International Conference on Profitable Condition Monitoring-Fluids and Machinery Performance Monitoring, BHR Group Publication (BHR Group, Cranfield), vol 22, pp 169–182
9. Alan D (1998) Handbook of the condition monitoring techniques and methodology. Chapman and Hall, London
10. Vahaoja P, Lahdelma S, Leinonen J (2006) On the condition monitoring of worm gears. Springer, London, pp 332–343
11. Elasha F, Ruiz-Cárcel C, Mba D, Kiat G, Nze I, Yebra G (2014) Pitting detection in worm gearboxes with vibration analysis. Eng Fail Anal 42:366–376
12. Elforjani M, Mba D, Muhammad A, Sire A (2012) Condition monitoring of worm gears. Appl Acoust 73(8):859–863
13. Jamaludin N, Mba D, Bannister RH (2001) Condition monitoring of slow-speed rolling element bearings using stress waves. J Process Mech Eng 215(E4):245–271
14. Jamaludin N, Mba D, Bannister RH (2002) Monitoring the lubricant condition in a low-speed rolling element bearing using high frequency stress waves. J Process Mech Eng 216(E):73–88
15. Mba D, Bannister RH, Findlay GE (1999a) Condition monitoring of low-speed rotating machinery using Stress Waves: part I. Pro Inst Mech Eng 213(E):153–185
16. Mba D (2002) Applicability of acoustic emissions to monitoring the mechanical integrity of bolted structures in low speed rotating machinery: case study. NDT & E Int 35(5):293–300
17. Mba D, Bannister RH, Findlay GE (1999b) Condition monitoring of low-speed rotating machinery using Stress Waves: Part II. Pro Inst Mech Eng 213:153–185

Program Control-Flow Structural Integrity Checking Based Soft Error Detection Method for DSP

Yangming Guo, Hao Wu, Guochang Zhou, Shan Liu, Jiaqi Zhang, and Xiangtao Wang

Abstract Nowadays, for the equipment in the outer space, the system safety and reliability are largely affected by soft errors which are caused by the single high energy particles. This is frequently reported in DSP and the other memory devices. Thus, the detection of soft error becomes an interesting research topic. For the purpose of detecting the occurrence of soft error occurred in storage areas of a DSP program, a control flow integrity based checking scheme for soft error detection is presented in this work. In this work, the DSP program implemented in assembly language is mainly focused. Firstly, the program is divided into a number of basic blocks with corresponding structure information being stored in a partition table. Then, for each basic block, a checkpoint is set at the end. The program control flow error can be easily determined by examining the consistency between the information at runtime and that recorded information in the partition table. Compared with the signature-based method, the proposed method is able to achieve almost 100% of error detection coverage. Furthermore, the proposed detection scheme has better cross platform portability under almost same detection efficiency and detection overhead.

Keywords DSP • Soft error • Control-flow error • Integrality checking

1 Reasearch Background of Soft Error Detection Method for DSP

At present, high-performance digital signal processor (DSP) is widely used in the spacecraft electronic systems. However, in outer space, there exist a large number of high-energy charged particles which result in the occurrence of constant single event

Y. Guo (✉) • H. Wu • S. Liu • J. Zhang • X. Wang
School of Computer Science and Technology, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, PR China
e-mail: yangming_g@nwpu.edu.cn; woohow@mail.nwpu.edu.cn; 840241140@qq.com; 424515842@qq.com; 472391837@qq.com

G. Zhou
Academy of Space Technology (Xi'an), Xi'an, Shaanxi 710100, PR China
e-mail: zhouguochang2000@163.com

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_5

effect (SEE) in DSP. Consequently the transient or intermittent errors incurred by SEE are referred to as soft errors. Due to the existence of such errors, the reliability and security of the system is severely affected [1]. Hence, investigation of the effects of soft error has drawn the attention of numerous researchers and engineers. Public information indicates that more than 70% of soft errors occurring in DSP are likely to incur control flow errors to the software [1]. Control flow errors may cause the corresponding program to deviate from the correct execution flow; this is likely to result in incorrect outputs or even complete breakdown of the system [2]. Therefore, control-flow checking on the program implemented in the DSP is able to effectively detect whether soft errors occurred or not, especially in the memory blocks of the program.

The main idea of the existing control-flow error detection approach is signature based monitoring. To be specific, the program is divided accordingly into several basic blocks with assigned static signatures. These signatures are then compared with run-time signatures, which are calculated dynamically with the execution of the program. The signature based monitoring approach can be implemented by either hardware or software. The hardware implementation is implemented through the utilization of a watchdog, aiming to detect the errors that are occurred during the executing process of the program in real time. The main features of the approach using a watchdog are high speed and small overhead required. However, this inevitably introduces extra hardware, which increases the necessary overhead; this definitely increases the difficulty of the system development.

As the development of human society, there is an increasing demand of developing embedded system with high performance, low cost and short development cycle. Due to the reason that the developing process can be greatly accelerated with the application of Commercial-Off-The-Shelf (COTS), this technique has been increasingly and widely applied in the development of embedded system. However, due to the unchangeable property of manufactured COTS architecture, researchers have to figure out a new technique for the purpose of detecting control-flow errors on the top of watchdog. Aiming at the limitation of watchdog, a variety of software-based detection schemes have been proposed for control-flow errors, such as Enhanced Control-flow Checking using Assertions (ECCA) [3], Control-Flow Checking by Software Signatures (CFCSS) [4], Enhanced Control-Flow Checking (ECFC) [5], Edge Control-Flow Checking (ECFC) and Region Based Control-Flow Checking (RBCFC) [6], and etc. Software-based detection schemes are independent on the underlying hardware at the expense of performance to some extent, which greatly enhances the reliability of COTS-based system. Although the signature based monitoring approach implemented by software also depends on no hardware, it still is highly platform-dependent. Due to different instruction set for different platforms such as ARM, MIPS and x86, the techniques of signature generation and monitoring for one platform cannot be directly applied to another. Thus the portability of the software system is greatly limited.

To overcome the shortcomings mentioned above, this paper presents a control-flow error detection method based on structural integrity checking (SIC). Whether the control-flow errors occur in the program is determined by checking the execution integrity of each basic block. The most desirable advantage of the proposed

scheme is portability, that is, only minor modifications are necessary for adapting the method for one platform to another. In terms of overhead required for detection, the proposed technique and signature-based control-flow detection scheme are at the same order of magnitude. Furthermore, the proposed method based on SIC is able to reach nearly 100% of error detection coverage.

2 Principle of SIC

2.1 Analysis of the Detection Principle

Aiming to monitor the state of the DSP in operation, the failure modes and characteristics of the soft errors incurred by SEE should be analysed first. Then corresponding program for the DSP is reasonably divided into several modules, between which the checkpoints can be set.

For the program of a DSP inside a basic block, the instructions are executed se-quentially. If no control-flow error occurs, the program starts to execute at the entry port of the block and then jump out at the exit port. Therefore, the instruction branch error is the main reason for the occurrence of the control-flow error. After the basic block partition process is completed, all the branch instructions are located at the end of the basic blocks and all non-branching instructions are inside the basic blocks. Thus the incorrect execution of a branch instruction is likely to cause the control flow being transferred to an incorrect address; thus an anticipated control-flow transition that should have happened does not occur or vice versa.

In addition, soft errors occurring in a DSP may cause data bits in memory cells to be flipped. If the flip happens inside the storage area for the program, the instruction code may be changed. It is indicated by the public document [7] and practical analysis that, the unexpected change in program storage area may cause a non-branching instruction to be transformed into a branch instruction, or vice versa. Hence, if a non-branching instruction is converted to a branch instruction, the control-flow error will definitely occur in the program. Due to the fact that, all non-branching instructions are inside the basic blocks (including the first instruction of the basic block). Then the control-flow error incurred by the unexpected conversion is reflected as the incorrect control-flow shift from the basic block to other blocks.

According to previous analysis, the control-flow errors are being classified into two categories. The first one is that, the control flow is not transferred where there should be an anticipated transfer, usually incurred by the incorrect transfer address. The other one is referred to the scenario that, the control flow is transferred unexpectedly where there should not be any transfer.

In Fig. 1, five basic blocks are presented for an illustration; here, the rectangles represent corresponding basic blocks being divided. Here, the black and red arrows stand for correct and incorrect control-flow transfers respectively. A and B indicate the type of control-flow errors incurred by the conversion from a non-branching

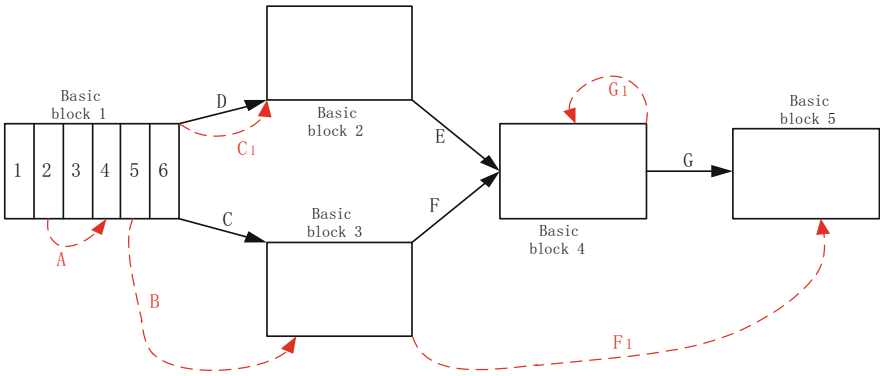


Fig. 1 An illustration of control-flow errors

instruction to a branch instruction; here, A indicates the incorrect control-flow transfer inside the same basic block while B denotes the incorrect transfer between two basic blocks. Furthermore, C and D indicate two branches that are anticipated to occur; while C1 denotes an incorrect control-flow transfer because of the branch instruction error; nevertheless, this seems to be correct due to the anticipated branch of D. F1 represents an incorrect transfer from one block to another, due to the branch instruction error. G1 refers to the incorrect transfer inside the basic block itself.

2.2 Handling of Delay Slot

The delay slot is usually referred to as one or several instructions after the branch instruction. The instructions in the delay slot are always executed and the result is submitted before the execution of the branch instruction; this behavior is regardless of the occurrence of the branch.

The delay slot is used to improve the efficiency of the pipeline on the early version of platforms which are incapable of predicting the branching. Since modern processors are capable of predicting the branching, the delay slot is useless. Nevertheless, aiming to achieve the compatibility of software system, the delay slot is also kept on the platform like MIPS, PowerPC and DSP. Thus, for the purpose of reducing the required storage overhead, the proposed signature based detection method is stored in the branch delay slot [1].

The instructions in the delay slot are supposed to be an independent basic block according to the definition. Nevertheless, in the function level, the instructions have a strong correlation with those in former basic block. Hence, the instructions should be merged into the former basic block. Furthermore, some delay slots are composed of completely empty operation instructions. Thus, there is nonsense of classifying the delay slot as a basic block; this is done also for the purpose of reducing the required detection overhead. To sum up, the delay slot is supposed to belong to the former basic block in this work.

2.3 Basic Block Partition Process

The control-flow error detection is largely affected by the basic block partition. For each basic block, it is a section of code that is executed sequentially. Thus the control flow can only enter the basic block at the first instruction and leave after the last instruction is executed. Usually, there are no more control-flow branches in the basic block except the last instruction; and there is no other control-flow entry into a basic block except the first instruction.

The basic block partition process can be implemented on two levels: C language and assembly one [8, 9]. Due to the high-level property of C language, it is independent of platforms. Therefore, the C program is capable of being executed on different platforms, after being divided into basic blocks and hardened by control-flow detection technique. However, the high-level language code has to go through the compilation and linking process to become executable; otherwise, the program is unable to be executed run on the machine. In the process of converting to low-level language, the basic block implemented by C language is not well mapped to lower level languages.

At present, most basic block partition techniques are implemented based on the assembly language. Mnemonics are utilized by the assembly instructions to translate the machine instructions. These two kinds of instructions (assembly and machine ones) correspond to each other well; thus the basic blocks divided on the assembly level are well mapped to the executable ones consisting of machine instructions. Hence, the purpose and appeal (adding signature etc.) of the partition on the language level can be satisfied during the execution of the program. In addition, the control-flow error detection technique based on SIC is required to get the structure information of each basic block, including the number of instructions and the address of the last instruction. Thus, this paper chooses to implement the basic block partition on the assembly level.

For different instruction sets, they all include the instructions like function calls, function returns and jumps, regardless of platforms. The basic block partition process is completed by seeking corresponding positions of the transfer instructions, such as function calls, function returns and jumps; these instructions are utilized to locate the boundary of the basic block. Here, the system function calls and interrupts are not considered as the transfer instructions to dispose. Therefore, the two types of instruction are both regarded as the ordinary instructions in the partition process [9].

2.4 Block Table Design Process

To realize the SIC based control-flow detection scheme, various information related with each basic block is required, including the entry port (i.e., the address of the first instruction), the exit port (i.e., the address of the last instruction), the length of

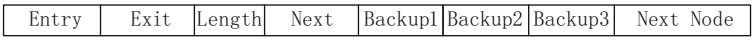


Fig. 2 An illustration of the data structure of the block table

each block (i.e., the number of instructions included by the block), next address (i.e., the target address for the jump operation or the address where the first instruction of the sub function locates if there is a function call operation). In addition, it is also necessary to back up the instructions and store three copies of the last instruction in the block, for the purpose of setting checkpoints. Hence, we need to design the appropriate data structure.

Nevertheless, the number of basic block divided by the program is unable to be determined as it is dynamic and the process happens at run time. Therefore, the designed data structure needs to be growing dynamically. In this paper, the block table is designed as shown in Fig. 2.

Based on the above analysis, the partition rules are listed as follows:

1. Taking the positions of the jump instruction, the function call instruction and the function return instruction as the exit to the current basic module;
2. Regarding the location of the following instruction (the first instruction after the delay slot if it exists) after the current jump instruction (call or call return) as the entrance to the next basic block;
3. Taking the position where the destination instruction of the current jump locates as the entrance to another basic block.

In fact, the basic block partition process is equivalent to constantly seeking the position of transfer instructions by scanning the assembly program. These positions are utilized to locate the boundary of the basic block. Furthermore, for the convenience of the detection, a working pointer is also set for the block table in this work. The pointer always points to the next node where the information about the block that is able to be checked completely by SIC.

Integrity checking refers to the final inspection of each basic block did the current basic block are complete

3 SIC Based Soft Error Detection Method via Program Control-Flow

3.1 Detection Mechanism Based on SIC

The SIC refers to check whether the basic block is completely executed at the end. If the execution is completed, then control-flow errors did not happen in the execution process of the program in the previous block. The SIC based detection scheme determines the occurrence of control-flow errors by checking by checking whether each basic block is completely executed or not.

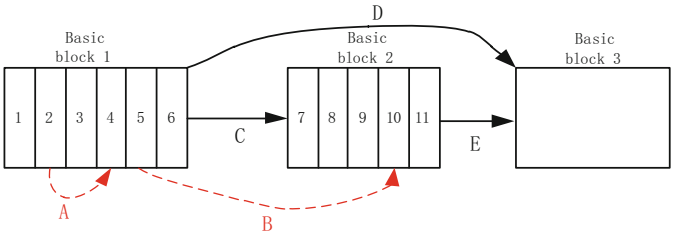


Fig. 3 An example of SIC

In order to implement SIC on each basic block, the checkpoints should be set at the exit of the block. In Fig. 3, when the program is executed in basic block 1, the control flow will jump from the fifth instruction to the tenth one given the occurrence of the control-flow error B. Due to the control-flow error B, the checkpoint being set at the sixth instruction is missed but the working pointer in the block table still points to the node. Hence, the structure information of basic block 1 is obtained. When the program is executed to the eleventh instruction in basic block 2, then the SIC based error detection process starts after the program is running into the checkpoint being set at basic block 2. Here, the detection module compares the address of '11' for the current checkpoint with the exit information '6' of the basic block indicated by the working pointer, and obviously this information is in different. As a result, the control-flow error occurs between two basic blocks and this is figured out by the detection module through comparison. Most control-flow errors can be detected successfully by that address comparison.

The control-flow error occurring inside a basic block cannot be detected by comparing addresses. As illustrated in Fig. 3, A indicates a control-flow error inside a basic block. Then after executing the second instruction in basic block 1, the program transfers to the fourth instruction. After executing the fifth instruction, the control-flow error B occurs. Thus, the checkpoint being set at the sixth instruction is unable to detect this error. Hence, a counter with an initial value of 0 is utilized here when the control flow enters a basic block. The corresponding value of counter is added by 1 if an instruction is executed.

After the control flow enters the checking module, the address of the checkpoint is compared with the block exit information in the node pointed by the working pointer first, and then the value of counter is compared with the length of corresponding block. If the two kinds of information are consistent, then the basic block is considered to be executed completely without any control-flow errors. Nevertheless, if the address information is not the same as the exit information, then there is a control-flow error between two basic blocks; if the value of counter is not the same as the block length, this indicates a control-flow error is occurred inside the basic block.

Hence, the SIC based detection technique is able to detect all control-flow errors as presented in Fig. 1 except that of C1. Thus, additional fault-tolerant mechanism needs to be applied to avoid the influence of the error such as C1.

3.2 Checkpoints Positioning

For software-based detection schemes, the source code is necessary to be modified to insert the checkpoints. The signature-based scheme is implemented by directly inserting the codes for comparing and updating signatures operations between two basic blocks. Thus, this will definitely increase the memory overhead required by the system.

One of the key steps to realize SIC based scheme is comparing the address of the checkpoint with the block exit information pointed by the working pointer of the block table. The block partition process is finished before the checkpoint insertion. Then, if the checkpoints are inserted by directly inserting the checking codes into the source code, the instruction address in the program is definitely affected by the insertion. Thus, the address change makes it invalid to detect errors by comparing addresses.

Based on the above discussion, the checking code for the scheme proposed in this work formulates into to an independent module, namely the checking module. This module is set at the end of the source code. At the same time, the last instruction of the basic block is replaced by a jump instruction to the checking module. As illustrated in Fig. 4, the solid arrows represent the control flows that are unprotected; while the dotted ones stand for the control flows being protected. In basic block 1, the checkpoint is being set at the original sixth instruction; then it is replaced by a jump instruction to the checking module. After the execution of the checking module, the instruction of the checking module is executed to realize the 'true' transfer of the anticipated control flow. As to the other basic blocks, the setting of the checkpoints and the transfer of the control flow are the same as those for basic block 1.

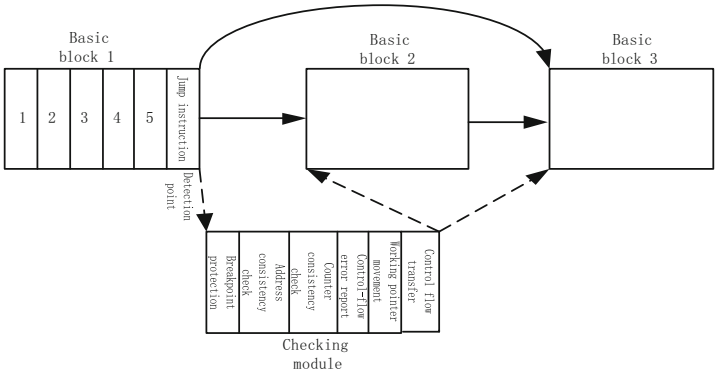


Fig. 4 Checkpoint setting and checking module

3.3 Design of Checking Module

In order to realize SIC based detection, the normal execution of program should be 'interrupted' by transferring the control flow to the checking module. To ensure the control flow returning to the original program correctly, we need to save the breakpoint information after the flow enters the checking module. Then, the consistency check of address and counter are to be completed by the checking module. Based on the analysis of the result for the comparison, we can easily figure out whether the current basic block is executed completely or not, i.e., whether there is a control-flow error occurred in the block or not. If the error occurs, the execution of program is paused and the error is reported to the system; otherwise, the next SIC starts to be initialized, i.e., moves the working pointer in the block table and assigns the value of counter to be 0.

Then, this indicates that the SIC based checking of the previous block is completed successfully, and the working pointer must be pointed to the next node for performing next operation. However, there is no real transfer of the control flow in the checking module and which block to execute next is unknown. To solve this problem, we introduce another process, named 'pre-transfer'.

The 'pre-transfer' process is to restore the program breakpoint previously saved, and then execute three instructions stored in block table. After that, the results are processed by a voter, namely triple modular redundancy (TMR). According to the result of voting, the block to execute next is determined, and then the working pointer repoints to the node where information of next block is stored.

After all the steps are executed, the control flow is still inside the checking module. Therefore, it is necessary to transfer the control flow back to the program at the end of the checking module. This can be done just by restoring the breakpoint saved and performing TMR on the replaced instructions.

4 Performance Analysis

4.1 Detection Coverage Analysis

The proposed detection scheme classifies the control-flow errors as between-block and inside-block, which are caused by branch and non-branching instructions respectively. Hence, this completely covers all the types of control-flow errors. With the help of TMR, the SIC based detection technique can detect all errors and the fault detection coverage is higher than most signature-based detection methods.

4.2 Analysis of Fault Detection Efficiency

The fault detection efficiency is referred to the number of instructions executed between the occurrence and the successful detection of the control-flow error. The efficiency is only affected by the length of the basic block and the positioning of the checkpoints. For the SIC based detection scheme, a checkpoint is set up at the end of each basic block. Therefore, all control-flow errors can be detected in the length of a basic block. However, due to the higher efficiency of semantic expression, the basic blocks divided on the higher language level are averagely shorter than those that are implemented on the assembly level. In conclusion, although the method proposed by this paper is no more efficient than other detection schemes on the assembly level, it still performs better than the method implemented on the higher language level.

4.3 Overhead Analysis

The overhead is defined as the ratio of the execution time of detection mechanism and that of the original program. In order to reveal the relationship of the overhead and number of circles, analyses of several scenarios are performed. For the scenarios presented here, different numbers of checkpoints are set. In Table 1, T30 indicates the original executable file; T31 contains one checkpoint and the checking module is executed 256 times; T32 has two checkpoints and the checking module is executed 512 times; T33 includes 15 checkpoints and the most of circles have been optimized; T34 contains 27 checkpoints and there is no optimization. Here, it costs 34 clocks if the checking module is executed for one time. Corresponding simulation results are presented in Table 1. The overhead is mainly influenced by the number and time of circles, as revealed by the results in Table 1.

For the SIC based detection scheme, four different processes are introduced listed as breakpoint protection, detection implementation, initialization of next detection (moving working and reset the value of counter) and return to the original program. Similarly for the signature-based detection methods, there are also four processes, listed as breakpoint protection, signature checking, signature update and return to the original program. Therefore, the overhead of SIC based detection scheme is roughly the same as that of signature-based detection methods.

Table 1 The overheads required by different test set

	T30	T31	T32	T33	T34
Clock	263298	271256	280128	302306	375681
Overhead	–	3.1%	6.4%	14.8%	42.7%

5 Conclusions

This paper proposes an SIC based soft error detection method for the DSP. Firstly, the program for the DSP in assembly language is divided into a number of basic blocks with structure information stored in a partition table. Then, a checkpoint is set at the end of each basic block. By examining the consistency of the address information and the counter value, the program control flow errors can be easily detected. Compared with signature-based method, the proposed method can reach almost 100% of error detection coverage. As it is implemented by assembly language, it has better cross platform portability while detection efficiency and detection overhead can also be ensured.

Acknowledgments This work is supported by National Natural Science Foundation of China under Grant No. 61371024 and No. 61601371, Aviation Science Fund of China under Grant No. 2016ZD53035, the Industry-Academy-Research Project of AVIC No.cxy2013XGD14, and the Open Research Project of Electronic components reliability physics and application technology Key Laboratory.

References

1. Nodoushan MJ, Miremadi SG, Ejlali A (2008) Control-flow checking using branch instructions. In: Proceeding of the IEEE/IFIP international conference on embedded and ubiquitous computing (EUC 2008), Shanghai, China
2. Xing KF (2007) Single event effect detection and mitigation techniques for spaceborne signal processing platform. National University of Defence Technology, Changsha
3. Alkhalifa Z, Nair VSS, Krishnamurthy N et al (1999) Design and evaluation of system-level checks for on-line control-flow error detection. IEEE Trans Parallel Distrib Syst 10:627–641. doi:10.1109/71.774911
4. Oh N, Shirvani PP, McCluskey EJ (2002) Control-flow checking by software signatures. IEEE Trans Reliab 51:111–122. doi:10.1109/24.994926
5. Reis GA, Chang J, Vachharajani N et al (2005) SWIFT: Software implemented fault tolerance. In: Proceedings of the third international symposium on code generation and optimization (CGO), San Jose, CA
6. Borin E, Wang C, Wu YF et al (2006) Software-based transparent and comprehensive control-flow error detection. In: Proceedings of the international symposium on code generation and optimization (CGO), New York, NY
7. Benso A, Di Carlo S, Di Natale G, Prinetto P (2002) Static analysis of SEU effects on software applications. In: Proceedings of the international test conference (ITC), Baltimore, MD
8. Goloubeva O, Rebaudengo M, Sonza Reorda M, Violante M (2003) Soft-error detection using control flow assertions. In: Proceedings of the 18th IEEE international symposium on defect and fault tolerance in VLSI systems (DFT'03), Boston, MA
9. Huang ZY (2006) Research and implementation of software error detection technique for on-board computers. Harbin Institute of Technology, Harbin

Research on the Fault Diagnosis of Planetary Gearbox

Tian Han, Zhen Bo Wei, and Chen Li

Abstract As a main transmission composition, planetary gearbox is widely used in wind turbine generation system. Due to the complicated working environment, sun gears, planet gear, ring gear and other key components are prone to failure. Therefore, the researches on the fault features of the planetary gearbox have significance to understand the operation of wind turbines, timely discover fault position and predict the trend of running status. In this paper, a method is proposed for the fault diagnosis of the planetary gearbox combined maximum correlation kurtosis deconvolution and frequency slice wavelet transform. The maximum correlation kurtosis deconvolution with particle swarm optimization algorithm is used to improve the signal to noise ratio. The frequency slice wavelet transform transferred the denoised signal into time-frequency domain to identify the gear fault. The feasibility of the proposed method is verified by testing the experimental signal through the test rig.

Keywords Fault diagnosis • Planetary gearbox • Vibration signal • Maximum correlation Kurtosis deconvolution

1 Introduction

Planetary gearbox is a main component of the wind turbine that transmits heavy loads from the driving motor to the driven machine with compact structure, large transmission ratio. Based on the fault statistics of the wind turbine, the outages caused by planetary gearbox accounts for 5.0%. However its downtime accounts for a very large share, about 16.3% due to the bad working conditions. Thus the

T. Han (✉) • Z.B. Wei
University of Science and Technology Beijing, 30 Xueyuan Road, Haidian District, Beijing, China
e-mail: hantian@ustb.edu.cn; weizhenbo202@126.com

C. Li
Key Laboratory of Operation Safety Technology on Transport Vehicles, Ministry of Transport, No.8 Xitucheng Road, Haidian District, Beijing, China
e-mail: c.li@rioh.cn

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_6

researches on the fault recognition of planetary gearbox have been obtained more and more attention to timely discover accurate fault position, predict the trend of running status [1].

According to the planetary gearbox vibration signal characteristics, the domestic and foreign scholars have applied various methods to carry on the exploratory research. During the 1990s, foreign scholars began to study planetary gears fault diagnosis. For example, the time average method was generalized by McFadden [2, 3] and Samuel and Pines [4] etc., the calculation method of the vibration components of the planet wheel and/or the solar wheel by using single sensor and multiple sensors are proposed respectively. Samuel [5] proposed a constrained adaptive algorithm to detect the damage of the gearbox based on lifting wavelet method. Williams and Zalubas [6] used Wigner-Vine distribution to detect the helicopter planetary gear box fault. Bartelmus and Zimroz [7] applied the cyclic stationary analysis method to study the modulation characteristics of planetary gearbox. Barszcz and Randall [8] applied spectral kurtosis to detect the crack of gear teeth in the wind planetary gearbox. Hu Niaoqing [9, 10] proposed energy characteristics in the neighborhood of meshing frequency method based on Hilbert Huang transform (HHT) and gray correlation analysis method to detect the sun gear fault in the planetary gearbox. Feng Zhipeng [11, 12] proposed the fault vibration model based on the meshing frequency of the planetary gearbox, amplitude demodulation analysis method and its vibration signal model. The fault diagnosis of planetary gear box is realized by the frequency demodulation method based on empirical mode decomposition.

2 Proposed Diagnosis Method

The flow chart of fault diagnosis method that combines maximum correlated kurtosis deconvolution (MCKD) with frequency slice wavelet transform (FSWT) of planetary gearbox is shown in Fig. 1. This paper firstly acquires the gear fault vibration signal via the signal acquisition equipment and selects the appropriate shock period, venue and the number of iteration steps, while using particle swarm optimization algorithm to find out the optimal filter length. Then, the fault signal of gear is denoised by MCKD in time domain. Finally, the gear fault feature is extracted and the fault resource is detected by applying FSWT method to analyse the denoised signal.

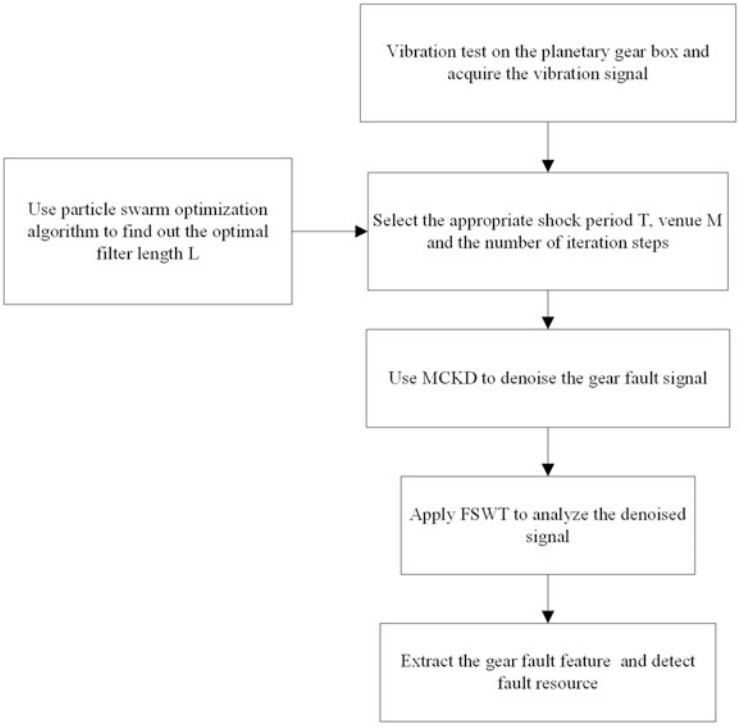


Fig. 1 Flow chart of fault diagnosis method of planetary gearbox

2.1 The Maximum Correlated Kurtosis Deconvolution Based on Particle Swarm Optimization Algorithm

The maximum correlation deconvolution [13] takes related kurtosis as the evaluation index, give full consideration to the signal containing the periodicity of the impact component. The cycle impact of gear fault will appear in the measured signal, in order to extract the signal of periodic impact characteristics, the maximum correlation kurtosis solution convolution method get the signal correlation kurtosis maximum value by filtering through the optimal filter.

Correlated Kurtosis (CK):

$$CK_M(T) = \max_f \frac{\sum_{n=1}^{N}\left(\prod_{m=0}^{M} y_{n-mT}\right)^2}{\left(\sum_{n=1}^{N} {y_n}^2\right)} f = [f_1 f_2 \ldots f_L]^T \tag{1}$$

In the formula, y_n is the periodic function, T is the period of signal y_n, f is the filter coefficient matrix, L is the length of the finite impulse filter, and M is the shift number.

In MCKD algorithm, the choice of filter length parameter L and reconciliation periodic convolution parameter T has an important influence on the deconvolution result. If a parameter remains the same, only take another parameter as the optimization object to discuss the effects of parameters on the calculation results, the local search methods ignore the interaction of two parameters, the obtained parameters are relatively optimal.

Particle swarm algorithm as a swarm intelligence optimization algorithm, can effectively reduce the influence and has good global searching ability, this paper use particle swarm optimization algorithm to optimize the two parameters of MCKD algorithm simultaneously, the adaptive filtering of the filter length parameter L and the reconciliation periodic convolution parameter T is realized.

Its principle is as follows, search the dimension Q space which consisting of the groups of N particles, each correlated particle can be expressed as: $x_i = (x_1, x_2, x_3, \ldots, x)$, the velocity of every correlated particle can be expressed as: $v_i = (v_{i1}, v_{i2}, v_{i3}, \ldots, v_{iQ})$. Each particle iterated updates its velocity and position by the local extreme value and population global extreme value, its velocity and position updating formula in particle swarm optimization algorithm are as follows:

$$\begin{cases} {v_{id}}^{k+1} = w{v_{id}}^{k} + c_1\xi\left({p_{id}}^{k} - {x_{id}}^{k}\right) + c_2\eta\left({g_{id}}^{k} - {x_{id}}^{k}\right) \\ {x_{id}}^{k+1} = {x_{id}}^{k} + {v_{id}}^{k+1} \end{cases} \tag{2}$$

In the formula, W is inertia weight, d = 1, 2. . .Q. c1 represents the recognition of the particle itself; c_2 represents the recognition of the particle group, ξ, η are random numbers which are the uniform distribution between 0 and 1.

In the particle swarm optimization algorithm, it need determine a fitness value. This paper selects the filtered signal envelope entropy that is used as the fitness value, which has smaller envelope entropy and the better effect. Envelope entropy E_g is shown as follows:

$$\begin{cases} E_p = -\sum_{j=1}^{N} p_j \lg p_j \\ p_j = a(j) / \sum_{j=1}^{N} a(j) \end{cases} \tag{3}$$

In the formula, p_j is a (j) in the normalized form, a (j) is the envelope of the signal y (j) after Hilbert demodulation.

The main words in all headings (even run-in headings) begin with a capital letter. Articles, conjunctions and prepositions are the only words which should begin with a lower case letter.

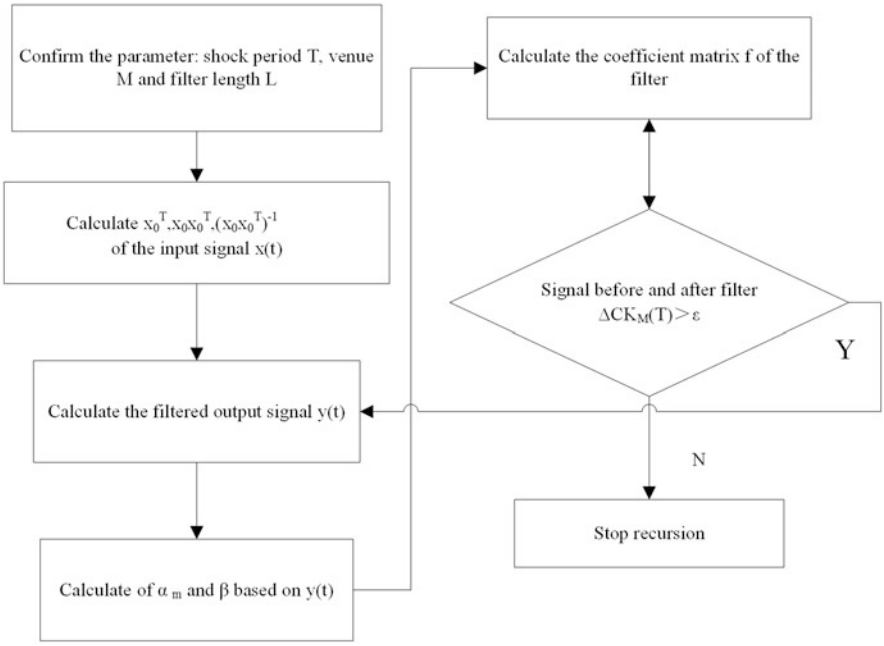


Fig. 2 Flow chart of iterative solution of filter parameters

Operate the MCKD to the signal and calculate the entropy in the envelope of the convolution solution signal with the position of any particle X_i (i.e. parameter L and T), the entropy $E_g(X_i)$ represents the fitness value of the particle. When the solution of convolution appears cyclical shocks, the envelope entropy becomes smaller and the solution the convolution effect becomes less effective and the regularity of impact becomes less obvious, but envelope entropy is relatively large.

Therefore, take the minimization of convolution signal envelope entropy as the final optimization goal, the optimization process of determining the reconciliation periodic convolution parameter T and the filter length parameter L though MCKD is shown in Fig. 2.

2.2 Frequency Slice Wavelet Transform

When mechanical and electrical equipment monitoring conditions, signals obtained from the site are non-stationary characteristics, separately using of the methods of time domain or frequency domain analysis cannot effectively isolate the correct feature information. Therefore, the time-frequency analysis method is more widely used to reveal the information from site signal. Yan presents a new method of time-frequency analysis which called frequency slice wavelet transform [14]. The frequency slice function inherits the characteristics of wavelet function. Adjustable

scale factor makes the time-frequency resolution of FWST controlled by time frequency resolution. Moreover, the inverse transform doesn't like depend on wavelet function like wavelet transform anymore. Thus, time frequency domain segmentation can be flexibly performed in time frequency space, and the desired signal component is separated from the reconstruction [15]. The principle of the method is as follows:

Assume the signal $f(x) \in L^2(R)$, if there exists a Fourier transform $\widehat{p}(w)$ of $p(t)$, the frequency slice wavelet transform of it can be expressed as:

$$W(t,\omega,\lambda,\delta) = \frac{1}{2\pi}\lambda \int_{-\infty}^{+\infty} \widehat{f}(u)\widehat{p}^*\left(\frac{u-\omega}{\delta}\right)e^{iut}du \tag{4}$$

In the formula, δ is the scale factor, $\delta = 0$; λ is the energy coefficient, $\lambda = 0$. δ, λ is constant or a function of the frequency w, u and t. In FSWT, $\widehat{p}(u)$ is frequency domain form of mother wavelet function $p(t)$, the wavelet function $\lambda\varphi\left(\frac{u-\omega}{\delta}\right)$ is the results of telescopic translation in frequency domain, $\widehat{p}^*(w)$ is the conjugate function of $\widehat{p}(w)$. Though the formula above, it's calculated that $\nu = e^{-\mu/2}, \kappa = \sqrt{2}/(2\eta)$.

3 The Validation of the Proposed Method

3.1 Experimental Introduction

In this paper, the gear fault vibration signal in the NGW11-12.5 planetary gear box is obtained on the test rig with the DH5927 vibration acquisition instrument. The test rig contains the motor, planetary gearbox, transmission shaft, magnetic powder brake and tension controller (loaded), computer and the DH5927 vibration acquisition instrument, as shown in Fig. 3a. Planetary gear box and the left end of the motor are connected, and the right end is connected to the transmission shaft to reduce speed. The speed of the motor is 1260 r/min and the secondary planetary gear transmission ratio is 12.5:1. The frequency parameters of planetary gearbox are shown in Table 1.

In order to validate the feasibility of the proposed method, sun gear is line cut in a tooth to simulate tooth breaking fault, as shown in Fig. 3b.

3.2 Signal Acquirement

In this test, the sampling frequency of DH5927 vibration acquisition instrument is set 1000 Hz, sampling time is set 67.2 s, sampling points is set 67,200 and the rotating frequency is 21 Hz. The measured waveform in time domain and frequency domain diagrams of the fault vibration signal is presented in Fig. 4. As is shown in

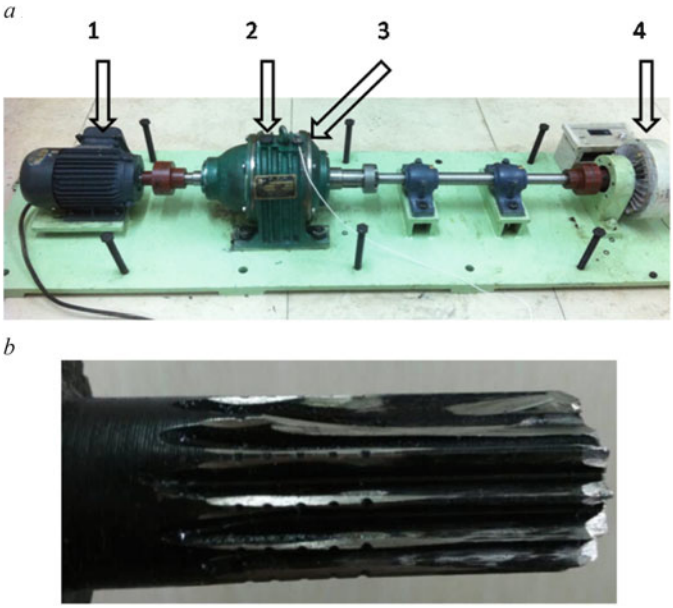


Fig. 3 Planetary gearbox test rig. (**a**) Planetary gearbox test rig. (**b**) Tooth broken sun gear

Table 1 Frequency parameters of planetary gearbox

Input shaft rotating frequency f_0	21 Hz
Gear meshing frequency f_1	249.5 Hz
Planet carrier frequency f_2	1.72 Hz
Self-rotating frequency of planetary gear f_3	2.08 Hz
Fault frequency of sun gear f_4	57.85 Hz

the diagram that the waveform in time domain and frequency domain diagrams is complex and difficult to distinguish the specific features of signals. Thus, the fault feature is submerged in the background noise signal. It could be clearly seen that the second harmonic component of the rotating frequency is prominent and the meshing frequency also have appeared in the diagram, but the amplitude is not obvious and the fault feature frequency hasn't appeared.

3.3 Signal Denoising

In the process of de-noising with the application of MCKD, the parameters are selected as follow: the impact period T is 43, the shift number M is 5 and the iteration number is 30.The filter length L is calculated to set 30 by using particle

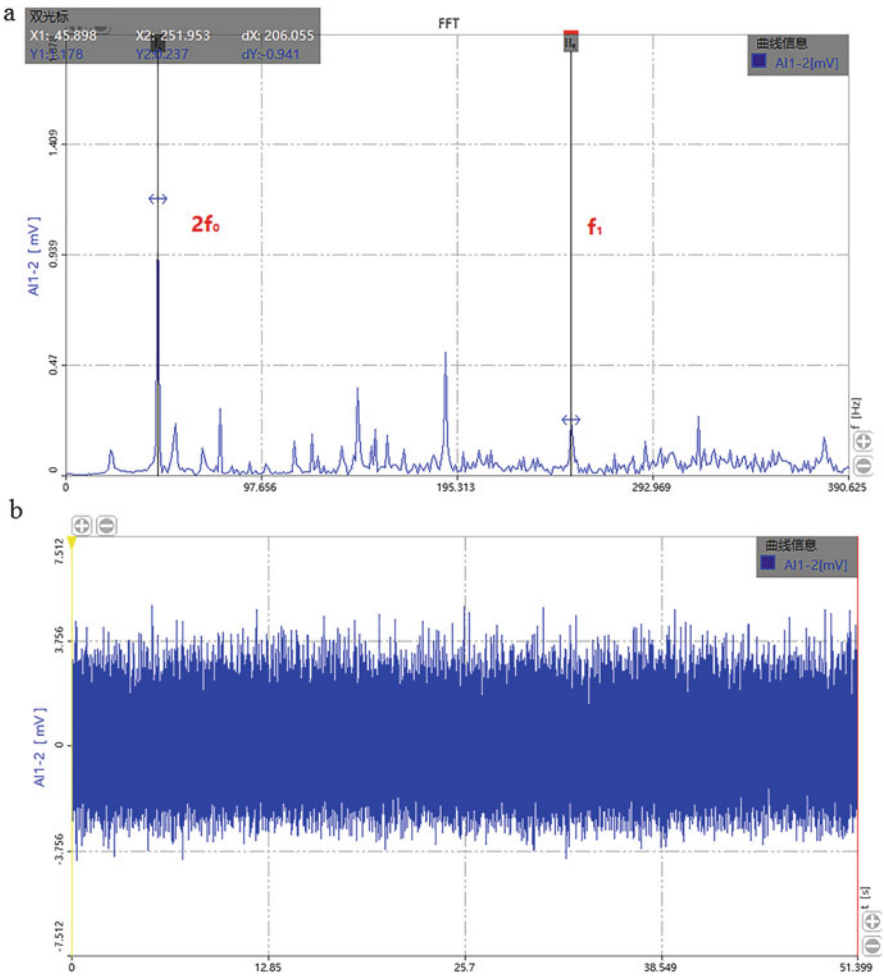


Fig. 4 Diagrams of the waveform in time and frequency domain. (**a**) Waveform in frequency domain. (**b**) Waveform in time domain

swarm optimization algorithm when the envelope entropy value of the convolution signal is minimum.

After determining the parameters, the maximum correlation kurtosis deconvolution is used to reduce the measured waveform of solar gear broken teeth in time domain, and the original signal and the denoised signal are shown in Fig. 5.

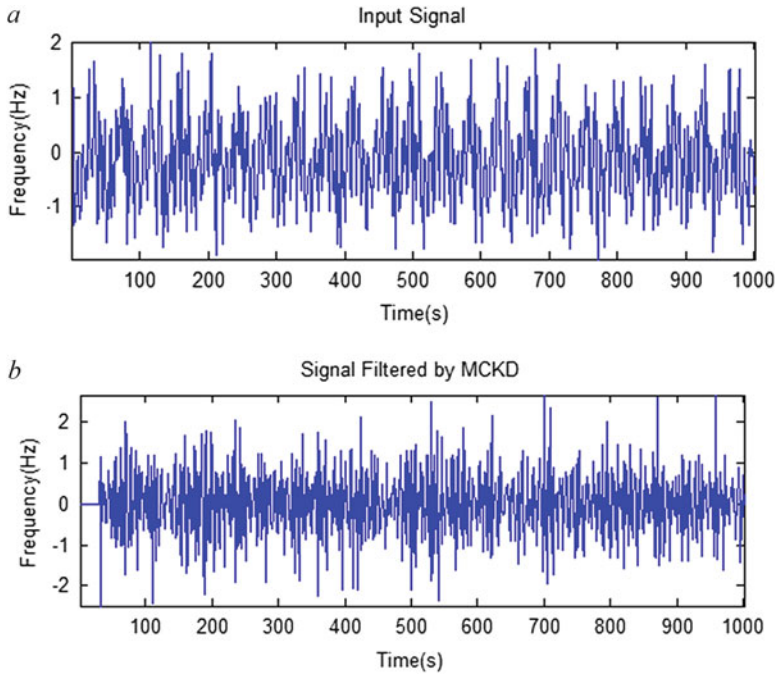


Fig. 5 Signals before and after noise de-noising. (**a**) Input signal. (**b**) Signal filtered by MCKD

3.4 Signal Time-Frequency Analysis

Analyse the filtered signal with frequency slice wavelet transform after MCKD, assume $\kappa = \sqrt{-2\ln v}/\eta$, $v = \sqrt{2}/2, \eta = 0.010$. Then, it is calculated that κ equals 83.26. So the parameters in the frequency slice wavelet transform can be selected: $\widehat{p}(\omega) = e^{-\omega^2/2}, \lambda = 1, \delta = \omega/\kappa$, the frequency slice interval selection is chosen [0,350] Hz.

The FSWT results both before and after signal denoising are presented in Figs. 6 and 7, it is obviously found in the diagrams that after MCKD denoised, the features of shock is strengthened, and the sun gear fault feature frequency 57.69 Hz and its 2X–5X harmonic generation are shown in the latter diagram and the energy is very high. In addition, the quadruple rotating frequency of sun gear and the meshing frequency 249.8 Hz also appear in frequency diagram. So the denoising and fault diagnosis method for signals in time domain in planetary gearbox are comparably applicable.

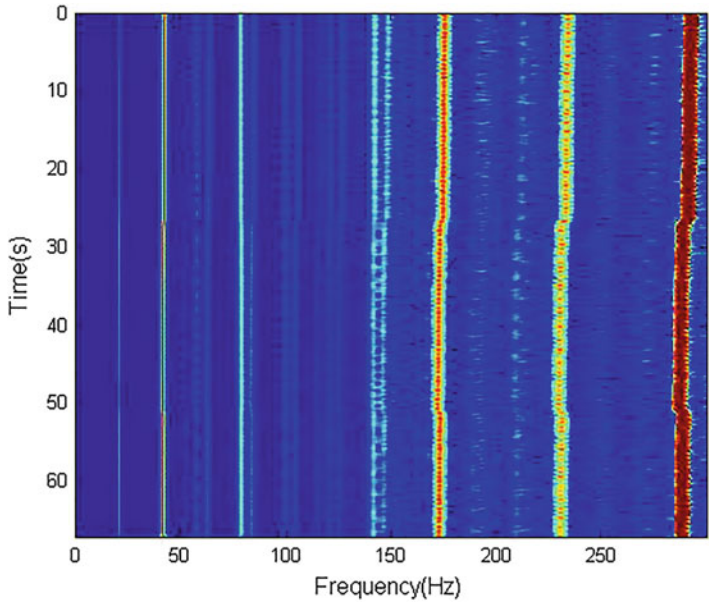


Fig. 6 FSWT time-frequency diagram of signal before denoising

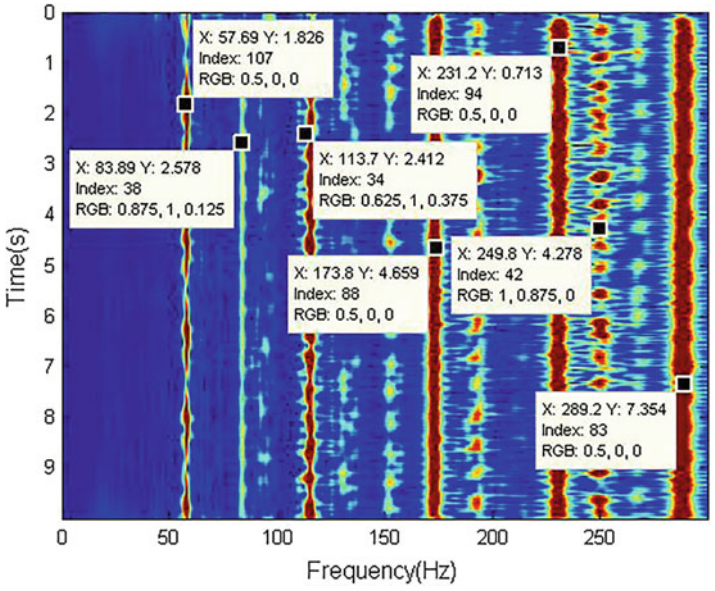


Fig. 7 FSWT time-frequency diagram of signal after denoising

4 Conclusions

This paper presents a new method for fault diagnosis of the planetary gearbox combined MCKD with particle swarm optimization algorithm and FSWT. In the proposed method, MCKD is used to de-noise the raw signal, the features of periodic impulse components are strengthened after noise reduction. The time-frequency distribution of the vibration signals could be done using FSWT to identify the gear fault. Based on above study, the following conclusions are drawn:

For early fault of planetary gearbox, the fault signal feature is relatively weak and largely effected by environmental noise and transmission path. Thus, the fault feature extraction is relatively difficult. Therefore, the maximum correlation kurtosis deconvolution is used to efficiently prominent periodic impulse component in the signal and to suppress other noise involved.

The frequency slice wavelet function is introduced to enable the traditional Fourier transform realize time-frequency analysis function, which makes the signal filtering and segmentation more flexible and the fault feature of the planetary gear box can be clearly identify.

Acknowledgments Supported by the Opening Project of Key Laboratory of operation safety technology on transport vehicles, Ministry of Transport, PRC.

References

1. Long Q, Liu Y, Yang Y (2008) Applications of condition monitoring and fault diagnosis to wind turbines. Modern Electric Power
2. Mcfadden PD (1991) A technique for calculating the time domain averages of the vibration of the individual planet gears and the sun gear in an epicyclic gearbox. J Sound Vib 144(1): 163–172
3. Mcfadden PD (1994) Window functions for the calculation of the time domain averages of the vibration of the individual planet gears and sun gear in an epicyclic gearbox. J Vib Acoust 116(2):179–187
4. Pines DJ (2000) Vibration separation methodology for planetary gear health monitoring. Proc SPIE – Int Soc Opt Eng 3985:250–260
5. Samuel PD, Pines DJ (2009) Constrained adaptive lifting and the CAL4 metric for helicopter transmission diagnostics. J Sound Vib 319(1–2):698–718
6. Williams WJ, Zalubas EJ (2000) Helicopter transmission fault detection via time-frequency, scale and spectral methods. Mech Syst Signal Process 14(4):545–559
7. Zimroz R, Bartelmus W (2009) Gearbox condition estimation using cyclo-stationary properties of vibration signal. Key Eng Mater 413(1):471–478
8. Barszcz T, Randall RB (2009) Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine. Mech Syst Signal Process 23(4):1352–1365
9. Feng Z, Hu N, Cheng Z (2010) Faults detection of a planetary gear based on condition indicator in time-frequency domain. Mechanical Science & Technology for Aerospace Engineering
10. Zhe C, Niao-Qing HU, Gao JW (2011) Scuffing damage quantitative detection of planetary gear set based on physical model and grey relational analysis. J Vib Eng 24(2):199–204

11. Feng Z, Zhao L, Chu F (2013a) Amplitude demodulation analysis for fault diagnosis of planetary gearboxes. Zhongguo Dianji Gongcheng Xuebao/Proc Chin Soc Elect Eng 33(8): 107–111
12. Feng Z, Zhao L, Chu F (2013b) Vibration spectral characteristics of localized gear fault of planetary gearboxes. Proc CSEE 33(2):118–125
13. Tang G, Wang X, Energy SO (2015) Adaptive maximum correlated kurtosis deconvolution method and its application on incipient fault diagnosis of bearing. Zhongguo Dianji Gongcheng Xuebao/Proc Chin Soc Elect Eng 35(6):1436–1444
14. Zhong X (2014) Research on time-frequency analysis methods and its applications to rotating machinery fault diagnosis. Wuhan University of Science and Technology
15. Duan C, Qiang G, Xianfeng XU (2013) Generator unit fault diagnosis using the frequency slice wavelet transform time-frequency analysis method. Proc CSEE 33(32):96–103

Bridge Condition Assessment Under Moving Loads Using Multi-sensor Measurements and Vibration Phase Technology

Hong Hao, Weiwei Zhang, Jun Li, and Hongwei Ma

Abstract This paper presents a bridge condition assessment approach under moving loads with multi-sensor measurements and vibration phase technology. The phase trajectories of multi-sensor responses are obtained and a damage index is defined as the separated distance between the trajectories of undamaged and damaged structures to identify the damage existence and location. The damage will induce the local structural stiffness change, and the distance between the corresponding points on those two trajectories of undamaged and damaged beams when a moving load travels across the damaged location will also change. Experimental studies demonstrate the proposed approach can be used to successfully identify the shear connection failure in a composite bridge model.

Keywords Bridge • Damage detection • Vibration phase • Multi-sensor

1 Introduction

Dynamic responses of bridge structures subjected to moving load could be used for assessing structural conditions [1]. In practical applications, the properties of the moving vehicle could not be obtained and thus they are usually assumed as unknown parameters. Zhang et al. [2] proposed a method for simultaneous identification of moving masses and structural damage from measured responses. Zhu and Law [3] performed the simultaneous identification of the moving forces and

H. Hao • J. Li (✉)
Center for Infrastructural Monitoring and Protection, Curtin University, Kent Street, Bentley, WA 6102, Australia
e-mail: hong.hao@curtin.edu.au; junli@curtin.edu.au

W. Zhang
Department of Mechanics, Taiyuan University of Science and Technology, Taiyuan, Shanxi 030024, China
e-mail: zwwps@tyust.edu.cn

H. Ma
College of Science and Engineering, Jinan University, Guangzhou, Guangdong 510632, China
e-mail: tmahw@jnu.edu.cn

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_7

structural damage iteratively by using a two-step identification procedure. Later, Law and Li [4] conducted structural damage identification of a three-span box-section concrete bridge deck subjected to a moving vehicle modelled by a three-dimensional mathematical model. To improve the accuracy in structural damage identification, which may be influenced by the accuracy of the identified moving loads, Li et al. [5] proposed an improved damage identification approach for the bridge structures subjected to moving loads with numerical and experimental validations, without the need to identify the moving forces as well as the properties of the moving vehicle.

The above studies mainly investigate the damage detection problems with modal information or vibration testing measurements for bridge structures under moving loads with model updating methods. The main difficulty of model based methods is to obtain an accurate finite element model to well represent the real bridge structure. With the development of advanced signal processing techniques, there have been a growing number of studies in the recent two decades for non-model based damage detection in bridge structures. For example, numerous studies have been conducted by performing the time-frequency analysis, such as the wavelet transform [6] and Hilbert-Huang transform (HHT) [7, 8] of displacement response, and wavelet transform of acceleration response [9]. The ideas of the above-mentioned studies are based on the fact that the damage can be indicated by the abnormality or singularity in the dynamic responses and it is visualized through the appearance of a highlighted oscillation in the signal features when the moving load passes the damage location in the bridge structure.

Usually only an individual type of dynamic response quantity, i.e. displacement or acceleration, is used in the above-mentioned studies to detect the structural damage. This could limit the sensitivity and accuracy of damage identification. Law and Zhu [10] investigated the dynamic behaviour of damaged concrete bridge structures under moving loads. The vibration "displacement-velocity" phase plane at the mid-span of a simply-supported beam subjected to a moving load is illustrated. It was observed that there is a clear difference in the phase plane plots between undamaged and damaged structural states. This indicates that damage detection using multi-type vibration measurements may perform better than that only using a single type of response measurement. Pakrashi et al. [11] applied continuous wavelet transform to observe the significant distortion in wavelet coefficients and detect the presence of damages. The measured strain and its derivative were plotted in a phase plane to track the evolution of damage conditions. Nie et al. [12] analysed measured strain signals on a steel arch to reconstruct the phase space and detect structural local damage. Later, this method was used to process the acceleration response of a continuous RC beam under a hammer impact load. A damage index termed as Change of Phase Space Topology (CPST) with multiple embedding dimensions was defined and used to identify the structural damage [13]. It was demonstrated in the study that this method is very sensitive to the structural damage. In the latter study only a single type of vibration response, i.e., strain or acceleration is used. It is believed that the development of a good index to describe the distortion of the phase space is of great potential to identify the

damage. However, this has not been fully explored and tested, particularly when multi-type vibration measurements are acquired from bridge structures under moving loads and used for damage detection.

2 Vibration Phase Technology

In a phase space, every system variable is represented as an axis of a multidimensional space. A one-dimensional system is called as a phase line, while a two-dimensional and a three-dimensional one as a phase plane and a phase space, respectively. Any dynamic system can be described in a phase space, which may be reconstructed from the measured time domain responses. In this paper, a multi-dimensional phase space will be constructed, for example, a three dimensional one with the three axes corresponding to the displacement, velocity and acceleration respectively. Taking a simply-supported beam subjected to a moving load with a constant moving speed as an example, the phase space trajectory can be plotted directly with the displacement, velocity and acceleration responses measured at the mid-span of the beam, which are the function of time t,

$$u = u(t), v = v(t), a = a(t) \tag{1}$$

For convenience, (u_0, v_0, a_0) and (u_d, v_d, a_d) denote the spatial coordinates on the phase trajectory from the undamaged and damaged beam models, respectively. The damage will induce the local structural stiffness change, and the distance between the corresponding points on those two trajectories of undamaged and damaged beams when a moving load travels across the damaged location will also change. The phase trajectory of the undamaged beam is considered as the baseline, a damage index based on the change in the phase space trajectories is used to identify the structural damage.

With the above mentioned bridge-vehicle system, the time axis could be normalized as the location of the moving load by $x = ct$, where c and t are the travelling speed and time, respectively. In a general case, the location of the moving load is often expressed as the normalized location by $x' = x/L$, in which L is the length of the beam. The damage index is defined as the distance between the spatial coordinates of the normalized moving load locations from the undamaged and damaged phase trajectories

$$DI = \sqrt{\left(u_d^{x'} - u_0^{x'}\right)^2 + \left(v_d^{x'} - v_0^{x'}\right)^2 + \left(a_d^{x'} - a_0^{x'}\right)^2} \tag{2}$$

where $\left(u_d^{x'},\ v_d^{x'},\ a_d^{x'}\right)$ and $\left(u_0^{x'},\ v_0^{x'},\ a_0^{x'}\right)$ denote the spatial coordinates on the damaged and undamaged phase space trajectories of the normalized locations of the moving load. It is noted that by using the normalized locations of the moving load on the bridge to define the damage index, the proposed approach can be used for

scenarios with different travelling speeds of the moving loads before and after damage.

To overcome the impact effect when the vehicle starts and stops as well as the measurement noise effect, a simple moving average (SMA) [14] process is used to smooth the responses. It is the unweighted mean of a specific number of data points. SMA is commonly used with time series data to smooth out short-term fluctuations and highlight longer-term trends. Supposing x is a time series, SMA is written as

$$x_s(i) = \frac{1}{N}\sum_{j=0}^{N-1} x(i+j) \tag{3}$$

where x_s and x represent the smoothed result and the original series, respectively, and N is the number of points used to calculate the average. SMA is used to process the dynamic vibration responses and then the phase trajectory is plotted.

Multiple types of vibration responses are used to construct the multi-dimensional vibration phase space and plot the trajectory, such as a three dimensional phase space "displacement-velocity-acceleration", or a two dimensional phase space "displacement-acceleration". By analyzing the phase space trajectories and calculating the damage index in the three dimensional or two dimensional phase space, the local maximum value could be used to indicate the damage location

3 Experimental Verifications

3.1 Experimental Testing Model

To demonstrate the applicability and effectiveness of the proposed approach, experimental verifications on a composite bridge model in the laboratory are conducted. The composite bridge model was constructed with a reinforced concrete slab supported on two steel I-type girders, as shown in Fig. 1. Thirty-two bolts

Fig. 1 Experimental testing model

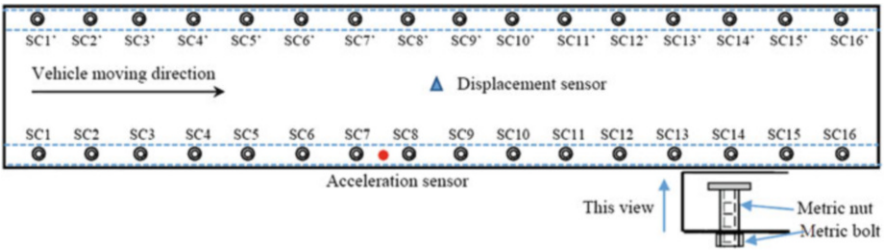


Fig. 2 Design of shear connectors and placed sensors

Fig. 3 Bridge-vehicle system

screwing into metric nuts which were cast in the slab were used to link the slab and girder together. The nuts were welded onto the reinforcement bar in the slab before pouring. Each girder has 16 connectors with an equal space of 200 mm and there are 32 in total in the bridge model, which are denoted as SC1~SC16 and SC1′~SC16′, as shown in Fig. 2. The bridge model is placed on two steel frames which are fixed to the laboratory strong floor. Figure 3 shows the experimental "bridge-vehicle" system. The vehicle model is simplified as a steel beam carrying two concrete blocks. The distance between the front axle and rear axle of the vehicle is 1 m. A roller system is designed and the crane is used to pull the vehicle to travel on the top of the bridge model with a constant speed as the pulling force of the crane is steady and stable. A fabricated track is placed and clamped on the top of the concrete slab to make sure the vehicle is moving on the predetermined traveling path. Three transverse connecting plates are located at the two ends and the center of the track to ensure the stability and rigidity of the track, as shown in Fig. 1. Subjected to the test conditions, the velocity response is not measured in the experimental tests, and the displacement and acceleration responses are used for the analysis. Only a displacement sensor and an accelerometer are used, and their locations are marked in Fig. 2. The sampling rate is set as 2000 Hz.

3.2 Condition Assessment Results

When all the connectors are tight into the nut, the bridge is defined as the undamaged state. If a specific connector is removed from the nut, it indicates the failure of the particular shear connector which also shows that the bridge is under the damaged state. In this study, two damage scenarios and their associated damaged shear connectors are shown in Table 1. The regions of those introduced damages in these two damage scenarios are also listed. The measurements from both the undamaged and damaged states of the bridge are taken. Figure 4 shows the measured displacement and acceleration responses from the undamaged model. The time instants when the vehicle starts and stops can be observed from the measured responses. In addition, there are some unexpected oscillations in acceleration responses. They are probably induced by the impact on the track when the front and rear axles of the vehicle are crossing the connecting transverse plate in the centre of the track. The Fourier spectrum of the acceleration response from the undamaged state and Damage Scenario 1 are shown in Fig. 5. The fundamental

Table 1 Damage scenarios in experimental studies

Damage scenario		Removed shear connectors	Damage region
Scenario 1	damage 1	SC15 and SC15′	2.8 m
Scenario 2	damage 1	SC15 and SC15′, SC14 and SC14′	2.6–2.8 m
	damage 2	SC2and SC2′, S3 and SC3′	0.2–0.4 m

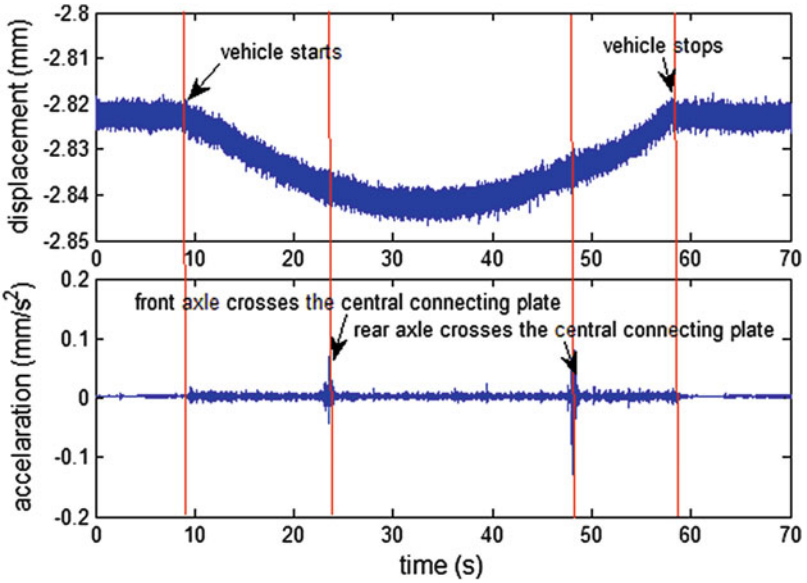


Fig. 4 Measured displacement and acceleration from the undamaged state

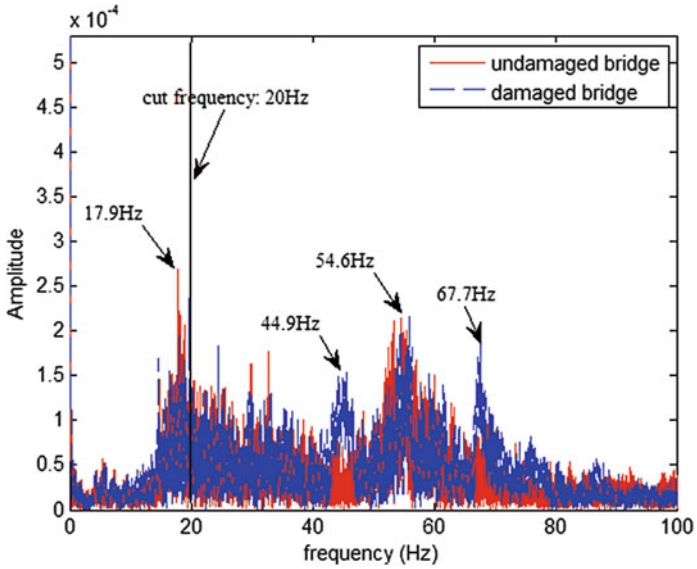


Fig. 5 Fourier spectrum of the acceleration response from the undamaged state and Damage Scenario 1

frequency of the bridge-vehicle system is identified as 17.9 Hz from the measurements of the undamaged bridge. There are two main frequencies components at 17.9 and 54.6 Hz, as observed from Fig. 5. It is noted that no significant differences are observed at these two frequencies, therefore it may be difficult to use the frequency changes to locate the shear connector damage since the damage of the local shear connector might not prominently affect the global stiffness of the bridge and the natural frequency. In this study, the measured responses are pre-processed with a low pass filter with the cut-off frequency of 20 Hz and then used for reconstructing the phase space and calculate the damage index.

Figure 6a, b shows the comparison of the phase trajectories between the undamaged and damaged states for Damage Scenario 1 and Scenario 2, respectively. It is noted that clear difference can be observed from the undamaged and damaged phase trajectories in those two scenarios, indicating the sensitivity of the phase trajectory on changes of structural conditions. Since Damage Scenario 2 has more severe damages with more loosen bolts, the difference between the phase trajectories is more prominent than that in Scenario 1. Equation (2) is followed to calculate the damage index, and the damage detection results will be discussed.

Since only the displacement and acceleration responses are measured, a two-dimensional phase trajectory is reconstructed and used to calculate the damage index. 250 data points is used with the SMA scheme to reduce the noise effect and improve the performance of damage detection. The damage detection results for those two damage scenarios are shown in Figs. 7 and 8, respectively.

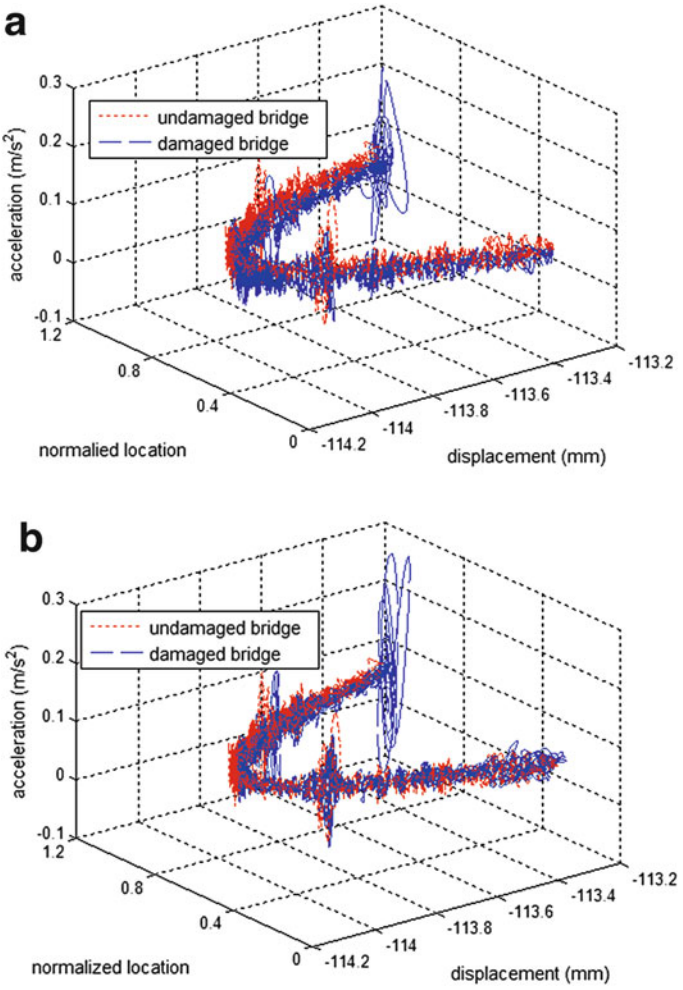


Fig. 6 Comparison of the phase trajectories between the undamaged and damage states. (**a**) Scenario 1 and (**b**) Scenario 2

The travelling path of the bridge-vehicle system, the distance of the front and rear axles of the vehicle and the number of shear connectors are shown in Fig. 9. Damage scenario 1 is associated with the removal of shear connectors SC15 and SC15′, and Damage scenario 2 associated with shear connectors SC2 and SC2′ being removed. Only the front axle will pass along the region of removed shear connectors in Damage scenario 1, and only the rear axle will pass the area of removed shear connectors in damage scenario 2. The highest damage index as shown in Fig. 7 indicates the location of damage 1 when the front axle moves on the top of damage. It is roughly located at the normalized location 0.91, which matches

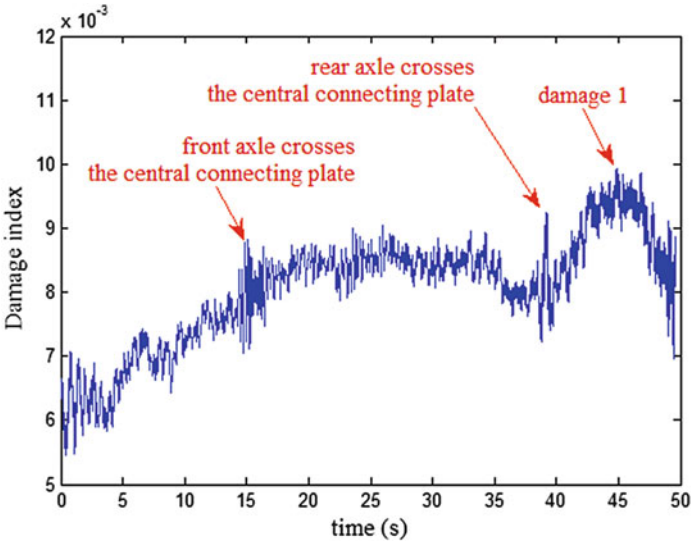


Fig. 7 Damage assessment results for Scenario 1

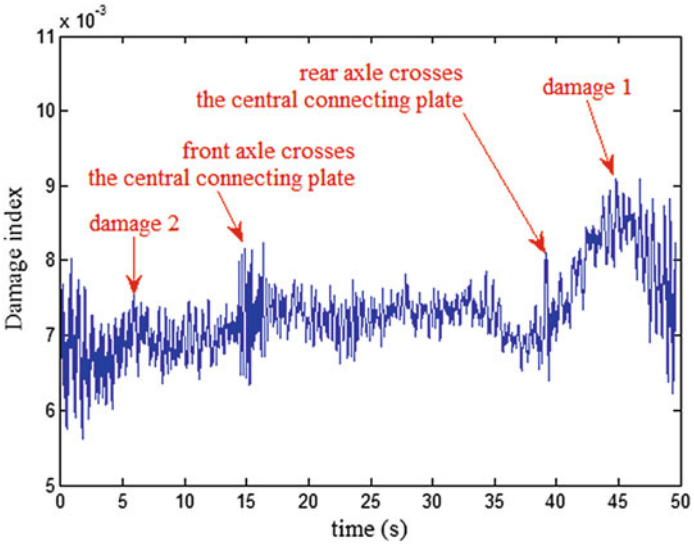


Fig. 8 Damage assessment results for Scenario 2

well with the true introduced damaged location at 2.8m/3 m = 0.93, indicating that the damage can be identified effectively for the Damage Scenario 1 with the proposed approach. The other two peak damage index values at the normalized locations of 0.3 and 0.8 are corresponding to the instants when the front and rear

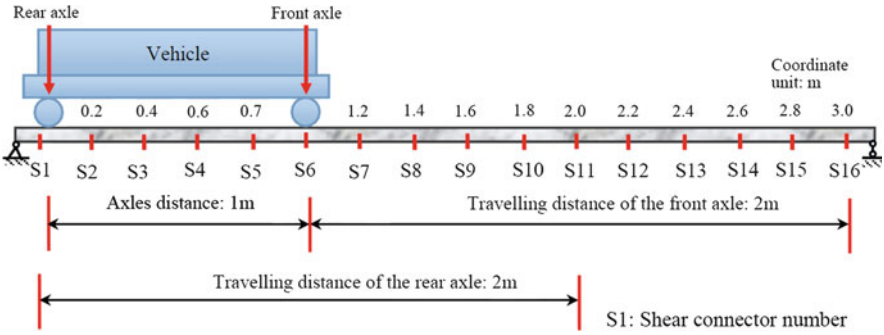


Fig. 9 The schematic diagram of the bridge-vehicle system and travelling path

Table 2 Damage assessment results in experimental studies

Damage scenario		True damage location	Identification results	
			Time instant (s)	Location (m)
Scenario 1	Damage 1	2.8 m	44.89	2.79
Scenario 2	Damage 1	2.6–2.8 m	44.81	2.79
	Damage 2	0.2–0.4 m	5.90	0.24

axles impact the track at the central transverse plate. This is also observed in the acceleration measurements as shown in Fig. 4. For Damage Scenario 2, the detection result is shown in Fig. 8. Besides the high damage index values associated with the locations of damage 1, the instants of the front and rear axles moving across the central transverse plate of the track as mentioned above, one more relatively high damage index value at the normalized location 0.1 corresponds to the instant when the rear axle moves on the top the location of damage 2, which is at 0.2m/3 m = 0.67. This can be used to identify the location of damage 2.

When the time instants of the identified damaged locations as shown in Figs. 7 and 8 are obtained, the damaged locations could be determined based on the vehicle speed from the measurement record. It is worthy to note that the real travelling distance of the vehicle is about 2 m instead of 3 m because the distance between the two axles of the vehicle is 1 m, as shown in Fig. 9. In this regard, damage 1 can be detected by using the response when the front axle crosses this region, and damage 2 may be detected by the rear axle. Supposing c and t_1 are the moving speed of the vehicle and the identified time instant when the front axle is crossing the location of damage 1, the damage region can be obtained as $L_{x1} = ct_1 + L_v$, in which L_v is the distance between the two axles of the vehicle. For damage 2, the location can be calculated as $L_{x2} = ct_2$ with t_2 identified as the time instant of the rear axle crossing the location of damage 2. The speed of vehicle c can be obtained as the average speed when the vehicle moves along the bridge. The total time during the vehicle travelling on the top of the bridge model is 49.44 s, the vehicle speed is obtained as 0.04 m/s. Table 2 shows the true and identified damage locations in the

abovementioned two damage scenarios. The identification results demonstrate a good accuracy and effectiveness of the proposed approach to identify the shear connection damages in composite bridges.

The above results demonstrated that the proposed damage index based on changes in phase trajectories are very sensitive to changes in structural conditions. Introduced structural damage by removing shear connectors in the present examples can be clearly identified, while such damages are difficult to be identified by using more traditional vibration-based parameters such as change in vibration frequencies. However, it is also noted that while this damage index is very sensitive to structural damage, it is also strongly influenced by other factors, such as bridge surface conditions that induce acceleration responses of vehicles as shown in Figs. 7 and 8 of the peaks associated with the axles passing the center of connecting plate. Further studies are deemed necessary to refine the definitions of the damage index or pre-processing the recorded signals to make the damage index insensitive or less sensitive to noises and surface conditions of bridge pavement but only sensitive to structural damages.

4 Conclusions

Vibration tests of bridges subjected to moving vehicle loads could be an effective approach to examine the bridge conditions and improve the performance of the current damage detection approaches for bridge structures. Uncertainties in testing, environmental noise effect and coupling vibrations of the bridge-vehicle system may significantly affect the accuracy and effectiveness of damage detection. This paper presents a condition assessment approach for bridge structures under moving loads with multi-sensor responses and vibration phase technology. Taken the phase trajectory from the undamaged bridge as a reference, the separation distance between undamaged and damaged bridge states could be calculated as a damage index to indicate the damage location.

Experimental studies demonstrate the proposed damage index is very sensitive to structural condition change and can be used to successfully identify the shear connection failure in a composite bridge model with measured displacement and acceleration responses. Two damage scenarios, namely, a single damage and two local damages, are considered by removing specific shear bolts in the composite bridge model. Damage locations in both scenarios can be effectively detected. It is well demonstrated that the proposed approach has a good performance to detect the local damage in bridges under moving loads while such local damages are difficult to be successfully detected with the traditional vibration-based damage indices such as vibration frequencies. However, it is also noted that the proposed damage index is also sensitive to other bridge parameters such as surface conditions that induce acceleration responses. Further study is deemed necessary to refine the approach, making it sensitive to only bridge damages.

Acknowledgments Financial supports from National Natural Science Foundation of China (Grant No. 11102125), and Australian Research Council Discovery Early Career Researcher Award DE140101741 are acknowledged

References

1. Li J, Hao H (2016) A review of recent research advances on structural health monitoring in Western Australia. Struct Monit Maint 3(1):33–49
2. Zhang Q, Jankowski L, Duan Z (2010) Simultaneous identification of moving masses and structural damage. Struct Multidiscip Optim 42(6):907–922
3. Zhu XQ, Law SS (2007) Damage detection in simply supported concrete bridge structure under moving vehicular loads. J Vib Acoust 129(1):58–65
4. Law SS, Li J (2010) Updating the reliability of a concrete bridge structure based on condition assessment with uncertainties. Eng Struct 32(1):286–296
5. Li J, Law SS, Hao H (2013) Improved damage identification in a bridge structure subject to moving vehicular loads: numerical and experimental studies. Int J Mech Sci 74:99–111
6. Zhu XQ, Law SS (2006) Wavelet-based crack identification of bridge beam from operational deflection time history. Int J Solids Struct 43(7–8):2299–2317
7. Li J, Hao H (2015) Damage detection of shear connectors under moving loads with relative displacement measurements. Mech Syst Signal Pr 60–61:124–150
8. Roveri N, Carcaterra A (2012) Damage detection in structures under traveling loads by Hilbert-Huang transform. Mech Syst Signal Pr 28:128–144
9. Hester D, Gonzalez A (2012) A wavelet-based damage detection algorithm based on bridge acceleration response to a vehicle. Mech Syst Signal Pr 28:145–166
10. Law SS, Zhu XQ (2004) Dynamic behavior of damaged concrete bridge structures under moving vehicular loads. Eng Struct 26(9):1279–1293
11. Pakrashi V, O'Connor A, Basu B (2010) A bridge-vehicle interaction based experimental investigation of damage evolution. Struct Health Monit 9(4):285–296
12. Nie Z, Hao H, Ma H (2012) Using vibration phase space topology changes for structural damage detection. Struct Health Monit 11(5):538–557
13. Nie Z, Hao H, Ma H (2013) Structural damage detection based on the reconstructed phase space for reinforced concrete slab: experimental study. J Sound Vib 332(4):1061–1078
14. Smith SW (1999) The scientist and engineer's guide to digital signal processing, 2nd edn. California Technical Publishing, San Diego

EcoCon: A System for Monitoring Economic and Technical Performance of Maintenance

Anders Ingwald and Basim Al-Najjar

Abstract Maintenance has been treated as a cost-center although it has obvious direct impact on company production, delivery on time, product quality and consequently company business. In this paper we develop and test a new system (EcoCon) for monitoring and assessing the economic impact of maintenance on a production process. EcoCon is tested using real industrial data collected over 3 years. The main findings are that by using EcoCon it is possible to assess maintenance economic impact on a production process. Also, EcoCon provides data for analyzing causes behind deviations in maintenance and production performance. It can be applied on companies of similar production process/machine. But, in some cases, the system demands marginal accommodation to suit the differences in the maintenance economic related factors among companies. EcoCon provides production and maintenance mangers reliable overview of maintenance importance through monitoring maintenance economic impact by linking technical production performance, e.g. downtime, to economic savings/losses generated due to maintenance performance and other activities.

Keywords Maintenance management • Cost-effectiveness • Decision support

1 Introduction

Production processes has over the years become more complex, applying production philosophies like just in time (JIT) and involving automatic and advanced production techniques, [1–3]. This, together with harsh competition has increased the severity of downtime in the production processes. This makes the importance of maintenance greater due to its role in keeping and improving availability, performance efficiency, quality, on-time deliveries, environment, safety and total plant productivity, see for example [4–9].

The positive financial effects of maintenance are widely recognized in the scientific community, for example [7, 10–13]. In [7] the authors conducted a

A. Ingwald (✉) • B. Al-Najjar (✉)
Linnaeus University, P G Vejdes väg, 351 95 Växjö, Sweden
e-mail: anders.ingwald@lnu.se; basim.al-najjar@lnu.se

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_8

study in north east of England during 1997 and 1998 including 298 companies followed by a smaller study among 23 companies and they found that companies systematically adopting best practices in maintenance do achieve higher performance. Also, among companies it is more and more realized that effective maintenance can contribute significantly to companies profitability, even if it traditionally has been considered as a non-core function, see for example [4, 14]. Yet, a study conducted among Swedish industry gave as a result that 70% is considering maintenance only as a cost [15].

Several authors have pointed out the importance of decision making in maintenance management, e.g. what maintenance policy to use and when to stop production for conducting maintenance actions [16–19]. However, to be able to make efficient decisions in maintenance require, among other things, knowledge about maintenance technical and economic performance. When following up maintenance performance maintenance costs are traditionally divided into direct and indirect costs. Some of these costs, such as spare parts and direct labour, can comparably easy be related to maintenance. Also some of the indirect costs are also relatively easy to relate to maintenance, see for example [20]. However, savings and profit generated by maintenance are often neglected, and are not easy to find in current accounting systems. In order to assess the economic impact of maintenance its impact on life cycle cost/profit needs to be considered [18, 21]. This is usually not an easy task. Due to the complex nature of maintenance it is not possible to measure maintenance impact only in the maintenance function. Also its impact on other areas such as quality and production must be considered.

Consequently, in order to be able to use maintenance as a cost-effective tool to reduce production cost and increase company's profitability a system based on a holistic view of maintenance for monitoring its technical and economic impact, considering both costs and savings of maintenance and providing both an overview of maintenance performance and possibility to trace causes behind deviations in maintenance performance, is required, [16, 22]. Such a system will enable cost-effective decisions regarding maintenance, make it possible to follow up actions and support continuous improvement of the maintenance function [23]. In [24] is reported that maintenance initiatives often fail to deliver because the supportive systems such as management and information systems are not in place. Thus, the main purpose of this study is to develop and test a new system for monitoring and assessing the economic performance of the production process with respect to maintenance.

2 Maintenance and Its Technical and Financial Impact

In the standard [25], maintenance is defined as a combination of all technical, administrative and managerial actions during the life cycle of an item intended to retain it in, or restore it to, a state in which it can perform the required function. In this definition function is in focus, but maintenance is economically motivated. Maintenance can be performed according to several approaches, ranging from reactive and preventive maintenance to pro-active approaches involving condition

monitoring techniques and advanced methods for prognosis and prediction, see for example [18, 26, 27]. In [28] the authors summarized the strategically elements of maintenance decisions into structural decision elements regarding capacity, facilities, technology and vertical integration and into infrastructural decision elements regarding organization, policy and concept, planning and control systems, human resources, modification and performance measurement and reward systems. Decisions in each of these areas will have an impact on the technical and financial performance of the company. The technical consequences can be for example in form of changed availability and reliability. Furthermore, some of these effects are delayed and may only be seen over time. These technical effects of maintenance all have financial consequences, e.g. low availability, reliability and quality rate lead to increased production costs, increased costs for buffers and penalties for late delivery. In addition, the low internal effectiveness that can be a result from not having proper maintenance may lead to decreased market shares, profitability, etc. Several researchers have described maintenance impact on productivity and profitability of a company, for example [15, 29, 30].

In order to assess the real impact from maintenance the life cycle cost (LCC) concept need to be applied [18]. LCC is defined as the sum of research and development costs, production and construction costs, operation and maintenance support costs and retirement and disposal costs [31]. In this definition it is obvious that maintenance influence LCC. The maintenance cost includes several maintenance related cost factors, e.g. direct labour and spare parts but also costs due to unavailable, inefficiency, delivery delay penalties and redundant equipment. In addition to LCC it is often also necessary to assess the life cycle income (LCI) in order to estimate the economic importance of maintenance, which can be difficult. However, instead it is possible to assess the savings that can be gained from maintenance, e.g. reduced downtime, increased quality, and reduced capital tied in inventory and spare parts [10]. A prerequisite to apply cost-effective maintenance is a well-functioning maintenance performance measuring system [32]. The maintenance performance measuring system should assess the total economic impact of maintenance as well as indicating where to invest in improvements. This allows following up maintenance performance measures more frequently, thereby be able to intervene on deviations earlier and thereby avoid unnecessary costs. In [33] the authors discuss how to implement maintenance performance measuring system, and describe how different performance indicators (PI) can be constructed. The described PI:s are focused on the overall cost of maintenance, e.g. maintenance cost/production output. In a review of state-of-the art in maintenance performance metrics, authors in [34] refer to several frameworks for measuring maintenance performance, but it seems that none of these framework directly link decreased number of disturbances or improved quality to maintenance economic impact.

3 A System for Monitoring Maintenance Economic Performance

In the following the development of a first prototype of a new system, denoted EcoCon, for monitoring maintenance economic performance is described.

3.1 Requirements on a System for Monitoring Maintenance Economic Performance

There are mainly two reasons for measuring performance that are to follow up and improve performance. This is also reflected in the requirements of a system, EcoCon, for monitoring economic performance of maintenance. Consequently, the output from EcoCon should provide the user with information regarding the economic impact, including costs, losses and savings from maintenance at operative as well as strategic level. In addition to this the system should also provide more detailed information regarding causes of deviation in maintenance performance. Then the output will be useful for both control purposes and also enable continuous improvements. These abilities of EcoCon may be achieved by showing aggregated values and allow the user to bring up the data that can be found under these aggregated values, see for example [22, 35].

Furthermore, to provide the manager with an overview of maintenance performance it is necessary to translate technical maintenance performance into economic terms. This makes it possible to compare and prioritise where improvements are required and how much to invest in improvements. However, translating non-economic measures into economic is not without problems. The authors of [36] cite several sources that points out the difficulties in expressing operative performance into economic terms, e.g. difficulties in putting a correct economic value on downtime and those operational changes will not immediately occur as profits in the accounting system. Problems with assessing the consequences of downtime are also reported by [37]. Maintenance actions are performed at operative level and influence the technical performance of a production system or a machine, i.e. failures, short stoppages and quality rate. To show the economic impact from maintenance all costs, losses and savings due that can be related to technical performance of maintenance must be considered. This means that when implementing EcoCon in a certain situation it is necessary to find the links between operational maintenance performance and impact on tactical and strategic levels, see Fig. 1.

The basis for measuring performance is that it is improvements in non-economic measures that create economic improvement, see for example [38]. Consequently, in order to assess the economic impact from an investment or action in maintenance first, maintenance technical performance at operative level need to be determined, i.e. number and length stoppages, length of short stoppages and quality level. Next,

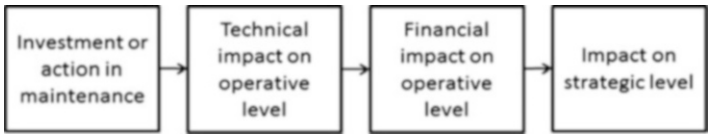


Fig. 1 Schematic view of maintenance impact

the economic impact, on operative as well as strategic level, from maintenance needs to be traced. For example, on operative level the economic impact from enhanced maintenance may be reduced stoppage time. If stoppage time leads to reduced sale, profit margin can be used to translate the operative technical performance into operative economic impact. Reduced stoppage time can also lead to reduced number of late deliveries and consequently reduced penalties for late deliveries, see [21, 39]. Because a maintenance measuring system also should provide data for continuous improvement activities, it is necessary that more detailed information is available regarding the measured performance, see for example [35, 40].

Furthermore, EcoCon should be flexible and allows defining what to maintenance related economic factors that should be monitored because maintenance impact may vary in different companies. Furthermore, a certain level of flexibility in EcoCon also allows for continuous improvement of the system itself.

3.2 System Development

A conceptual view showing the logic behind EcoCon is illustrated in Fig. 2.

1. All maintenance related economic factors are identified through identifying maintenance relate events for example stoppages, stoppage time, short stoppages and quality rate. Then, these economic factors that can be related to maintenance technical performance are identified.
2. Identifying the required data reflecting changes in these factors for monitoring.
3. Determined where to find data and how it should be gathered.
4. Data is collected into EcoCon. This can be done either automatically or manually.
5. Regularly assess the economic performance of the different maintenance related economic factors that were identified. These factors are denoted S_1 to S_{14}, see Fig. 2. The assessment is done by comparing performance in consecutive periods see Fig. 3. For example, savings or losses due to maintenance are assessed by comparing period maintenance performance in period t_1–t_2 with the performance of period t_1–t_0.

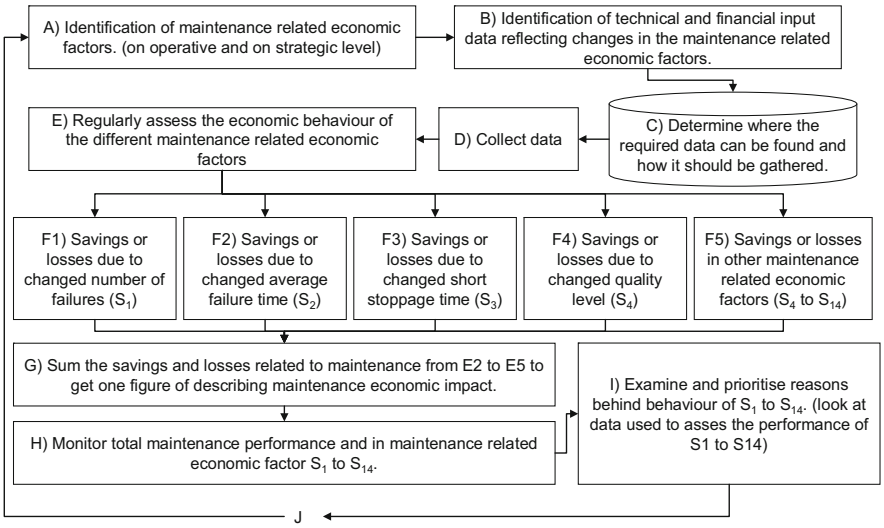


Fig. 2 Conceptual view of Eco-Con

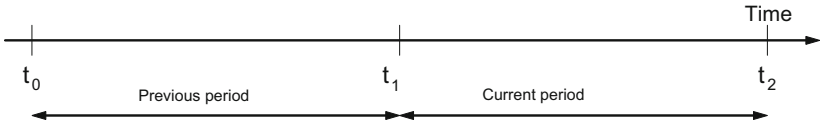


Fig. 3 Periods used when following up performance

1. (F1 to F5) The first two factors that are monitored and assessed are regarding losses due to failures. S_1: Savings or losses due to changed number of failures, and S_2: Savings or losses due to changed average stoppage length.

   Two factors are used to describe the losses; due to the number of failures and stoppage length.

   Savings or losses in S_1 and S_2 are calculated as:

$$S_1 = \left[(Y - y)^* L_1\right]^* PR^* PM \tag{1}$$

$$S_2 = \left[(L_1 - l_1)^* y\right]^* PR^* PM \tag{2}$$

where

Y: Number of failures during previous period,
y: Number of failures during current period,
L_1: Average failure time during previous period,
l_1: Average failure time during current period,

PR: production rate during current period, and
PM: profit margin during previous period.

S_3: Savings or losses due to short stoppages. Only one factor is used to describe these losses, because usually there are not much data recorded regarding type and length these stoppages. In many cases the impact of short stoppages are not realized, and they are sometimes referred to as chronic losses, and considered as normal [41]. S_3 is calculated as:

$$S_3 = \left[(B - b)^* L_2\right]^* PR^* PM \tag{3}$$

where

B: Number of short stoppages during previous period,
b: Number of short stoppages during current period, and
L_2: Average short stoppage length.

S_4: Savings or losses due to changed quality level. S_4 can be assessed by:

$$S_4 = (p - P)^* WH^* WD^* PM \tag{4}$$

where

p: Production of good quality during current period,
P: Production of high quality during previous period,
WH: work hours per day, and
WD: work days per year.

Maintenance economic impact in other identified maintenance related economic factors S_5 to S_{14}, is assess by subtracting the cost for the current period with the cost for the previous period. This method of assessing maintenance economic impact is developed and described in [16, 21, 22, 39].

2. Assess maintenance total economic impact by adding S_1 to S_{14}.
3. Monitor maintenance performance by following the development of maintenance total economic impact and the development in the maintenance related economic factors, S_1 to S_{14}.
4. If a deviation occurs in either maintenance total economic impact or in any of the factors S_1 to S_{14} a deeper analyze of S_1 to S_4 using more detailed information is required to find the causes of the deviations. For more detailed information, see for example [35, 40].
5. Because the dynamic nature of a production system the system should regularly be reviewed and updated to ensure that it is always reflecting the current situation.

3.3 EcoCon

To develop a prototype of EcoCon Microsoft Excel was used, see Fig. 4. In order to make it easy to understand we decided that it was important to show as much as possible of the data used to assess the losses and saving due to maintenance. This was an important issue because it makes it easier to understand for persons involved in testing EcoCon. For the prototype of EcoCon data was gathered from two sources: a system for registering disturbances in a production process, MUR developed by Adductor. From MUR EcoCon fetch data regarding failures and short stoppages. All other required data was entered manually in the prototype. Below is a description of the most important features of EcoCon as shown in Fig. 4.

1. **Setup:** The data regarding failures and short stoppages that come from MUR are mapped to S_1 Savings or losses due to changed number of failures, S_2 Savings or losses due to changed average stoppage length and S3 Savings or losses due to short stoppages. Because MUR is a measuring system and registers all stoppages and disturbances in a production process. It was also necessary to associate each disturbance type from MUR with maintenance or any other working area, e.g. disturbance type number 5, Repair, can be connected to Maintenance while disturbance type number 4 can be associated classified as Logistic and so on. Also during setup it is necessary to manually enter the economic consequences of downtime. During setup maintenance related economic factors S_5 to S_{14} is defined. Depending on the type of production these can vary and have to be decided for each implementation.

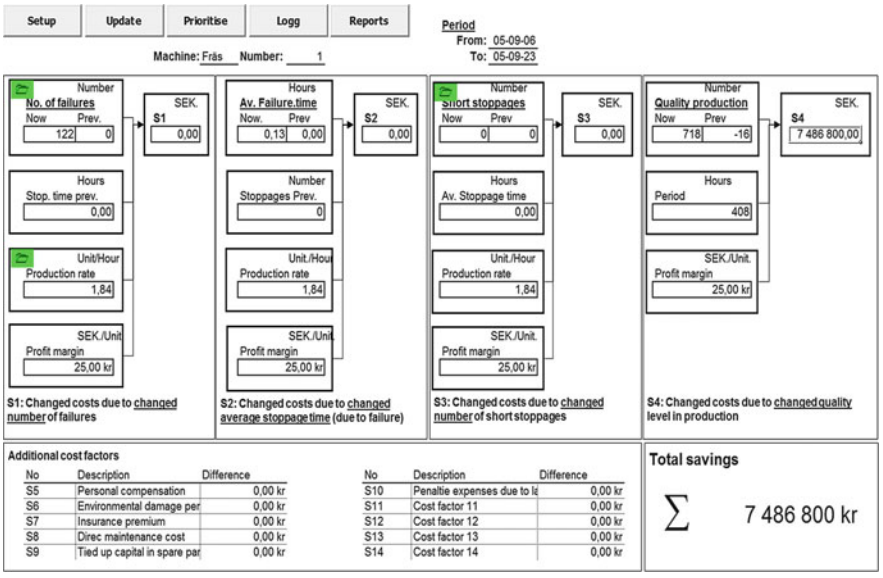


Fig. 4 EcoCon

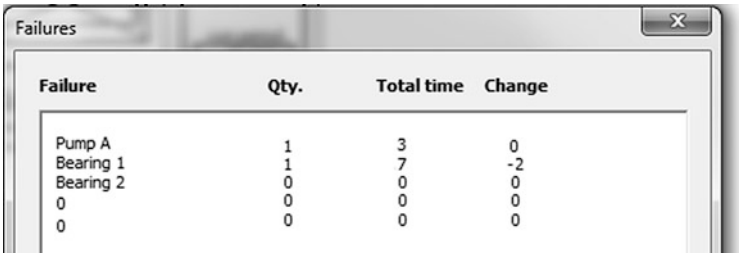


Fig. 5 Individual failure types included in S_1 and S_2

No	Area	Loss	Part	Total
1	Logistics	468 kr	2%	2%
2	Maintenance	293 kr	1%	3%
3	Other	3 kr	0%	3%
4	Organisation	0 kr	0%	3%

No	Cause of loss	Loss	Part	Total
1	1 Pump A	15 245 kr	63%	63%
2	5 Maintenance	7 559 kr	31%	94%
3	Failure 10	1 080 kr	4%	99%
4	4 Material	0 kr	0%	99%

Fig. 6 Prioritization of losses per working area and per failure type

2. **Update:** To monitoring maintenance impact Update is used. First the user have to enter start and end date for the current period, then data related to stoppages and short stoppages are fetched from MUR for the current period. Data related to quality level, S_4, and additional economic factors, S_5 to S_{14}, are also entered now. When the required data is entered EcoCon immediately shows the result in the excel sheet, see Fig. 4. The savings (profit) or losses gained because of improved maintenance compared with last review period in each of the monitored economic factors, S_1 to S_{14}, and total.
3. **Prioritize:** If significant deviation in any of the monitored maintenance related economic factors occurs it is possible to further analyze data to find the cause, see Fig. 5. If S_1 or S_2 deviates significantly it is possible to directly look at the impact from individual failure types. It is also possible to prioritize according to working area, i.e. to see if deviations can be related to maintenance or if there are other reasons such as logistics or organizational causes behind the deviation. This requires that the failure types are associated with working areas, see Fig. 6.

4 Test Using Real Data

EcoCon have been tested using data from the mail terminal in Alvesta Sweden. More than one million letters pass through and are sorted at the terminal each 24 hours. For this test historical data regarding stoppages and quality, covering a period for approximately 3 years, from one sorting machine which handles about

Table 1 Description of data

Historical stoppage data		Maintenance related economic factor in EcoCon
Category	Description	
Technical failures	Short stoppages of various types that usually are corrected by the operator or maintenance personnel	S3, Short stoppages
Jam	Letters got stuck in the machine, also mostly corrected by operators	S1 Number of failures, S2 Average failure time
Other	Longer stoppages which require more advanced repair by maintenance personnel	S1 Number of failures, S2 Average failure time
Number of letters entered		S4, Quality production
Number of letters sorted		S4, Quality production

100,000 letters each 24 hours was used. The data are stored in an Access data base. Also, a few other economic factors that were possible to relate to stoppages and quality were identified, i.e. extra costs for transportation when the sorting were not finished on time, spare parts and direct maintenance cost. However, the historical data did not allow us to include these factors in this test. Table 1 describe the data in the access database and how it is associated with the maintenance related economic factors in EcoCon. It could be argued that jam should not be associated with S_1 and S_2 because the stoppage times are short. However, it was decided to do that because this stoppage type and stoppage time was properly registered, while technical failures include several different types of short stoppages without proper registration of the causes. In this test only S_1 to S_4 was considered. All collected data are related to maintenance.

Three adaptations of EcoCon had to be made for this test:

1. Since the mail terminal have no influence over the amount of letters they have to sort each day and do not sell a product it was not possible to use profit margin to measure the economic consequences of lost production. Instead it was decided to use the cost of using extra resources, in this case the sorting machine and the operators. However, this is an area that needs to be further researched in order to get an accurate cost.
2. Because the number of letters varies from day to day, the savings or losses compared to previous period was shown both as normalized figures, i.e. savings or losses per two million sorted letters, and as the actual saving or loss.
3. Also, because the number of letters varies from day to day the formula used for assessing S_4 savings or losses due to quality level, see equation 4, had to be modified to compare based on a normalized number of letters entered.

Table 2 shows the result from using EcoCon on data covering a period from 2005-03-21 to 2008-05-28. Please note that the figures in the table are masked.

Table 2 Changed cost per 2 million sorted letters

Period		S_1 S_2	S_3		S_4
P1	2005-03-21–2005-12-29	Reference			
P2	2006-01-02–2006-12-19	94	571	−195	435
P3	2007-01-08–2007-12-27	101	−91	28	2446
P4	2008-01-02–2008-05-28	875	86	−361	−2896
Total change per maintenance related economic factor		1070	566	−528	−15
Total changed cost		1093			

From Table 2 it can be seen that savings have been made due to decreased number of failures, S_1, and also due to reduced average stoppage time, S_2. EcoCon can show that the savings made in S1 and S2 are mostly due to reduced numbers of letters getting stuck in the machine, i.e. the machine has become more reliable. On the other hand costs have increased due to increased short stoppages and reduced quality level.

The result was discussed with the maintenance manager at the mail terminal and the improvements made to S_1 and S_2 is the result of continuous improvement activities in maintenance. The increased losses shown in S_3 and the increased losses in S_4, between periods P3 and P4 may be due to introduction of new technology in the sorting machine. Even if this was an initial limited test and we only considered S_1 to S_4, EcoCon provided the maintenance manager with a new view of maintenance performance by showing financial losses and savings that can be associated with maintenance activities. The test of EcoCon will continue including additional economic factors that can be traced to maintenance performance of this machine.

5 Result and Conclusions

The main result presented in this paper is a system, EcoCon, for monitoring maintenance economic performance. EcoCon present maintenance total increased or decreased maintenance performance in economic terms and also the contribution from the underlying maintenance related economic factors are presented in economic terms. This allows both for monitoring maintenance technical and economic impact. Furthermore, EcoCon can provide more detailed information to use for analysing causes behind deviations in maintenance performance which support cost-effective continuous improvement of maintenance.

By expressing maintenance performance in economic terms, EcoCon makes it easier for maintenance managers to communicate with higher management. A test that was performed showed that EcoCon can provide useful financial information regarding maintenance performance. Also, during the test it was emphasized that determine the cost of down time can be a problem.

A system for measuring economic maintenance performance relies heavily on finding an adequate economic measure of downtime, which can be difficult. The

latter was revealed during the test of this system as well as previously being pointed out by several authors. Consequently this is an area that needs further research. Additionally the mechanism that makes savings that are gained on operational level e.g. due to reduced number of failures and short stoppages, turn up as profit in the accounting system needs to be further investigated in order to increase the confidence in this type of systems.

Acknowledgements The authors would like to acknowledge Adductor for letting us use MUR for this project, and also Postterminalen in Alvesta.

References

1. Holmberg K (2001) Competitive Reliability 1996–2000, Technology Programme Report 5/2001, Final Report Edited by Kenneth Holmberg, National Technology Agency, Helsinki
2. Luce S (1999) Choice criteria in conditional preventive maintenance: short paper. Mech Syst Signal Process 13(1):163–168
3. Vineyard M, Amoako-Gyampah K, Meredith J (2000) An evaluation of maintenance policies for flexible manufacturing systems: a case study. Int J Oper Prod Manag 20(4):299–313
4. Al-Najjar B, Alsyouf I (2003) Selecting the most efficient maintenance approach using fuzzy multiple criteria decision-making. Int J Prod Econ 84(1):85–100
5. Bevilacqua M, Braglia M (2000) The analytic hierarchy process applied to maintenance strategy selection. Reliab Eng Syst Saf 70(1):71–83
6. Mckone K, Weiss E (1998) TPM: planned and autonomous maintenance: bridging the gap between practice and research. Prod Oper Manag 7(4):335–351
7. Mitchell E, Robson A, Prabhu VB (2002) The impact of maintenance practices on operational and business performance. Manag Audit J 15(5):234–240
8. Riis J, Luxhoj J, Thorsteinsson U (1997) A situational maintenance model. Int J Qual Reliab Manag 14(4):349–366
9. Swanson L (2001) Linking maintenance strategies to performance. Int J Prod Econ 70 (3):237–245
10. Basim A-N, Imad A (2004) Enhancing a company's profitability and competitiveness using integrated vibration-based maintenance: a case study. Eur J Oper Res 157(3):643–657
11. Groote PD (1995) Maintenance performance analysis: a practical approach. J Qual Maint Eng 1(2):4–24
12. Markeset T, Kumar U (2001) R&M and risk-analysis tools in product design, to reduce life-cycle cost and improve attractiveness. In: Proceedings of the annual reliability and maintainability symposium, 2001
13. Ollila A, Malmipuro M (1999) Maintenance has a role in quality. TQM Mag 11(1):17–21
14. Nikolopoulos K, Metaxiotis K, Lekatis N, Assimakopoulos V (2003) Integrating industrial maintenance strategy into ERP. Ind Manag Data Syst 103(3):184–192
15. Alsyouf I (2004) Cost effective maintenance for competitive advantages. Doctoral Thesis, Växjö University, Växjö, Sweden
16. Holmberg K, Jantunen E, Adgar A, Mascolo J, Arnaiz A, Mekid S (2009) E-Maintenance. Springer, London
17. Pujadas W, Chen FF (1996) A reliability centered maintenance strategy from a discrete manufacturing facility. Comput Ind Eng 31(1/2):241–244
18. Sherwin D (2000) A review of overall models for maintenance management. J Qual Maint Eng 6(3):138–164

19. Waeyenbergh G, Pintelon L (2002) A framework for maintenance concept development. Int J Prod Econ 77:299–313
20. Cavalier M, Knapp GM (1996) Reducing preventive maintenance cost error caused by uncertainty. J Qual Maint Eng 2(3):21–35
21. Al-Najjar B (2007) The lack of maintenance and not maintenance which costs: a model to describe and quantify the impact of vibration-based maintenance on company's business. Int J Prod Econ 107(1):260–274
22. Al-Najjar B (2009) A computerised model for assessing the return on investment in maintenance; Following up maintenance contribution in company profit. In: Proceedings of the fourth World Congress on Engineering Asset Management (WCEAM) 2009, Greece, Athens, 08 Sep 2009. Springer, London, pp 137–145
23. Al-Najjar B, Kans, M, Ingwald, A, Samadi, R (2003) Decision support system for monitoring and assessing maintenance financial impact on company profitability. In: Proceedings of business excellence I, performance measures, benchmarking and best practices in new economy, Portugal
24. Tsang AHC (2002) Strategic dimensions of maintenance management. J Qual Maint Eng 8 (1):7–39
25. EN 13306:2001 (2001) Maintenance terminology. Comite Europeen de Normalisation
26. Al-Najjar B (1997) Condition-based maintenance: selection and improvement of a cost-effective vibration-based policy for rolling element bearings. Doctoral thesis, Lund University, Lund Sweden
27. Kelly A (1997) Maintenance strategy. Butterworth Heinemann, Oxford
28. Pinjala SK, Pintelon L, Vereecke A (2006) An empirical investigation on the relationship between business and maintenance strategies. Int J Prod Econ 104(1):214–229
29. Alysouf I (2007) The role of maintenance in improving companies' productivity and profitability. Int J Prod Econ 105(1):70–78
30. Kutucuoglu KY, Hamali J, Irani Z, Sharp JM (2001) A framework for managing maintenance using performance measurement systems. Int J Oper Prod Manag 21(1):173–195
31. Fabricky WJ, Blanchard BS (1991) Life cycle cost and economic analysis. Prentice-Hall, Englewood Cliffs
32. Pintelon L (1997) Maintenance performance reporting systems: some experiences. J Qual Maint Eng 3(1):4–15
33. Stenström C, Parida A, Kumar U, Galar D (2013) Performance indicators and terminology for value driven maintenance. J Qual Maint Eng 19(3):222–232
34. Kumar U, Galar D, Parida A, Stenström C, Berges L (2013) Maintenance performance maetrics: a state-of-the-art review. J Qual Maint Eng 19(3):233–277
35. Pintelon L, Van Puyvelde F (1997) Maintenance performance reporting systems: some experiences. J Qual Maint Eng 3(1):4–15
36. Caplice C, Sheffi Y (1995) A review and evaluation of logistics performance measurement systems. Int J Logist Manag 6(1):61–74
37. Moore WJ, Starr AG (2006) An intelligent maintenance system for continuous cost-based prioritisation of maintenance activities. Comput Ind 57(6):595–606
38. Kaplan RS, Norton DP (2007) Using the balanced scorecard as a strategic management system. Commun Manag 85(7):150–161
39. Ingwald A, Al-Najjar B (2006) System for monitoring machines financial and technical effectiveness. In: Proceedings for the 19th international congress of Condition Monitoring and Diagnostic Engineering Management (COMADEM 2006), pp 219–229
40. Martorell S, Sanchez A, Muñoz A, Pitarch JL, Serradell V, Roldan J (1999) The use of maintenance indicators to evaluate the effects of maintenance programs on NPP performance and safety. Reliab Eng Syst Saf 65(2):85–94
41. Ljungberg Ö (1998) Measurement of overall equipment effectiveness as a basis for TPM activities. Int J Oper Prod Manag 18(5):495–507

An Adaptive Power-Law Degradation Model for Modelling Wear Processes

R. Jiang

Abstract The power-law model is often used as the mean value function of the non-homogeneous gamma process for describing nonlinear degradation behaviour of a product population. Since there is unit-to-unit variability among individual degradations, it would be inaccurate to use the power-law process of the population to predict the time to a degradation limit for a given individual. To address this issue, we present an adaptive power-law degradation model in this paper. The scale parameter of the power-law model is dynamically updated on the basis of the recent degradation observation. A method that is somehow like the exponential smoothing approach is adopted to modify the scale parameter. A real-world example that deals with a non-homogeneous wear process is included to illustrate the appropriateness of the proposed model. A graphical method is also presented to intuitively display the relative health level of a condition monitored item.

Keywords Non-homogeneous gamma process • Power-law model • Adaptive process • Wear process • Relative health level

1 Introduction

Poisson processes with non-constant intensity functions (i.e., non-homogeneous Poisson processes, NHPP) have been widely used for modelling degradation processes (e.g., wear process, see [1]). The system is regarded as failed when the degradation process reaches a critical threshold (e.g., wear limit). A well-known NHPP is the power-law process. In the context of modelling degradation processes, the NHPP with independent increment is widely used [2]. The independent increment is simple but restrictive since the rate of degradation is generally dependent on the current state (e.g., see [3–5]). When the deterioration increment is state-dependent, an adaptive approach is needed to update the parameters of the degradation process in order to improve the accuracy of residual life prediction (e.g., see

R. Jiang (✉)
Changsha University of Science and Technology, Changsha, China
e-mail: jiang@csust.edu.cn

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_9

[6–8]). In the literature, adaptive approaches are also applied to update maintenance decision rules (e.g., see [9–12]).

In this paper, we consider the power law wear process and assume that the shape parameter of the power-law model is a constant but the scale parameter is dynamically updated based on the recent wear observation. A method, which is somehow like the exponential smoothing approach, is developed to modify the scale parameter. A graphical method is presented to intuitively display the evolution trend and relative health level of a wear process. A real-world example is used to illustrate the proposed approach. The applications of the resulting model are also discussed. As such, the main contributions of this paper are: (a) a new method to dynamically update the scale parameter of the power-law model and (b) a graphical method to show the relative health level of an item.

The paper is organized as follows. The proposed adaptive model is presented in Sect. 2. Section 3 deals with the model parameter estimation. A numerical example is presented in Sect. 4. The applications of the resulting model are discussed in Sect. 5. The paper is concluded in Sect. 6.

2 Proposed Model

2.1 Mean Value Function

A wear process generally undertakes three phases: the rate of wear first increases quickly, then roughly maintains in a constant and finally increases quickly again. This implies that the mean wear curve is inverse S-shaped with an inflection point. That is, the rate of wear is bathtub-shaped. Jiang [13] presents a three-parameter model that is suitable for representing the whole wear process. Its mean value function is given by

$$w(t) = (t/\eta)^{\beta} e^{t/\gamma}, \quad 0 < \beta < 1, \quad \eta > 0, \quad \gamma > 0. \tag{1}$$

Another bathtub-shaped intensity model is the superposed power-law model [14]. Its mean value function is given by

$$w(t) = (t/\eta_1)^{\beta_1} + (t/\eta_2)^{\beta_2}, \quad 0 < \beta_1 < 1 < \beta_2. \tag{2}$$

The model has four parameters.

When the wear limit is defined in a relatively small wear value, the observation period may be in the time interval before the inflection point. In this case, the power-law process and log-linear process can be appropriate for fitting the wear data [1]. In this paper, we focus on the power-law model.

Let $W(t)$ denote a stochastic wear process. Observed wear processes are given by:

$$\left(t_{ij}, w_{ij}; i = 1, 2, \ldots, n, j = 0, \ 1, \ 2, \ \ldots, n_i\right) \tag{3}$$

where $t_{i0} = 0$, $w_{i0} = 0$, and w_{ij} is the cumulative wear amount measured at time t_{ij} for the ith wear process. The mean value function of $W(t)$ is given by the power-law model:

$$w(t) = (t/\eta)^{\beta} \tag{4}$$

where β is the shape parameter and assumed to be unchanged; η is the scale parameter and can vary with time. Specifically, for time interval (t_{i0}, t_{i1}), the scale parameter is η_0 for any wear process. At time t_{ij}, the j-th observation is obtained. Based on this observation and Eq. (3), the scale parameter can be estimated as

$$s_{ij} = t_{ij}/w_{ij}^{1/\beta}. \tag{5}$$

The scale parameter is updated by

$$\eta_{ij} = \alpha\eta_0 + (1 - \alpha)s_{ij} \tag{6}$$

where α is a parameter to be simultaneously estimated with the other model parameters. Equation (6) will be used in time interval $t \in [t_j, t_{j+1})$. The idea of the adaptive power-law process is graphically displayed in Fig. 1, where the dotted points are the wear observations and the dotted lines are the mean value functions of dynamically adjusted power-law model.

It is noted that Eq. (6) is similar to the exponential smoothing [15] in form. The main differences between them are: (a) η_0 is a constant rather than the earlier forecasting value, and (b) α is no necessarily confined in interval (0, 1). We call α the adjustment factor and further examine its properties as follows.

From Equation (6) we have

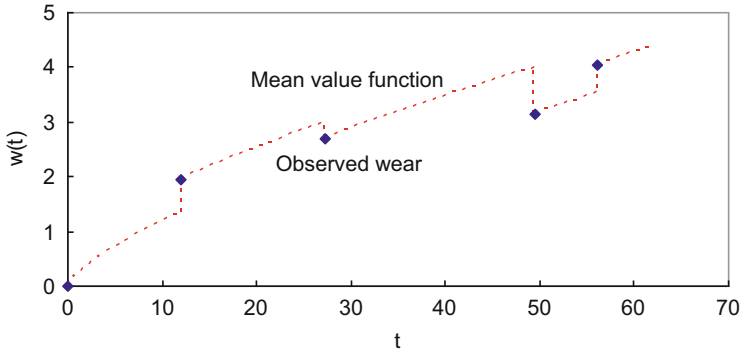


Fig. 1 An illustration of adaptive power-law process

$$\eta_{ij} - s_{ij} = \alpha(\eta_0 - s_{ij}), \quad \eta_{ij} - \eta_0 = (1-\alpha)(s_{ij} - \eta_0) \tag{7}$$

We consider three cases: $\alpha < 0$, $\alpha \in (0,\ 1)$ and $\alpha > 1$ as follows:

1. If $\alpha < 0$, we have $\eta_{ij} < s_{ij} < \eta_0$ when $s_{ij} < \eta_0$; or $\eta_{ij} > s_{ij} > \eta_0$ when $s_{ij} > \eta_0$. This may result in an over-adjustment for the scale parameter and hence is not allowed.
2. If $\alpha \in (0,\ 1)$, $\min(s_{ij}, \eta_0) < \eta_{ij} < \max(s_{ij}, \eta_0)$. This implies that the adjustment given by Eq. (6) is proper and hence is desired. Specially, when $\alpha = 1$, $\eta_{ij} = \eta_0$. In this case, the increment is independent of the current state and hence the model reduces into the independent increment model.
3. If $\alpha > 1$, we have $\eta_{ij} > \eta_0$ when $s_{ij} < \eta_0$, or $\eta_{ij} < \eta_0$ when $s_{ij} > \eta_0$. This implies that Eq. (6) adjusts the scale parameter in the opposite direction of s_{ij} relative to η_0. Under some situations, such an adjustment is reasonable and hence is allowed.

2.2 *Non-homogeneous Gamma Increment Process*

The power-law mean value function represents non-linearity of the wear process. For a given value of t, we assume that $W(t)$ follows the gamma distribution with the scale parameter v and shape parameter $u(t)=w(t)/v$. The mean and variance of $W(t)$ are given respectively by [16]

$$E[W(t)] = vu(t) = w(t), \quad \text{var}[W(t)] = v^2 u(t) = vw(t). \tag{8}$$

Equation (8) indicates that the scale parameter reflects the variability of the wear process.

The proposed non-homogeneous gamma process model has four parameters: β, η_0, α and v. Since the scale parameter is a function of t_{ij} and w_{ij}, $w(t)$ is state-dependent and hence the increment is no longer independent.

3 Maximum Likelihood Estimation

We use the maximum likelihood method (MLM) to estimate the model parameters. Consider the i-th wear path given by Eq. (3). The wear increment $\Delta W_{ij} = W_{ij} - W_{i,j-1}$ $(j = 1,\ 2,\ \ldots)$ is a random variable with mean given by

$$\mu_{ij} = E(\Delta W_{ij}) = \left(\frac{t_{ij}}{\eta_{i,j-1}}\right)^{\beta} - \left(\frac{t_{i,j-1}}{\eta_{i,j-1}}\right)^{\beta}, \quad \eta_{i,0} = \eta_0. \tag{9}$$

Under the assumption that the wear increment ΔW_{ij} follows the gamma distribution with shape parameter $u_{ij} = \mu_{ij}/v$ and the scale parameter v, the likelihood function for path i is given by

$$L_i = \prod_{j=1}^{n_i} g\left(\Delta w_{ij}; u_{ij}, v\right), \quad \Delta w_{ij} = w_{ij} - w_{i,j-1} \tag{10}$$

where $g(x; u, v)$ is the gamma cdf evaluated at x. The overall log-likelihood function is given by

$$\ln(L) = \sum_{i=1}^{n} \ln(L_i). \tag{11}$$

The parameter set $(\eta_0, \beta, \alpha, v)$ can be estimated by directly maximizing $\ln(L)$ using Solver of Microsoft Excel. The interval estimates of the parameters can be obtained using standard statistical methods.

4 Illustration

4.1 Wear Data and Earlier Works

The data shown in Table 1 come from Bocchetti et al. [1], and deal with accumulated wear amounts of 33 cylinder liners of 8-cylinder SULZER RTA 58 engines. In the table, t_{ij} (in 1000 h) is the jth inspection time or operating age ($j = 1, 2, \ldots, n_i$) for the ith liner ($i = 1, 2, \ldots, 33$), and w_{ij} is the corresponding accumulated wear amount in mm. There is no wear record for the 32nd liner. The maximum admissible wear is 4 mm. The 12th, 25th and 28th liners were replaced at the last inspection due to a large accumulated wear.

The data have been modelled using different models. For example, Bocchetti et al. [1] fit the data to stochastic process model with a log-linear mean function and a power-law mean function, respectively; Giorgio et al. [3] fit the data to three different stochastic models, one of which is state-dependent homogeneous Markov chain; Giorgio et al. [4] fit the data to a parametric Markov chain model, in which the transition probabilities between states depend on both the current age and wear level; Guida and Pulcini [5] model the data using a non-stationary inverse Gamma process, which is state-dependent.

Table 1 Wear data of cylinder liners

i	$t_{ij}, j = 1, 2, \ldots$	$w_{ij}, j = 1, 2, \ldots$
1	11.3, 14.68, 31.27	0.90, 1.30, 2.85
2	11.36, 17.2	0.80, 1.35
3	11.3, 21.97	1.50, 2.00
4	12.3, 16.3	1.00, 1.35
5	14.81, 18.7, 28	1.90, 2.25, 2.75
6	9.7, 19.71, 30.45	1.10, 2.60, 3.00
7	10, 30.45, 37.31	1.20, 2.75, 3.05
8	6.86, 17.2, 24.71	0.50, 1.45, 2.15
9	2.04, 12.58, 16.62	0.40, 2.00, 2.35
10	7.54, 8.84, 9.77, 16.3	0.50, 1.10, 1.15, 2.10
11	8.51, 14.93, 21.56	0.80, 1.45, 1.90
12	18.32, 25.31, 37.31, 45	2.20, 3.00, 3.70, 3.95
13	10, 16.62, 30	2.10, 2.75, 3.60
14	9.35, 15.97	0.85, 1.20
15	13.2	2.0
16	7.7	1.05
17	7.7	1.60
18	8.25	0.90
19	3.9	1.15
20	7.7	1.20
21	9.096	0.50
22	19.8	1.60
23	10.45	0.40
24	12.1	1.00
25	12, 27.3, 49.5, 56.12	1.95, 2.70, 3.15, 4.05
26	8.8	1.40
27	2.2	0.40
28	33, 38.5, 55.46	2.90, 3.25, 4.10
29	8.8, 27.5	0.50, 2.15
30	8.25	0.70
31	18.755	1.15
33	8.49	0.95

4.2 Parameter Estimation

The maximum likelihood estimates (MLEs) of the model parameters are shown in the second row of Table 2; and the MLEs for case $\alpha = 1$ are shown in the last row. From the table we have the following observations:

- The smoothing parameter $\alpha > 1$, implying that the scale parameter is adjusted towards the opposite direction of s_{ij} relative to η_0. As mentioned earlier, this is allowed.

Table 2 MLEs of the model parameters

Case	η_0	β	α	v	$\ln(L)$
$\alpha \neq 1$	8.1177	0.7321	1.2168	0.1511	−22.733
$\alpha = 1$	8.1644	0.7428	1	0.1557	−23.876

- In terms of AIC, the state-dependent increment model outperforms the state-independent increment since the difference between the log-likelihood values is larger than 1. As such, the proposed model is validated.

5 Applications

In this section, we present two applications of the resulting model. One deals with prediction of time to the wear limit, and the other deals with evaluation of relative health level.

5.1 Prediction of Time to the Wear Limit

The distribution function of time to the wear limit ($L = 4$ mm) for the population is given by

$$F(t) = 1 - G(L; w(t)/v, v) \tag{12}$$

where $G(x; u, v)$ is the gamma cdf evaluated at x. For the special case $\alpha = 1$, we have

$$w(t) = (t/\eta_0)^{\beta}. \tag{13}$$

Using Eq. (13) into Eq. (12), the cdf of time to the wear limit can be directly evaluated, and the pdf can be numerically evaluated. For small value of Δt, the pdf is approximated by

$$f(t + 0.5\Delta t) \approx [F(t + \Delta t) - F(t)]/\Delta t. \tag{14}$$

For the current example, the pdf evaluated by Eq. (14) is shown in Fig. 2 (the solid line), and can be approximated by the gamma distribution with the parameters shown in the second row of Table 3. Its plot is also shown in Fig. 2 (the dotted line).

For the general case $\alpha \neq 1$, we first infer the time to the wear limit for each process using the last observation before the process reaches the wear limit. The inference is based on the mean value function (this can lead to an underestimate of dispersion). This yields a sample of time to the wear limit (which is sometimes

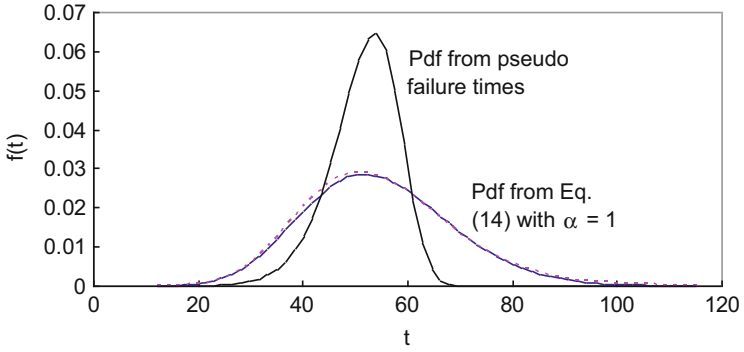


Fig. 2 Distribution of time to the wear limit for the population

Table 3 MLEs of time to wear limit

	u, μ, β	v, σ, η	ln(L)	B_X^*
Gamma, $\alpha = 1$	14.85	3.69		40.95
Gamma, $\alpha \neq 1$	46.36	1.11	−109.894	41.14
Normal, $\alpha \neq 1$	51.45	6.96	−107.510	41.81
Weibull, $\alpha \neq 1$	9.48	54.23	−104.994	42.78

called pseudo-failure time). Then, we fit the sample to a standard distribution using the MLM. The results are shown in the last three rows of Table 3. As seen, the Weibull distribution provides the best fit in terms of log-likelihood. The fitted Weibull distribution is also shown in Fig. 2.

A possible application of the population distribution is to determine a preventive replacement age for the population. This needs a decision model such as the cost model (when the cost parameters for preventive and failure replacements are known) or the tradeoff BX life B_X^* (when the replacement costs are unknown, see [17]). For the current example, the values of B_X^* from different models are shown in the last column of Table 3. As seen, the values of B_X^* are very close to each other though they come from different models.

We now apply the adaptive model to make the maintenance decision for Liners 12, 25 and 28. The time to the wear limit can be inferred using linear interpolation or extrapolation based on the last two records. The results are shown in the second row of Table 4. Actual maintenance types are shown in the third row.

The decision times t_d are shown in the fourth row. The values of scale parameter are shown in the fifth row, and the times for the mean wear curves to reach the wear limit are shown in the sixth row. As seen, the predicted failure times are close to the actual failure times for Liners 12 and 28 but larger than the actual failure times for Liner 25.

The seventh row shows the tradeoff B_X lifetimes of individual liners. As seen from the table, these liners would have been preventively replaced at B_X^* if the liners were preventively replaced at age B_X^*. Of course, some useful lifetimes would be lost.

Table 4 Maintenance decisions for Liners 12, 25 and 28

	Liner 12	Liner 25	Liner 28
Linearly inferred T_f	46.54	55.75	53.46
Actual maintenance type	PM	CM	CM
Decision time t_d	37.31	49.50	38.50
η_j	8.52	7.64	8.21
$E(T_f)$	42.59	64.69	51.76
B_X^*	37.62	55.52	43.75
Maintenance type based on B_X^*	PM	PM	PM
$B_X^* - t_d$	0.31	6.02	5.25
Maintenance decision	PM	Inspection at 54.50	Inspection at 43.50

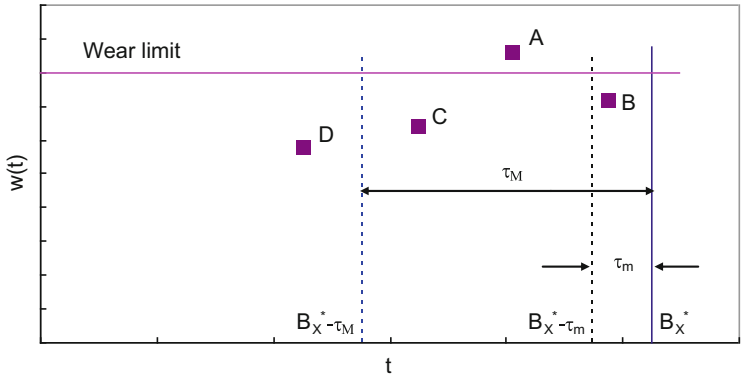


Fig. 3 Maintenance decision rules

The actual maintenance decision is based on decision rules. For the current example, the decision rules can be defined as below:

- Rule 1: a corrective replacement is immediately performed if the observed wear is larger than or equal to the wear limit (see Point A in Fig. 3).
- Rule 2: a preventive replacement (PR) is immediately performed if the observed wear is smaller than the wear limit and the time to the previously predicted PR time (say B_X^*) is smaller than a certain critical value τ_m (e.g., 1000 h) (see Point B in Fig. 3).
- Rule 3: a PR can be performed at B_X^* or opportunistically performed between $B_X^* - \tau_M$ and $B_X^* - \tau_m$ if the observed wear is smaller than the wear limit and the time to the PR time is larger than τ_m and smaller than the inspection interval τ_M (e.g., 5000 h) (see Point C in Fig. 3).
- Rule 4: the next inspection is scheduled if the observed wear is smaller than the wear limit and the time to the PR time is larger than τ_M (see Point D in Fig. 3).

Both τ_m and τ_M can be optimized [13].

The maintenance decisions made based on the above rules are shown in the last row of Table 4. As seen, the PR will be performed at $t = 37.31$ for Liner 12; and an inspection will be scheduled at $t = 54.50$ for Liner 25. Similarly, an inspection will be scheduled at $t = 43.50$ for Liner 28. As a result, the maintenance decision based on the proposed adaptive model generally can lead to a PR action.

In the wear monitoring context, there are other decision problems, such as determining a PM threshold, determining an opportunistic maintenance threshold (i.e., parameter τ_m), and dynamically determining the time of the next inspection (i.e., varying inspection interval, see [13, 18]).

5.2 Relative Health Level

The power-law model given by Eq. (4) can be linearized under the following transformations:

$$x = \ln(t), y = \ln(w). \tag{15}$$

The linearized power-law model can be written as below:

$$y(x) = A + \beta x, \quad A = -\beta \ln(\eta). \tag{16}$$

Since η varies with time and can be different for different wear processes, parameter A is a random variable.

Applying the transformations given by Eq. (15) to the data in Table 1, the observed data can be transformed as

$$x_{ij} = \ln\left(t_{ij}\right), \quad y_{ij} = \ln\left(w_{ij}\right). \tag{17}$$

Using Eq. (17), the data and resulting model can be visualized. Figure 4 shows the plot of the transformed data. As seen, the relationship between x and y is roughly linear, implying that the power-law model can be appropriate for fitting the data.

To evaluate the relative health level of a specific liner, we may build equal-health lines on x-y plane [19]. An h-equal-health line is given by

$$y_h(x) = a_h + \beta x. \tag{18}$$

The straight line $y_h(x)$ divides all the m data points into two parts: those data points that are above the line and the other data points that are on or below the line. Let k [$m - k$] denote the number of data points that are above [on or below] the line. The relative health level is defined as $h = (k - 0.3)/(m + 0.4)$. If an observation point falls on an h-equal-health line, its relative health level is h. The larger h, the healthier the process.

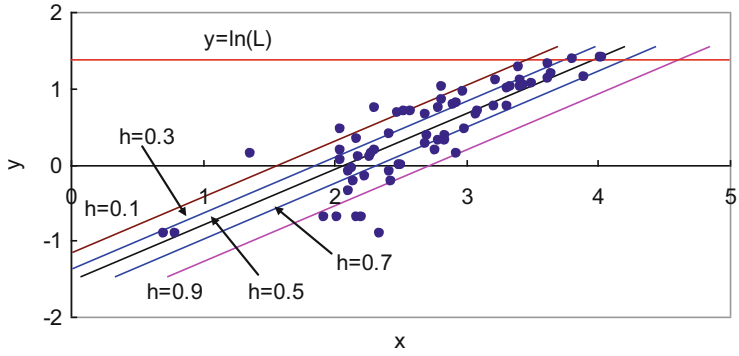


Fig. 4 Plot of transformed data

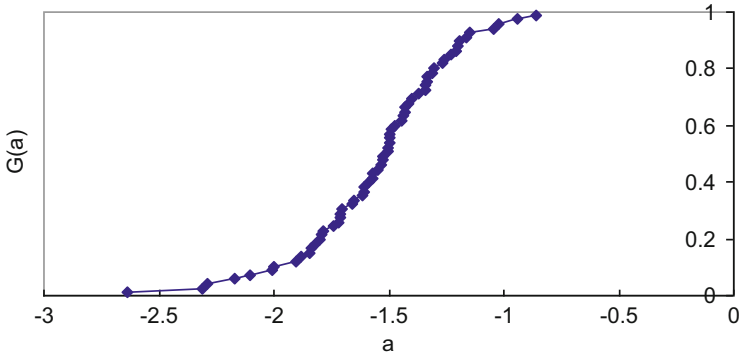


Fig. 5 Empirical distribution of A

From the data in Table 1, we can obtain a sample of A, given by

$$a_{ij} = \ln\left(w_{ij}\right) - \beta \ln\left(t_{ij}\right). \tag{19}$$

Figure 5 shows the empirical distribution $G(a)$ of A, where a is a realization of A. The random variable $A+\gamma$ can be well approximated by the Weibull distribution with parameters $(\beta,\ \eta,\ \gamma) = (8.39, 2.48, 3.90)$.

For a given h, we can obtain the value of a_h from the distribution of A. When a_h is known, we obtain the h-equal-health line. Figure 4 shows the h-equal-health lines for $h = 0.1(0.2)0.9$. The current relative health level of a wear process can be intuitively evaluated by drawing the observation point $(x_o,\ y_o)$ on the x-y plane and comparing this point with the equal-health lines.

The relative health level of a wear process can vary with time. For the point $(t_{ij},\ w_{ij})$, we can calculate $(x_{ij},\ y_{ij})$, from which we can calculate h_{ij}. The relative health level of process i can be evaluated as

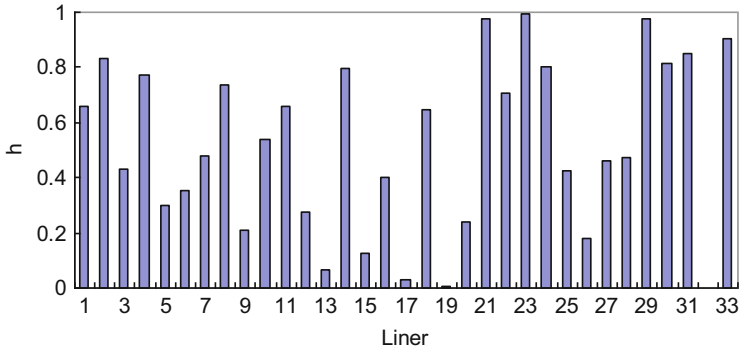


Fig. 6 Relative health levels of wear processes of all liners

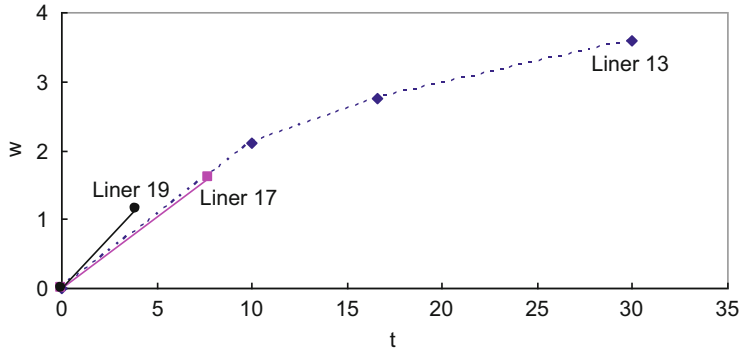


Fig. 7 Wear processes of Liners 13, 17 and 19

$$h_i = \frac{1}{n_i} \sum_{j=1}^{n_i} h_{ij}. \tag{20}$$

The stability of the process can be represented by

$$s_i = \left[\frac{1}{n_i - 1} \sum_{j=1}^{n_i} \left(h_{ij} - h_i \right)^2 \right]^{1/2}. \tag{21}$$

Figure 6 displays the relative health levels of wear processes of all the liners. As seen from the figure, the three worst processes are from Liners 13, 17 and 19. Their wear processes are shown in Fig. 7. As seen, they have similar wear trends and high initial wear rates. Therefore, the special attention should be paid on them.

6 Conclusions

In this paper, we have proposed an adaptive power-law process to model a set of wear processes based on the assumption the wear increment is state-dependent. The proposed model is expected to provide better prediction accuracy for the time to the wear limit. This has been illustrated through a real-world example.

A graphical representation has been developed to intuitively evaluate the relative health level of a wear process. More attention should be paid on those liners with low relative health level.

The proposed model and relative health level evaluation method can be applied in similar situations. A topic for future research is to develop the model to optimize a preventive replacement threshold, opportunistic replacement window and the time of the next inspection based on the resulting degradation model.

Acknowledgments The research was supported by the National Natural Science Foundation of China (No. 71371035).

References

1. Bocchetti D, Giorgio M, Guida M, Pulcini G (2009) A competing risk model for the reliability of cylinder liners in marine diesel engines. Reliab Eng Syst Saf 94:1299–1307
2. van Noortwijk JM (2009) A survey of the application of gamma processes in maintenance. Reliab Eng Syst Saf 94:2–21
3. Giorgio M, Guida M, Pulcini G (2009) Stochastic processes for modeling the wear of marine engine cylinder liners. In: Erto P (ed) Statistics for innovation. Springer, Berlin, pp 213–230
4. Giorgio M, Guida M, Pulcini G (2010) A parametric Markov chain to model age- and state-dependent wear processes. In: Mantovan P, Secchi P (eds) Complex data modelling and computationally intensive statistical methods. Springer, New York, pp 85–97
5. Guida M, Pulcini G (2013) The inverse Gamma process: a family of continuous stochastic models for describing state-dependent deterioration phenomena. Reliab Eng Syst Saf 120: 72–79
6. Laurenciu NC, Cotofana SD (2013) A nonlinear degradation path dependent end-of-life estimation framework from noisy observations. Microelectron Reliab 53(9-11):1213–1217
7. Si XS (2015) An adaptive prognostic approach via nonlinear degradation modeling: application to battery data. IEEE Trans Ind Electron 62(8):5082–5096
8. Xu W, Wang W (2012) An adaptive gamma process based model for residual useful life prediction. In: 2012 Prognostics and system health management conference (PHM-2012 Beijing), MU3019
9. Do P, Voisin A, Levrat E, Iung B (2015) A proactive condition-based maintenance strategy with both perfect and imperfect maintenance actions. Reliab Eng Syst Saf 133:22–32
10. Fouladirad M, Grall A (2011) Condition-based maintenance for a system subject to a non-homogeneous wear process with a wear rate transition. Reliab Eng Syst Saf 96:611–618
11. Li H, Deloux E, Dieulle L (2016) A condition-based maintenance policy for multi-component systems with Lévy copulas dependence. Reliab Eng Syst Saf 149:44–55
12. Zhu W, Fouladirad M, Bérenguer C (2015) Condition-based maintenance policies for a combined wear and shock deterioration model with covariates. Comput Ind Eng 85:268–283

13. Jiang R (2010) Optimization of alarm threshold and sequential inspection scheme. Reliab Eng Syst Saf 95:208–215
14. Guida M, Pulcini G (2009) Reliability analysis of mechanical systems with bounded and bathtub shaped intensity function. IEEE Trans Reliab 58(3):432–443
15. Gardner ES (2006) Exponential smoothing: the state of the art – part II. Int J Forecast 22(4): 637–666
16. Jiang R (2015) Introduction to quality and reliability engineering. Springer, Berlin
17. Jiang R (2013) A tradeoff Bx life and its applications. Reliab Eng Syst Saf 113:1–6
18. Zhang X, Zeng J (2015) A general modeling method for opportunistic maintenance modeling of multi-unit systems. Reliab Eng Syst Saf 140:176–190
19. Jiang R, Jardine AKS (2008) Health state evaluation of an item: a general framework and graphical representation. Reliab Eng Syst Saf 93(1):89–99

A Comparison Study on Intelligent Fault Diagnostics for Condition Based Maintenance of High-Pressure LNG Pump

Hack-Eun Kim and Tae-Hyun Jeon

Abstract High Pressure Liquefied Natural Gas Pump (HP-LNG Pump) is important equipment in the LNG receiving terminal process, and plays a key role in determining the total supply capacity of natural gas in LNG terminal. Therefore, the condition monitoring, fault diagnostics and prognostics technologies are applied to implement the CBM (Condition Based Maintenance) strategy for HP-LNG pumps, which can support for appropriate maintenance decision. Currently, a number of valuable diagnostic models and methods have been proposed in machine diagnostics. However, most intelligent diagnostic models have generally validated using an experimental data, and still not focus on industrial real data because of the complex nature in real industry. In this paper, intelligent fault diagnostic performances using three conventional classification algorithms such as SVM (Support Vector Machines), k-NN (k-Nearest Neighbor) and Fuzzy-ARTMAP have been evaluated for the CBM decision of HP-LNG pump. Comparative results indicate that the fault classification using SVM provides more accurate performance compared to other classification methods for HP-LNG pump.

Keywords Intelligent fault diagnostics • Condition based maintenance • High-pressure LNG pump

1 Introduction

The LNG receiving terminal receives Liquefied Natural Gas (LNG) from LNG carrier ships, stores the liquid in special storage tanks, vaporizes the LNG and then delivers the natural gas through distribution pipelines. The receiving terminal is designed to deliver a specified gas rate into a distribution pipeline and to maintain a reserve capacity of LNG. LNG takes up six hundredths of the volume of natural gas at/or below the boiling temperature (−162 °C), which is used for storage and easy transportation [1].

Figure 1 shows key facilities and process overview in an LNG receiving terminal. As shown in Fig. 1, the unloaded LNG from vessels is transported to ground storage

H.-E. Kim (✉) • T.-H. Jeon
Korea Gas Technology Corporation, Daejeon, Korea
e-mail: hackeunkim@gmail.com

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_10

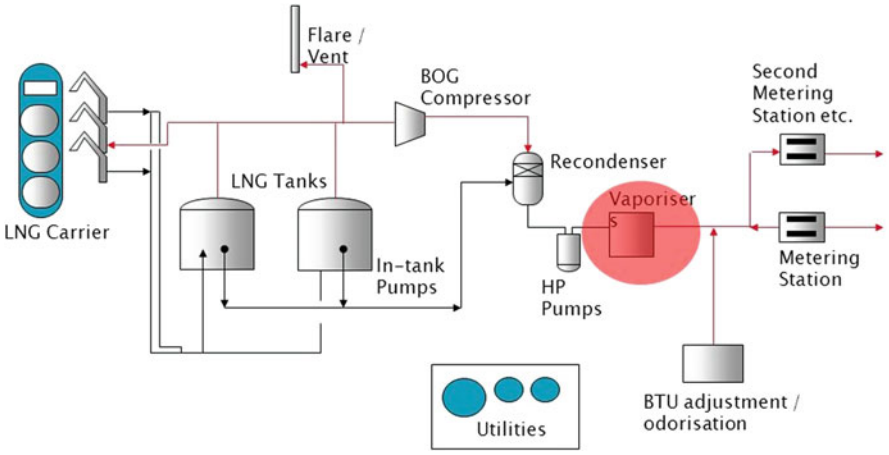


Fig. 1 Process overview in LNG receiving terminal

tanks via pipeline using cargo pumps on the LNG carrier vessel. In an LNG receiving terminal, primary cryogenic pumps that are installed in the storage tanks which supply the LNG to HP-LNG pumps with pressure around 8 bar. The HP-LNG pumps boost the LNG pressure to around 80 bar for evaporation and delivery of the highly compressed natural gas via a pipeline network across the nation.

CBM is a management philosophy that posits repair of replacement decisions on the current or future condition of assets; it recognizes that change in condition and/or performance of an asset is the main reason for exciting maintenance. Thus, Horner et al. noted that the optimal time to perform maintenance is deter-mined from actual monitoring of the asset, its subcomponent, or part. They also noted that condition assessment varies from simple visual inspections to elaborate automated inspections using a variety of condition monitoring tools and techniques [2].

In the history of development of machine diagnostics, signal processing method has become a central issue for CBM technology. However, most studies in the field of diagnostics have only focused on the experimental data given artificial fault. Hwang [3] has shown a novel approach for applying the fault diagnosis of rotating machinery. To detect multiple faults in rotating machinery, a feature selection method and support vector machine (SVM) based multi-class classifier are constructed and used in the faults diagnosis. Kim et al. [4] has developed a new fault detection method based on vibration signal for rotor machinery. Ahn et al. [5] has shown that signal processing and experiment under AE system is performed to evaluate the performance complex bearing fault classification.

In this paper, the comparative evaluations for intelligent fault diagnostics of HP-LNG pump have been performed using real industrial data including various

vibration signal processing techniques such as Hilbert Transform (HT) and Discrete Wavelet Transform (DWT) before extracting of dominant fault features from time domain and frequency domain signals. In other to estimate the abilities of fault classification techniques, three methods such as SVM (Support Vector Machine), k-NN (k-Nearest Neighbor) and Fuzzy-ARTMAP are used. To extract useful features, Distance Evaluation Technique (DET) has been applied in this study.

This paper is organized as follows: Sect. 2 briefly introduces the HP-LNG pump in LNG terminal. Section 3 describes the pump fault data used in this study through historical failure data analysis of HP-LNG pump. The proposed fault diagnostics methodology including feature extraction/selection techniques is explained in Sect. 4. Finally comparative evaluation results of intelligent fault diagnostics have been addressed in Sect. 5.

2 HP-LNG Pump

In LNG receiving terminal, HP-LNG pumps play a key role in determining the total supply capacity of natural. These HP-LNG pumps are submerged and operate at super-cooled temperatures. They are self-lubricated at both sides of the rotor shaft using LNG. Due to the low viscous value (about 0.16 cP) of LNG, the two bearings of the HP-LNG pump are poorly lubricated and bearing must be specially designed. Therefore, vibration and noise of HP-LNG pumps are regularly monitored and managed based on CBM (Condition based maintenance) techniques. Normally condition monitoring and fault diagnostic technology are applied for HP-LNG pump to evaluation of machine health and early detection before the catastrophic pump failures [1].

Table 1 shows the pump specifications. HP-LNG pump is installed in pump vessel with integrated shaft of the pump and motor. Two ball bearings are installed at top and bottom side of motor to support the entire dynamic load. The HP-LNG pump schematic, vibration measuring points are presented in the Fig. 2.

Vibration data are generally collected using two different condition monitoring techniques. One is on-line monitoring using two accelerometers installed motor housing and the other is off-line monitoring method using potable vibration data collection device from pump top plate.

Table 1 Pump specifications

Capacity	Pressure	Stages	Speed	Voltage	Rating	No. of pole	Current
241.8 m^3/h	88.7 bar	9	3585 rpm	6600 V	746 kW	2	84.5 A

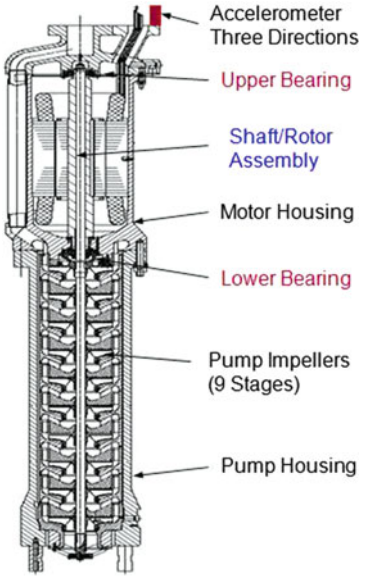


Fig. 2 HP-LNG pump schematic and vibration measuring points

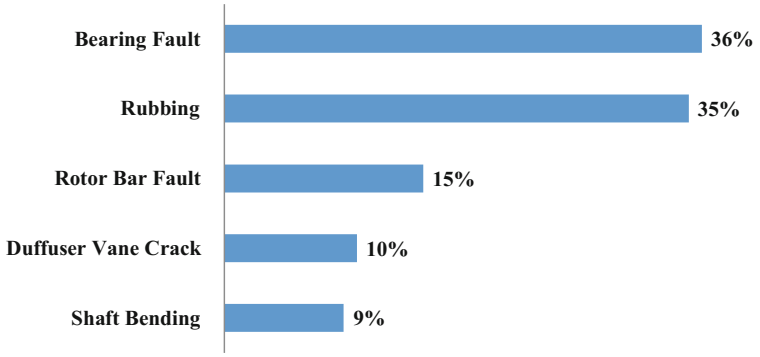


Fig. 3 Result of historical fault data analysis

3 Historical Data Analysis

In this research, the historical fault data have been analyzed to identify the main fault types of HP-LNG pump by using over 20 year's maintenance records. Figure 3 shows the statistical analysis result of HP-LNG pump faults. From this result, three predominant faults have been defined as main faults of HP-LNG pump during pump operation such as bearing fault, rubbing of wearing part and rotor bar fault as shown in Fig. 3.

To estimate the proposed methodology, the historical condition monitoring data collected on top plate points (Horizontal, Vertical, Axial) from nine HP-LNG

Table 2 Pump fault data set

Pump no.	Fault condition	Number of features	Sampling frequency
P702(F)	Rotor bar fault (RBF)	324	12,480
P702(A,B), P701(B,E,F)	Rubbing (RB)	324	12,480
P701(F), 702(D)	Bearing fault (BF)	324	12,480
P701(F), P702(A,C,D)	Normal (NOR)	324	12,480

pumps are used in this work. As shown in Table 2, total 80 vibration samples of four types (Rotor Bar Fault, Rubbing, Bearing Fault and Normal Condition) are obtained as described in Sect. 2. Each fault condition has 324 features which have calculated from frequency domain and time domain vibration data.

4 Proposed Fault Diagnostics Methodology

4.1 Diagnostics Process

In this paper, the proposed intelligent fault diagnostics for HP-LNG pump follows the typical procedure of intelligent fault diagnosis consisting of condition monitoring, signal processing, feature extraction and fault classification. The conventional feature-based diagnostics framework is illustrated in Fig. 4.

For signal processing of vibration monitoring data, mainly two pre-processing techniques such as DWT and HT have been employed for this research because these two methods has been extensively used in vibration signal analysis for fault conditions of machine. The statistical parameters form vibration time and frequency domain were calculated after pre-processing of vibration signals. Total 324 features from each condition of pump were calculated and used in this study. For outstanding performance of fault classification and reduction of computational effort, effective features for pump fault conditions were selected using the DET to evaluate the employed classification algorithms.

4.2 Feature Extraction

Feature extraction is a commonly used technique applied before classification when a number of measures, or features, have been taken from a set of objects in a typical statistical pattern recognition task. The goal is to define a mapping from the original representation space into a new space there the classes are more easily separable. This will reduce the classifier complexity, increasing in most cases classifier accuracy [6] (Table 3).

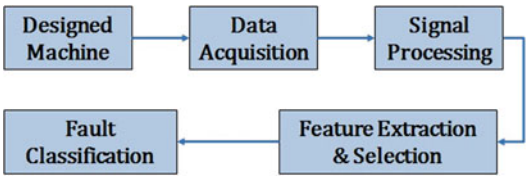


Fig. 4 Conventional feature-based diagnostics framework

Table 3 Features from time and frequency domains

Time-domain (12)		Frequency-domain (6)	
f_1	Mean	f_{13}	Frequency center
f_2	Root mean square	f_{14}	Mean square frequency
f_3	Shape factor	f_{15}	Root mean square frequency
f_4	Crest factor	f_{16}	Variance frequency
f_5	Skewness	f_{17}	Root variance frequency
f_6	Kurtosis	f_{18}	Mean average deviation frequency
f_7	Clearance factor		
f_8	Impulse factor		
f_9	Lower-bound of histogram		
f_{10}	Upper-bound of histogram		
f_{11}	Standard deviation		
f_{12}	Square mean		

For outstanding performance of fault classification and reduction of computational effort, effective features were selected using the distance evaluation technique of feature effectiveness introduced by Knerr et al. [7] as depicted below. The average distance ($d_{i,j}$) of all the features in state i can be defined as follows:

$$d_{i,j} = \frac{1}{N \times (n-1)} \sum_{m,n=1}^{N} \left| p_{i,j}(m) - p_{i,j}(n) \right| \tag{1}$$

The average distance ($d_{i,j}$) of all the features in different states is

$$d'_{i,j} = \frac{1}{M \times (M-1)} \sum_{m,n=1}^{M} \left| p_{ai,m} - p_{ai,n} \right| \tag{2}$$

where, m, n = 1, 2, …, m ≠ n, $P_{i,j}$: Eigen value, i: data index, j: class index, a: average, N: number of feature and M: number of class.

When the average distance ($d_{i,j}$) inside a certain class is small and the average distance ($d'_{i,j}$) between different classes is big, these averages represent that the features are well separated among the classes. Therefore, the distance evaluation criteria (α_i) can be defined as

$$\alpha_i = d'_{ai}/d_{ai} \tag{3}$$

The total 24 features were selected using distance evaluation technique (DET), these features were mainly made up of more dynamic and variable parameters such as mean square frequency (f_{14}), root variance frequency (f_{17}) and mean average deviation frequency (f_{18}) among number of features which were calculated after signal processing techniques such as DET and HT from pump vibration data.

4.3 Fault Classifiers

Three classification algorithms employed in this research for intelligent fault diagnostics of HP-LNG pump are briefly introduced in this section.

4.3.1 SVM

Several methods have been proposed regarding Multi Class Support Vector Machines (MCSVMs) such as one-against-one (OAO), one-against-all (OAA) and one-acyclic-graph (OAG). Among these methods, OAA method is employed in this research because this method is a simple and effective for multi-class classification, and has been successively demonstrated with its high classification performance. In addition, selection of appropriate kernel function is very important to classify in feature space. In various kernel functions, the radial basis function (RBF) is the most popular kernel type for fault diagnostics because it finds a set weights for a curve fitting problem in complex faults feature set as stated below:

$$\text{K\_RBF} = \exp(\|X - X'\|)/\gamma \tag{4}$$

where $\gamma = 1/(2\sigma^2)$, σ can decide more flexible degree of boundary.

4.3.2 k-NN

Nearest neighbors methods can be used as an important pattern recognition tool. In such methods, the aim is to find the nearest neighbors of an undefined test pattern within a hyper-sphere of pre-defined radius in order to determine its true class. Nearest neighbors methods can detect a single or multiple numbers of nearest neighbors. A single nearest neighbor method if primarily suited to recognize data where we have sufficient confidence in the fact that class distributions are non-overlapping and the features used are discriminatory. In most practical applications, however, the data distributions for various classes are overlapping and more than one nearest neighbors are used for majority voting [8].

4.3.3 Fuzzy ARTMAP

A number of ART neural network architectures have been progressively developed. Recently, a growing number of models computationally synthesize proper-ties of neural networks, and fuzzy logic. Fuzzy-ARTMAP is one such model, combined with ARTMAP and fuzzy logic. Fuzzy-ARTMAP utilizes a minimax learning rule that conjointly minimizes prediction error and maximizes generalization. As learning processing, the input and stored prototype of a category are said to resonate when they are sufficiently similar. When an input pattern is not sufficiently similar to any existing prototype a new category is formed having the input pattern as prototype [9].

5 Comparison Results of Fault Diagnostics

Tables 4, 5, and 6 show the comparison results of classification performance using three conventional classifiers such as SVM, k-NN and Fuzzy-ARTMAP for fault diagnostics of HP-LNG pump.

As shown in Tables 4, 5, and 6, SVM indicates higher classification performance than Fuzzy-ARTMAP and k-NN relatively. Especially, SVM has a good capability for fault classifications such as rotor bar fault (RBF) and bearing fault (BF) compared to rubbing (RB) and normal condition (NOR) of HP-LNG pump.

The comparison results of classification algorithms are summarized in Fig. 5. The SVM shows high accuracy among three classification algorithms except for the normal condition.

Table 4 Classification performance using SVM

	RBF	RB	BF	NOR
RBF	8	1	0	1
RB	0	6	0	1
BF	0	1	8	1
NOR	0	0	0	5
Accuracy (%)	100	75	100	62.5
Avg (%)	84.38			

Table 5 Classification performance using k-NN

	RBF	RB	BF	NOR
RBF	7	1	0	1
RB	1	3	2	0
BF	0	2	5	1
NOR	0	2	1	6
Accuracy (%)	87.5	37.5	62.5	75
Avg (%)	65.63			

Table 6 Classification performance using Fuzzy-ARTMAP

	RBF	RB	BF	NOR
RBF	3	3	0	0
RB	0	1	1	0
BF	0	3	2	0
NOR	5	1	5	8
Accuracy (%)	37.5	12.5	25	100
Avg (%)	43.75			

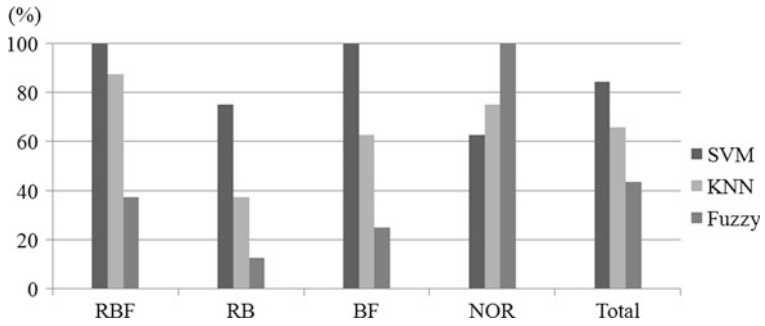


Fig. 5 Comparison result of SVM, k-NN and Fuzzy-ARTMAP for fault diagnostics

6 Conclusions

In this paper, three classification techniques for the intelligent fault diagnostics of HP-LNG pump were evaluated using real industrial data. For signal processing of vibration monitoring data, mainly two pre-processing techniques such as Discrete Wavelet Transform and Hilbert Transform have been employed in this research. To improve classification performance, distance evaluation technique (DET) has been used to select the features that reflect the dynamic characteristics of each fault condition.

The comparison results indicate that SVM has a better capability for HP LNG pump fault diagnostics than Fuzzy-ARTMAP and k-NN relatively. Especially, SVM provides more accurate performance for fault conditions such as rotor bar fault(RBF) and bearing fault(BF) compared to rubbing(RB) and normal condition (NOR) for the intelligent fault diagnostics of HP-LNG pump.

Acknowledgments This research was conducted by support of Korea Gas Technology Corporation (KOGAS-Tech).

References

1. H. E. Kim and J. S. Oh (2014) Comparative study of intelligent fault diagnostics for LNG pump failure. In: WCEAM 2014
2. Byron A. Ellis (2008) Condition based maintenance. In: TJP, November 10, 2008, pp 1–5
3. Yu W, Hwang W (2003) Fault diagnosis of machinery using multi-class support vector machines. In: KSNVE annual fall conference 2003, pp 537–543
4. Kim DH, Shon SM, Kim YW, Bea YC (2014) Rotating machinery fault diagnosis method on prediction and classification of vibration signal. In: KSNVE annual fall conference 2014, pp 90–93
5. Ahn BH, Kim YH, Lee JM, Ha JM, Choi BK (2014) Image fault classification for AE complex fault signal. In: KSNVE annual spring conference, pp 417–419
6. Huang HZ, Qu J, Zuo MJ (2009) Genetic-algorithm-based optimal apportionment of reliability and redundancy under multiple objectives. IEEE Trans 41:287–298
7. Knerr S, Personnaz L, Dreyfus G (1990) Single-layer learning revisited: a stepwise procedure for building and training a neural network. Neurocomputing 68:41–50
8. Yang BS, Widodo A (2010) Introduction of intelligent machine fault diagnosis and prognosis. Nova Science, New York
9. Carpenter GA, Grossberg S, Markuzon N, Reynolds JH, Rosen DB (1991) Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps. IEEE Trans Neural Netw 3:698–713

Rotating Machine Prognostics Using System-Level Models

Xiaochuan Li, Fang Duan, David Mba, and Ian Bennett

Abstract The prognostics of rotating machines is crucial for the reliable and safe operation as well as maximizing usage time. Many reliability studies focus on component-level prognostics. However, in many cases, the desired information is the residual life of the system, rather than the lifetimes of its constituent components. This review paper focuses on system-level prognostic techniques that can be applied to rotating machinery. These approaches use multi-dimensional condition monitoring data collected from different parts of the system of interest to predict the remaining useful life at the system level. The working principles, merits and drawbacks as well as field of applications of these techniques are summarized.

Keywords Rotating machinery • Prognostics and health management • Condition monitoring • Multivariate models

1 Introduction

Rotating systems such as gas turbines and compressors are widely used due to their high performance and robustness. However, many kinds of failures may occur during the operation of the machine. Those failures will cause unplanned downtime and economic losses as well as reduced reliability. One way to minimize the negative influence of these failures is to make maintenance strategies more predictive by using automated condition monitoring. Condition-based maintenance (CBM) is a preventive maintenance strategy that seeks to improve the reliability of engineering systems based on condition monitoring information [1]. CBM enables the diagnosis of impending failures and the prognosis of the future health state and remaining useful life (RUL) of a system [2]. Prognostic programs based on condition monitoring data provide a potent tool for practitioners in making

X. Li (✉) • F. Duan • D. Mba
London South Bank University, SE1 0AA London, UK
e-mail: lix29@lsbu.ac.uk; duanf@lsbu.ac.uk; mbad@lsbu.ac.uk

I. Bennett (✉)
Shell Global Solutions International, B.V., The Hague, The Netherlands
e-mail: ian.bennett@shell.com

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_11

appropriate maintenance decisions by estimating the future degradation trends and anticipating the failure time [3]. Suspensions or overhauls could be carried out before the estimated failure time, which allows for improved machine availability and reliability and reduced the overall operating cost.

Many condition-based reliability studies have focused on component-level prognostics, which enable the failure of critical rotating components to be predicted. However, in many cases, the desired information is the residual life of the system, rather than the lifetimes of its constituent components [4]. This necessitates the development of robust system-level prognostic techniques for rotating machines.

Due to the progress of sensing technology, condition monitoring data such as the oil debris, pressure values, temperature values and vibration is available at different parts of the rotating system [5]. The availability of data from multiple sensors has provided the possibility of developing multidimensional prognostic techniques. Since the RUL of a system is dependent upon its constituent components and how they interact [6], one could make use of the data collected from sensors distributed over the machine and multi-dimensional techniques to predict failure times. Subsequently, system maintenance schedules can be made based on the estimated failure.

Several review papers on prognostic techniques for engineering systems have been published [7–13]. However, limited numbers of these papers have highlighted the system-level prognostic options for complex rotating machines. This paper reviews prognostic techniques that can be applied to predict rotating machinery failures at the system level.

2 Discussions on System-Level Prognostic Models for Rotating Machinery

2.1 Characteristics of Rotating Machinery Prognostics

Compared with general industrial applications, rotating systems have several unique characteristics that should be considered when developing prognostic methods. First, various sources of nonlinearities can be present in complex rotating machines, such as nonlinearities in the bearings, aerodynamic effects and friction in rotating assemblies and seals [14]. The presence of these nonlinear elements can lead to nonlinear system dynamic characteristics. Moreover, most real-world rotating machinery works in non-stationary operating conditions, which can have profound effects on both diagnostic and prognostic signals. The non-stationary operations are generated by load/speed variations, system parameter adjustments, strong nonlinearities in the components, etc. [15, 16]. Moreover, rotating systems are composed of multiple sub-systems and components with various failure modes, which further introduce a degree of complexity in prognostic modelling. In addition to non-linearity, non-stationary and multiple failure modes, modelers should also consider the possible synergy among the different sensor signals collected from the

machine. Therefore, the multidimensional prognostic techniques are commonly used when developing an appropriate system-level method for rotating machinery.

2.2 *Prognostic Techniques Categorization*

The prognostic approaches reviewed in this paper can be divided into three categories: (1) statistical methods, (2) artificial intelligence methods, and (3) similarity-based methods (see Fig. 1).

For statistical methods, we review models based on Bayesian theory and proportional hazard models. These models predict the RUL based on past observed information and statistical models in a probabilistic manner. Therefore, a probability density function (PDF) of the RUL is formulated for uncertainty management. Moreover, in statistical methods, machine lifetime data might be required in addition to monitoring data for failure prediction.

When lifetime data is scarce or non-existent, artificial intelligence methods that make predictions using only monitoring data can be considered. These models can be treated as a nonlinear function approximator, which aims to determine dependencies in a training data set such that predictions of the outputs (e.g., RUL) can be made when new inputs are available. Unlike statistical methods, most artificial intelligence methods do not provide a PDF of the RUL. For artificial intelligence methods, we review models based on neural networks and support vector machines (SVMs).

Similarity-based methods are fundamentally different from the techniques in the first two categories because they do not perform trending or extrapolation of the degradation process. Instead, they construct a probabilistic health indicator, which characterizes the system health state via trajectories. Then, predictions are made based on evaluating similarities between the trajectories.

The working principles, merits, drawbacks and the applications of these techniques are discussed in the following sections.

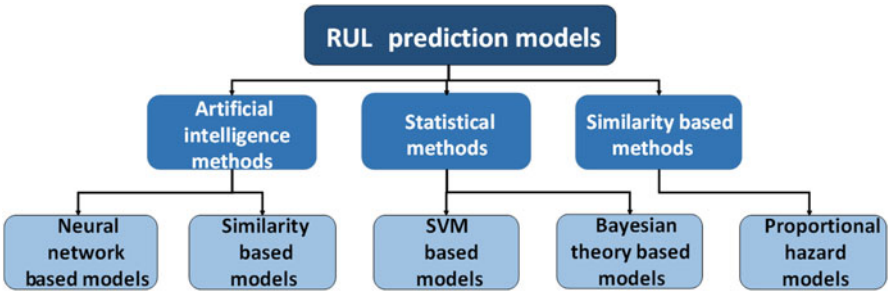


Fig. 1 Models categories for RUL prediction

2.3 Bayesian Theory Based Models

We assume the system model based on Bayesian theory can be defined as:

$$p(x_t|x_{t-1}) \tag{1}$$

$$p(y_t|x_t) \tag{2}$$

Where x_t refers to the unobservable health state of the system under study at time t, y_t refers to the observed information at time t. $p(x_t|x_{t-1})$ is the state equation, and $p(y_t|x_t)$ is the observation equation. Then, the prognostic tasks can be divided into two sequential stages: the estimation stage and the prediction stage. The purpose of the estimation stage is to find a state estimate $p(x_t|Y_t)$ given the observation history up to time t, where Y_t denotes the observation history up to time t. The prediction problem is to predict the health state $p(x_{t+k}|Y_t)$ at a given prediction time $t + k$. The RUL can be computed by extrapolating the predicted health state to a pre-set failure threshold [4].

Because most engineering facilities in practice undergo a non-linear and time-varying deterioration process, the parameters of the prediction model cannot remain constant over the entire prediction process (otherwise the predictions might be not accurate). Therefore, Bayesian updating methods are often adopted to jointly update the model parameters and the system health state when new observations are available. Three widely used Bayesian updating tools are particle filtering [4, 17], semi-stochastic filtering [18] and dynamic Bayesian updating [19]. To address systems with multiple sensory inputs, sensor fusion techniques, such as principal component analysis [20], independent component analysis (ICA) (Wang and Hussin [21]), linear regression [4] and path model [19] are often employed to merge the multivariate measurements into a one-dimensional variable—the system health indicator. Then the equations defined in the first paragraph can be used to map health indicator values to the RUL. Some common Bayesian theory techniques will be discussed below.

1. *Particle Filtering:* The main concept of particle filtering is to represent the required posterior distribution of the state variables, such as $p(x_t|Y_t)$, using a set of particles with associated weightings. These particles evolve and adapt recursively when new information becomes available [22]. Then, the unobserved health state is estimated based on these weighted samples.

In health state prediction, particle filter has three advantages: (a) it can be applied to nonlinear process with non-Gaussian noise; (b) it provides probabilistic results which is helpful to manage the prognostic uncertainties; (c) it allows information fusion such that data from multidimensional sensors could be employed collectively [23, 24]. However, one limitation of particle filtering is that a large number of samples might be required to accurately approximate the future state distributions, which may cause the filtering system to collapse. A good

solution to this problem is to adopt the efficiency monitoring method of filtering proposed by Carpenter [22].

Wang [17] presented an engine wear estimation model base on particle filtering. In his work, the relationship between condition monitoring measurements and system wear was modelled using the concept of a floating scale parameter. In this approach, the scale parameter of the observations is a function of both system degradation and time. The PCA was employed to produce a one-dimension representation of the metal concentration data, which were then processed by particle filtering to obtain the density function of system wear. Recently, Sun et al. [4] applied a state space model embedded with particle filtering to a gas turbine degradation data set obtained via simulation. A health indicator inferred using a linear regression method was used to represent the latent degradation state of the engine given multivariate sensory measurements. The authors combined the state estimation with model parameter estimation to reduce the prognostic uncertainty.

2. *Semi-stochastic Filtering:* Wang and Christer [5] firstly developed a state space prognostic model embedded with a semi-stochastic filtering technique. Based on the authors assumption, two relationships should be determined to model the probability density of a system's state given all of the observations: the relationship between x_t and x_{t-i} and the relationship between y_t and x_t. x_t is the residual life at time t, and y_t denotes the observation at time t. The two relationships could be described by $x_t = x_{t-i} - (t-i)$ and $y_t = g(x_t, \delta_t)$, where δ_t denotes a noise term and g is a function to be determined. Then, the posterior distribution of the residual life given all past observation history can be estimated based on the obtained conditional probabilities. Various extensions have been developed and applied to rotating system prognostics based on the above framework. A revision of this semi-stochastic filtering technique was applied to the lifetime data and monitored oil analysis data collected from an aircraft engine [20]. The predicted RUL is assumed to be proportional to the wear increment measured by the monitored measurements. PCA was employed to obtain a weighted average of the original monitored data. A similar model was given in (Wang et al. [25]). The authors combined lifetime data and accumulative metal concentration data to estimate the remaining useful life of a diesel engine. Similarly, Wang and Hussin [21] developed a stochastic filtering-based prognostic model and applied it to two data sets: engine lubricant and contaminant analysis data and metal concentration data. Instead of the commonly used PCA, they employed ICA to merge the latter. The results indicated that a higher accuracy was achieved based on the data lubricant and contaminant sets. Another extension of Wang's semi-stochastic filtering was given in [26]. This model extends the original filtering with respect to two aspects: the concept of a two-stage life model was introduced to achieve both fault detection and prediction, and a combination of categorical and continuous hidden Markov chain was used to model the underlying health state transitions. The authors suggested the use of a PCA algorithm in

combination with the proposed model to address multidimensional data in complex rotating systems.

3. *General Path Model with Dynamic Bayesian Updating:* Coble and Hines [19] developed a prognostic model, called the general path model (GPM), to predict the RUL of aircraft engines. First a deterioration measure is identified to represent the failure evolution. Then a linear regression fit of the measure is extrapolated to a preset failure threshold to predict the RUL. The results indicated that the dynamic Bayesian updating method greatly improved the prediction accuracy.

These approaches are based on the same assumption that there is no maintenance actions during two condition-check points or that the actions do not affect the system degradation pattern. However, this may not be the case in reality. Moreover, failure lifetime data may be required for parameter estimation for most models mentioned above. But this type of data might be scarce in reality.

2.4 Proportional Hazard Models

Machine failures can be predicted by analyzing either condition monitoring data or historical service lifetime data [27, 28]. Condition monitoring data, which are obtained continuously, have been widely used to predict fault evolutions and system RUL. The lifetime data, which indicate how long the machine has been operating since the last failure (or suspension), also provide supplementary information for RUL prediction [29]. Hence, it would be wise to develop proper prognostic models with a combination of condition monitoring data and lifetime data. The proportional hazard model (PHM), proposed by Cox [30], attempts to utilize both types of information for RUL prediction. The basic assumption of this method is that the failure rate of a machine depends on two factors: the baseline hazard rate and the effects of covariates (condition monitoring sensory variables). Hence, the hazard rate of a system at service time t can be written as $\lambda(t; z) = \lambda_0(t)$ $\exp(z\beta)$, where $\lambda_0(t)$ denotes the base line hazard, which is determined by the system lifetime data. $\exp(z\beta)$ denotes the covariate function, which describes the effect of the sensory variables on the degradation process [28]. Applying PHMs requires that the baseline hazard function $\lambda_0(t)$ and covariate function $\exp(z\beta)$ be identified.

Methods that have been used to estimate the $\exp(z\beta)$ including the maximum likelihood algorithm ([3]; Cox [30]) and Wald statistic [29]. The relative influences of sensory signals on the system hazard rate are first determined. Then, key variables with close correlation to the system failure are retained and employed to estimate the system failure probability density [31]. Once the covariate function has been determined, the baseline function parameters can be estimated. The PHM provides a distribution-free estimate of the baseline function $\lambda_0(t)$ which means that a specific distribution for $\lambda_0(t)$ is not needed to fit the lifetime data. Researchers

prefer this type of estimate because it can avoid the loss of accuracy caused by the assumption of a parametric distribution [32]. However, in practice, the baseline hazard function is often assumed to be a parametric distribution, such as the Weibull or exponential distribution [11]. Such assumptions might not be reasonable in many cases because of the confusing effects of different covariates [31].

To apply the PHMs, the sensory measurements and lifetime data are combined to fit the model. The PHM can identify the important risk factors from all input variables and their relative influence on the failure of the equipment [29]. Then, the system failure distribution at time service t can be estimated. Finally, the failure time at time t can be predicted according to the estimated probability distribution.

PHMs have been applied to many non-linear and non-stationary machinery prognostic problems. Jardine et al. [33] developed a PHM and employed it to estimate the remaining useful life of aircraft engines and marine gas turbines. The baseline hazard function was assumed to be a Weibull distribution and was estimated using lifetime data. The levels of various metal particles (such as Fe, Cu and Mg) in the oil were used as the covariates in both cases. The influence of the condition monitoring variables on the equipment RUL has been properly interpreted by the developed PHM. The authors also applied the PHM to estimate the RUL and optimize the maintenance decisions of haul truck wheel motors in [29]. The key covariates related to failures were identified from 21 monitored oil analysis variables using the developed PHM. The results showed that significant savings in maintenance costs could be achieved by optimizing the overhaul time as a function of lifetime data and oil analysis variables.

The above models are based on the assumption that the system under study is subject to a single failure mode. However, in practice, most complex mechanical systems consist of multiple sub-systems with various failure modes [28]. Therefore, prognostic models for determining only one type of failure mode cannot properly estimate the overall system failure time. Recently, Zhang et al. [28] proposed a mixed Weibull proportional hazard model (MWPHM) for complex mechanical system reliability assessment. In this model, the overall system failure probability density is determined by a mixture of failure densities of various failure modes. The influences of multiple monitoring signals on different failure modes are integrated using the maximum likelihood estimation algorithm. Real data from a centrifugal water pump were combined with lifetime data to test the robustness of the model.

The main problem with applying PHMs for failure prediction is that they require a large amount of lifetime data to determine the parameters of the baseline hazard function and the weighting of covariates [27], which may limit the applications of PHMs because the amount of lifetime data might be insufficient in many cases due to various reasons, such as missing or no records or transcription mistakes [34]. Another drawback of PHMs is that they rely on the choice of the failure threshold for RUL prediction. The threshold must be continuously updated when system maintenance is conducted [28].

2.5 Neural Network Based Models

Artificial neural networks (ANNs) have recently been widely used in modelling degradation processes. An ANN is a computing system that is able to capture, represent and compute mapping from the input multi-variable space to the output space [35]. ANNs have three layers: an input layer, one or more hidden layers and an output layer. ANNs are comprised of a large number of processing elements (known as neurons) that are connected to each other by weighted interconnections [36]. These neurons are organized into distinct layers, and their interconnections are determined through a training process. The network training involves presenting data sets collected from the degradation process. Then the network parameters are adjusted to minimize the errors between the model output and desired output [35]. Once the training is finished, ANNs process new input data to make predictions about the outputs (RUL).

Network architectures that have been used for prognostics can be classified into two types: feed-forward and recurrent networks [37]. In feed-forward networks, the signals flow in one direction; therefore, the inputs to each layer depend only on the outputs of the previous layer. However, applications in signal processing and prognostics should consider the system dynamics. Recurrent networks provide an explicit dynamic representation by allowing for local feedbacks [38]. Two types of networks: multi-layer perceptron (MLP) and recurrent neural networks (RNNs) (Fig. 2 shows the architecture of a simple RNN), which have been applied extensively by researchers, will be discussed below [35].

1. *Multi-Layer Perceptron (MLP):* MLPs are one of the most popular feed-forward neural networks used for prognosis [11]. MLPs utilize back-propagation (BP) learning technology for training. After the training, the MLP is capable of classifying the fault and predicting the RUL based on new measurements

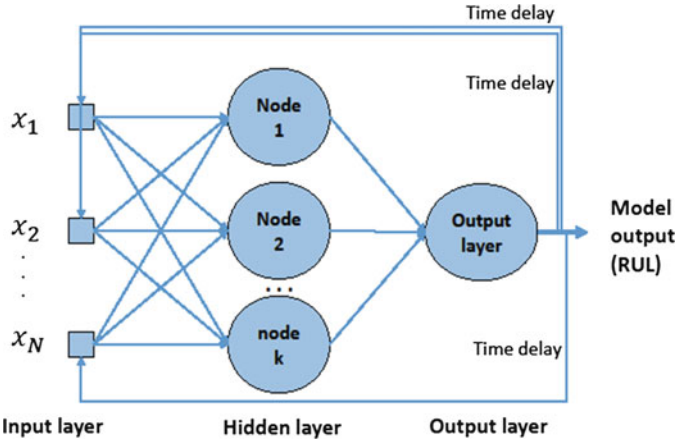


Fig. 2 Architecture of a simple RNN

collected from machines [39]. The benefit of back-propagation (BP) training is that it does not require knowledge of the precise form of the input-output mapping functions (e.g., function type, number of model parameters) of the model to be built, which makes it suitable for the analysis of multivariate complex systems [12, 13].

2. *Recurrent neural network (RNN):* Feed-forward neural networks have limitations in identifying temporal dependences in time series signals. RNNs solve this problem by including local or global feedback between neurons. Thus, they are suitable for a wide range of dynamic systems [40], such as time-varying and non-linear systems. However, the drawback of RNNs is the limitations in accurate long-term predictions arising from the frequently used gradient descent training algorithm [40].

ANNs can represent and build mappings from experience and history measurements to predict the RUL and then adapt it to unobserved situations. The strong learning and generalization capabilities of ANNs render them suitable for modelling complex processes, particularly systems with nonlinear and time-varying dynamics [41]. ANNs are superior in capturing and depicting relationships between many variables in high-dimensional data space [42, 43]. RNNs are suitable for approximating dynamic dependencies [40]. These distinct characteristics make ANNs promising candidates for modelling degradation processes in rotating machinery.

Xu et al. [44] successfully employed RNNs, support vector machines (SVMs) and Dempster-Shafer regression to estimate the RUL of an aircraft gas turbine. Echo state network (ESN), which is a variant of the RNNs, was employed by Peng et al. [45] to predict the RUL of engines using NASA repository data. The results indicated that the ESN significantly reduced the computing load of the traditional RNNs. ANNs have also been used in combination with Kalman filters and Extended Kalman filters in [46] and (Felix [47]) to perform failure predictions of aircraft engines.

Although ANNs have been shown the superior power in addressing complex prognostic problems which have multivariate inputs, there are some limitations. For example, the majority of the ANN based prognostic models aim to assume a single failure mode and do not relate lifetime data with the machine RUL. Moreover, the models rely on a large amount of data for training. The prognostic accuracy is closely dependent on the quality of the training data [44]. Furthermore, ANNs allow for few explanatory insights into how the decisions are reached (also known as the black box problem), which has become a concern to modelers because the causal relationship between the model variables is essential for explaining the fault evolution [48]. Attempts to solve the black box problem can be found in [49]. Moreover, ANNs lack a systematic approach to determine the optimal structure and parameters of the network to be established [12, 13].

2.6 Support Vector Machine Based Models

Previously, SVMs were mainly used for pattern recognition problems and have not been used for time series forecasting until the introduction of the Vapniks insensitive loss function [8]. SVM-based machine learning starts with a number of input variables $x(i), i = 1, 2, 3, \ldots, N$, and the corresponding target values $y(i)$, $i = 1, 2, 3, \ldots, N$. The idea is to learn the dependency of $y(i)$ on $x(i)$ and to define a function over the input variables. Then, predictions of $y(i)$ can be made given unseen $x(i)$ [50]. When applying SVMs to nonlinear prognostics, model inputs are first mapped onto a higher dimensional feature space by using a kernel function. The most commonly employed kernel function is the radial-based function (RBF) [51]. Then a linear model is constructed in the feature space to make estimation. Figure 3 shows the architecture of a simple SVM based prognostic model. SVMs are excellent in addressing prognostic problems regarding complex rotating machinery because they have no limitations on the dimensionality of the input vectors and have relatively a low computational burden [12, 13]. Besides, SVMs can achieve highly accurate results with nonlinear inputs.

Several different prognostic models based on SVMs have been used in modelling nonlinear and non-stationary dynamic systems.

1. *Relevance Vector Machine (RVM):* RVM has an identical function form as a SVM. Hu and Tse [52] proposed a model based on RVM for RUL prediction of a pump. This model was proved to be accurate when dealing with non-stationary vibration signals.
2. *Particle Swarm Optimization and SVM (PSO-SVM):* García Nieto et al. [53] developed an RUL framework based on the PSO-RBF-SVM technique. This model combines a SVM with particle swarm optimization (PSO) to enable the parameter adjustment of the RBF kernel function. The results show that the

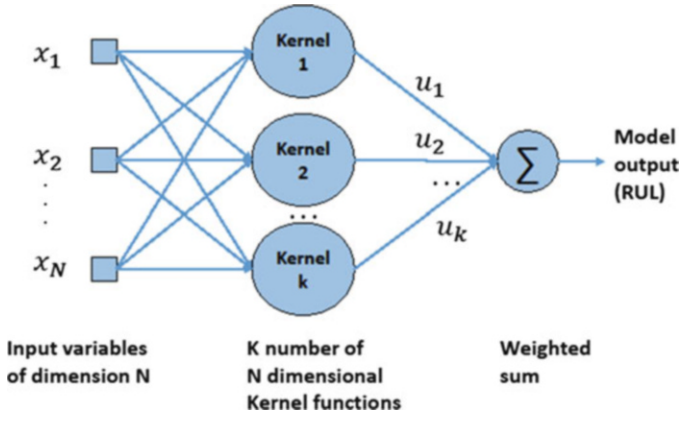


Fig. 3 Architecture of a simple SVM based prognostic model

proposed prognostic model accurately predicts the RUL of engines based on a simulation data set (collected from the MAPSS).

3. *Least Squares Support Vector Regression and Hidden Markov Model (LSSVR and HMM):* Compared with the traditional SVM, LSSVR can lead to better performance in addressing non-linear, small sample problems. Li et al. [54] proposed a hybrid model of the LSSVR and the HMM for RUL prediction. The RUL was calculated by the LSSVR model built on the health indexes obtained from HMM.
4. *SVM and RNN:* In [44], a SVM approach based on the RBF kernel function was employed together with a RNN and Dempster-Shafer regression to predict the RUL of engines. The integrated prognostic method demonstrated superior capacity in providing accurate predictions.

However, the problem with using SVM is that a standard method of choosing an appropriate kernel function for SVMs does not exist [8]. In addition, parameters should be specifically tuned for the case of interest and this might be challenging. Efforts should be made to choose the appropriate kernel functions and estimate the appropriate parameters.

2.7 Similarity Based Models

Similarity-based prognostic models are particular cases of data-driven models and have only recently been applied to complex rotating machinery. These models are essentially pattern matching approaches [55]. Similarity-based prognostic models are suitable for situations in which abundant run-to-failure data of a mechanical system are available [56]. Multivariate monitoring data collected from various failure modes and operating conditions [55] of the system are first processed to produce a health indicator (HI). The indicator represents the fault evolution of the system by trajectories. The methods to obtain the health indicator trajectories include logistic regression [56], weighted averaging methods [57], and flux-based methods [58]. If the un-processed data already capture the progression of the degradation process, the data can remain multi-dimensional [59]. Then, the monitoring data are converted into instances. An instance can be either a segment of the HI trajectories or a complete degradation trajectory. Therefore, a library of instances can be created from these run-to-failure data and then stored in the memory. If one wishes to predict the RUL using a new run-to-failure dataset, the same operations are applied to the new data to produce a new instance. Instead of extrapolating, the instance is compared with the stored instances to determine and select the instances with the best matching scores (i.e., the most similar ones [56]. Then, the best matching instance is used to extrapolate the RUL or the weighted multiple instances are added together to calculate the RUL [55].

Figure 4 shows the general framework of similarity-based prognostic models. Because the similarity-based approaches use training data to construct instances

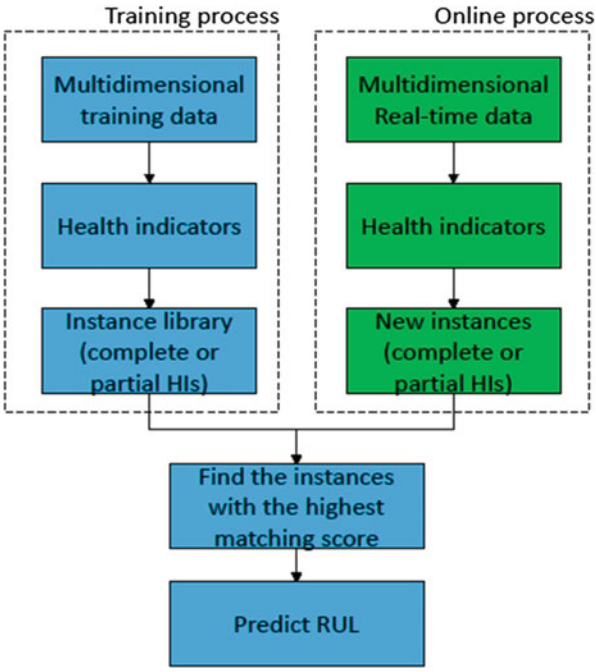


Fig. 4 General framework of similarity-based prognostic models

(health indicator trajectories or multidimensional monitoring variables), they are compatible with algorithms that extract health indicators for RUL prediction (Malinowski et al. [59]). The advantage of similarity-based approaches is that they can achieve satisfactory and accurate predictions when abundant data are collected from a variety of failure modes. However, the run-to-failure data are scarce in many cases [55]. Hence, efforts should be made to extend this type of approach to situations in which limited training data are available. Additionally, many similarity-based prognostic techniques suffer from computational inefficiency in terms of sorting a large amount of training data [60].

The ability to accommodate multidimensional sensory measurements collected from various failure patterns makes similarity-based methods suitable for the prognostics of complex rotating machinery. Examples are given below to set out how various similarity based models have been used for RUL prediction.

1. *Similarity Model Based on Shapelet Extraction:* Malinowski et al. [59] developed an RUL prediction technique that employs the Shapelet extraction process to extract failure patterns from multivariate sensory data obtained from a turbofan engine simulation program, C-MAPSS. The RUL was calculated as the weighted sum of the failure patterns that are highly corrected with the residual life.
2. *Similarity Model Based on Normalized Cross Correlation:* Zhang et al. [61] applied a prognostic method based on the similarity of phase space trajectory to

the monitoring data collected from a pump with six distinct degradation modes. The normalized cross correlation was employed to determine the optimal matching trajectory segments, which were then used to estimate the RUL.

3. *Similarity Model Based on PCA and K-NN:* Mosallam et al. [62] employed principle component analysis (PCA) and an empirical mode decomposition (EMD) algorithm to construct health indicators from turbofan engine deterioration simulation data. Then, the k-nearest neighbor (K-NN) classifiers were used to determine the most similar HIs for RUL prediction.
4. *Similarity Model Based on Belief Functions:* An improved technique based on belief functions was proposed by Ramasso and Gouriveau [60] and Ramasso et al. [63]. In this method, the authors only match the last points of the trajectories in the library with tested ones because the last points are more likely to be closely related to the degradation state. One of the main contributions of this method is its ability to manage labels (indicating degradation states) that would have been incorrectly assigned to the sensory data.
5. *Similarity Model Based on Linear Regression and Kernel Smoothing:* Wang et al. [56] proposed a prognostic model in which the health indicator is constructed from multiple sensors using linear regression. The best matching instances were selected by examining the Euclidean distance between the test and stored instances. This method was applied to data provided by the 2008 PHM Data Challenge Competition to predict the RUL of an unspecified system.
6. *Similarity Model Based on Regression Vector Machine (RVM):* Wang et al. [64] improved the previous model by incorporating the uncertainty information into the RUL estimation. The degradation curves of health indexes were estimated using RVM. The Challenge data was employed again to test the effectiveness of this method.
7. *Hybrid Similarity Model:* Hu et al. [65] proposed an ensemble prognostic model which combines five individual algorithms (i.e. similarity-based approach with the SVM, RVM and Exponential fitting, a Bayesian linear regression with the quadratic fitting and a RNN) with a weighted sum formulation. The integrated model shows higher accuracy in RUL prediction compared to any single algorithm.

3 Summaries of Prognostic Techniques of Rotating Machines

Table 1 summarizes the application of different RUL prediction models to various industrial rotating machines and the machines' common available data types.

Furthermore, the reviewed manuscripts (those in Table 1) are classified based on the type of data used in the article. There are two types of data, namely, simulated data collected from simulation programs, such as C-MAPSS, and field data (real-world condition monitoring data). Figure 5 compares the type of data being used in

Table 1 Applications of RUL prediction models

Rotating machine type	RUL prediction models	Common available data types
Gas turbine engines	Similarity model based on shapelet extraction [59]) Similarity model based on linear regression and kernel smoothing [56]) Similarity model based on PCA and K-NN [62]) Similarity model based on belief functions [63]) Similarity model based on RVM Wang et al. [64]) Hybrid similarity model based on SVM, RVM, Exponential fitting, Quadratic fitting and RNN [65]) Echo state network [45]) Multi-layer perceptron and kalman filter [46] Recurrent neural network and extended kalman filter Felix [47]) Recurrent neural network and support vector machine [44]) Particle filtering and linear regression Wang [17], [4]) Semi-stochastic filtering and PCA [20, 21] Linear regression and dynamic Bayesian updating [19]) Weibull proportional hazard model [33]) PSO-SVM [53]) Least squares support vector regression and HMM [54])	1. Condition monitoring data: Vibration, metal concentration, acoustic emission, ratio of fuel flow, temperature of Fan, low/high pressure compressor and low/high pressure turbine, pressure, fan speed, etc. 2. Lifetime data
Pumps	Similarity model based on normalized cross correlation [61]) Mixture of Weibull proportional hazard model [28]) Relevance vector machine [52])	1. Condition monitoring data: Mainly vibration, can involve pressure, temperature, etc. 2. Lifetime data
Diesel engines	Semi-stochastic filtering and PCA [25]	1. Condition monitoring data: Metal concentration, etc. 2. Lifetime data
Haul truck wheel motors	Weibull proportional hazard model [29])	1. Condition monitoring data: sediment, viscosity, voltage, load, vibration, etc. 2. Lifetime data

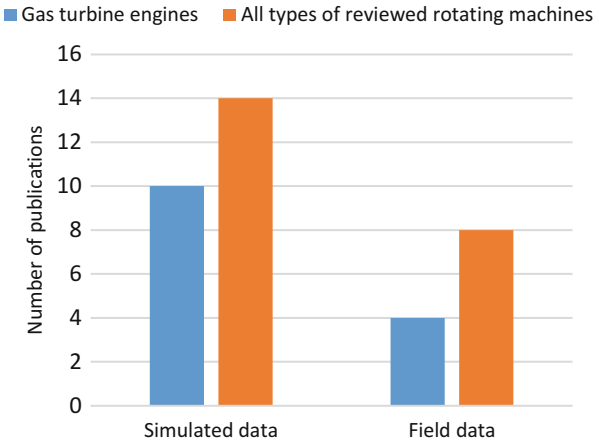


Fig. 5 Type of data used in studies regarding (**a**) Gas turbine engines only (*blue*), and (**b**) All kinds of reviewed rotating machines (*orange*)

studies regarding (a) gas turbine engines only and (b) all types of reviewed rotating machines.

According to the reviewed articles, the RUL estimate of gas turbine engines is the main application field. However, the proportion of studies using simulation data is higher than those using field data. The reason for this is the simplicity of using simulation programs and the difficulty of obtaining sufficient field data from operating machines.

4 Discussions on Reliability Analysis of System with Multiple Failure Modes

Most existing prognostic techniques were originally developed for a single failure mode. To predict RUL for systems with multiple failure modes, several models must be separately constructed for each failure mode. For example, Daigle et al. [6] developed a distributed method for failure prediction of a four-wheel rover. This method first decomposes the system-level prognostic problem into independent sub-failure problems through structural model decomposition. Thereafter, the Kalman filter and physical model are used to perform individual failure prognostics. Finally, the local prognostic results are merged to form a system-level result. However, the correlation between different failures may be overlooked via this approach. To solve this problem, several frameworks for reliability analysis in general engineering systems with competitive multiple failures have been proposed. Ahmad et al. [66] developed a failure analysis approach by integrating the failure mode effect and criticality analysis (FMECA) and PHM. This method was validated in a cutting process system with two failure modes. FMECA was applied to classify the censored and uncensored data based on the severity of the different failure modes. The FMECA output was then used in PHM to determine the

reliability of the system. Huang and Askin [67] proposed a method to analyze the reliability of an electronic device with two types of competing failure modes. Based on the competing failure rule, the mean time-to-failure of the device was estimated by jointly considering the failure rate of both failure modes. Bichon et al. [68] proposed a surrogate-based approach based on the Gaussian process model and physical laws. This method was used to analyze the failure probability of a liquid hydrogen tank with three failure modes. However, most of the frameworks discussed above have not yet been used to predict the RUL for rotating machinery. Therefore, efforts can be made to extend them to rotating machine prognostics.

5 Conclusions

This paper has explored prognostic models for predicting the remaining useful life of rotating machines at the system level. The reviewed prognostic models make predictions based on the multi-dimensional condition monitoring signals collected from the sensors distributed over the studied system. The relevant theories were discussed, and the advantages and disadvantages of the main prognostic model classes were explored. Examples were given to explain how these approaches have been applied to predict the RUL of rotating systems. The reviewed approaches generally require a large amount of historical data (condition monitoring data or lifetime data) to obtain accurate estimates. In addition, the implementation of the reviewed models in industry is still in the nascent stage and more work should be conducted to apply them in real-world operating machines. Moreover, most of the reviewed techniques were originally designed for a signal failure mode. Therefore, several frameworks for reliability analysis of general engineering systems with multiple failure modes were examined. In the future, more work should be conducted to apply these frameworks to rotating machinery prognostics.

References

1. Veldman J, Klingenberg W, Wortmann H (2011) Managing condition-based maintenance technology. J Qual Maint Eng 17(1):40–62
2. Peng Y, Dong M, Zuo MJ (2010) Current status of machine prognostics in condition-based maintenance: a review. Int J Adv Manuf Technol 50(1–4):297–313
3. Tran VT, Thom Pham H, Yang B-S, Tien Nguyen T (2012) Machine performance degradation assessment and remaining useful life prediction using proportional hazard model and support vector machine. Mech Syst Signal Process 32:320–330
4. Sun J, Zuo H, Wang W, Pecht MG (2012) Application of a state space modeling technique to system prognostics based on a health index for condition-based maintenance. Mech Syst Signal Process 28:585–596
5. Wang W, Christer AH (2000) Towards a general condition based maintenance model for a stochastic dynamic system. J Oper Res Soc 51(2):145–155

6. Daigle M, Bregon A, Roychoudhury I (2012) A distributed approach to system-level prognostics. PHM Society, Minneapolis, MN
7. Heng A, Zhang S, Tan ACC, Mathew J (2009) Rotating machinery prognostics: state of the art, challenges and opportunities. Mech Syst Signal Process 23(3):724–739
8. Kan MS, Tan ACC, Mathew J (2015) A review on prognostic techniques for non-stationary and non-linear rotating systems. Mech Syst Signal Process 31:1–20
9. Lee J, Wu F, Zhao W, Ghaffari M, Liao L, Siegel D (2014) Prognostics and health management design for rotary machinery systems – reviews, methodology and applications. Mech Syst Signal Process 42(1–2):314–334
10. Si X-S, Wang W, Hu C-H, Zhou D-H (2011) Remaining useful life estimation – a review on the statistical data driven approaches. Eur J Oper Res 213(1):1–14
11. Sikorska JZ, Hodkiewicz M, Ma L (2011) Prognostic modelling options for remaining useful life estimation by industry. Mech Syst Signal Process 25(5):1803–1836
12. Zhang L, Liu Z, Luo D, Li J, Huang HZ (2013a). Review of remaining useful life prediction using support vector machine for engineering assets. In: Proceedings of the 2013 international conference on quality, reliability, risk, maintenance, and safety engineering (QR2MSE), Chengdu, China, IEEE
13. Zhang Z, Wang Y, Wang K (2013b) Fault diagnosis and prognosis using wavelet packet decomposition, fourier transform and artificial neural network. J Intell Manuf 24 (6):1213–1227
14. Noah ST, Sundararajan P (1995) Significance of considering Nonlinear effects in predicting the dynamic behavior of rotating machinery. J Vib Control 1(4):431–458
15. Bachschmid N, Chatterton S (2014) Dynamical behavior of rotating machinery in non-stationary conditions: simulation and experimental results. Springer, Berlin
16. Bartelmus W, Chaari F, Zimroz R, Haddar M (2010) Modelling of gearbox dynamics under time-varying nonstationary load for distributed fault detection and diagnosis. Eur J Mech A Solids 29(4):637–646
17. Wang W (2007a) A prognosis model for wear prediction based on oil-based monitoring. J Oper Res Soc 58(7):887–893
18. Wang W (2000) A model to determine the optimal critical level and the monitoring intervals in condition-based maintenance. Int J Prod Res 38(6):1425–1436
19. Coble J, Hines JW (2011) Applying the general path model to estimation of remaining useful life. Int J Progn Health Manag 2:1–13
20. Wang W, Zhang W (2005) A model to predict the residual life of aircraft engines based upon oil analysis data. Nav Res Logist 52(3):276–284
21. Wang W, Hussin B (2009) Plant residual time modelling based on observed variables in oil samples. J Oper Res Soc 60(6):789–796
22. Carpenter J, Clifford P, Fearnhead P (1999) Improved particle filter for nonlinear problems. IEEE Proc Radar Sonar Navig 146(1):2
23. Chen C, Zhang B, Vachtsevanos G (2012) Prediction of machine health condition using neuro-fuzzy and Bayesian algorithms. IEEE Trans Instrum Meas 61(2):297–306
24. Orchard M, Wu B, Vachtsevanos G (2005) A particle filtering framework for failure prognosis. American Society of Mechanical Engineers, Washington
25. Wang W, Hussin B, Jefferis T (2012a) A case study of condition based maintenance modelling based upon the oil analysis data of marine diesel engines using stochastic filtering. Int J Prod Econ 136(1):84–92
26. Wang W (2007b) A two-stage prognosis model in condition based maintenance. Eur J Oper Res 182(3):1177–1187
27. Sun Y, Ma L, Mathew J, Wang W, Zhang S (2006) Mechanical systems hazard estimation using condition monitoring. Mech Syst Signal Process 20(5):1189–1201
28. Zhang Q, Hua C, Xu G (2014) A mixture Weibull proportional hazard model for mechanical system failure prediction utilising lifetime and monitoring data. Mech Syst Signal Process 43 (1–2):103–112

29. Jardine AKS, Banjevic D, Wiseman M, Buck S, Joseph T (2001) Optimizing a mine haul truck wheel motors' condition monitoring program use of proportional hazards modeling. J Qual Maint Eng 7(4):286–302
30. Cox RD (1972) Regression models and life tables (with discussion). J R Stat Soc 34:187–220
31. Bendell A (1985) Proportional hazards modelling in reliability assessment. Reliab Eng 11 (3):175–183
32. Li Z, Zhou S, Choubey S, Sievenpiper C (2007) Failure event prediction using the Cox proportional hazard model driven by frequent failure signatures. IIE Trans 39(3):303–315
33. Jardine AKS, Anderson PM, Mann DS (1987) Application of the weibull proportional hazards model to aircraft and marine engine failure data. Qual Reliab Eng Int 3(2):77–82
34. Tsang AHC, Yeung WK, Jardine AKS, Leung BPK (2006) Data management for CBM optimization. J Qual Maint Eng 12(1):37–51
35. Rafiq MY, Bugmann G, Easterbrook DJ (2001) Neural network design for engineering applications. Comput Struct 79(17):541–552
36. Rodríguez JA, Hamzaoui YE, Hernández JA, García JC, Flores JE, Tejeda AL (2013) The use of artificial neural network (ANN) for modeling the useful life of the failure assessment in blades of steam turbines. Eng Fail Anal 35:562–575
37. Atiya AF, El-Shoura SM, Shaheen SI, El-Sherif MS (1999) A comparison between neural-network forecasting techniques-case study: river flow forecasting. IEEE Trans Neural Netw 10 (2):402–409
38. Gençay R, Liu T (1997) Nonlinear modelling and prediction with feedforward and recurrent networks. Phys D 108(1–2):119–134
39. Ahmadzadeh F, Lundberg J (2013) Remaining useful life prediction of grinding mill liners using an artificial neural network. Miner Eng 53:1–8
40. Liu J, Djurdjanovic D, Ni J, Casoetto N, Lee J (2007) Similarity based method for manufacturing process performance prediction and diagnosis. Comput Ind 58(6):558–566
41. Senjyu T, Takara H, Uezato K, Funabashi T (2002) One-hour-ahead load forecasting using neural network. IEEE Trans Power Syst 17(1):113–118
42. Wang S (2003) Application of self-organising maps for data mining with incomplete data sets. Neural Comput Appl 12(1):42–48
43. Zhang S, Ganesan R (1997) Multivariable trend analysis using neural networks for intelligent diagnostics of rotating machinery. J Eng Gas Turbines Power 119(2):378–384
44. Xu J, Wang Y, Xu L (2014) PHM-oriented integrated fusion prognostics for aircraft engines based on sensor data. Sensors J 14(4):1124–1132
45. Peng Y, Wang H, Wang J, Liu D, Peng X (2012) A modified echo state network based remaining useful life estimation approach. IEEE, Denver
46. Peel L (2008) Data driven prognostics using a Kalman filter ensemble of neural network models. IEEE, Denver
47. Felix OH (2008) Recurrent neural networks for remaining useful life estimation. IEEE, Denvor
48. Olden JD, Jackson DA (2002) Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. Ecol Model 154 (1–2):135–150
49. Sussillo D, Barak O (2013) Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. Neural Comput 25(3):626–649
50. Saha B, Goebel K, Christophersen J (2009) Comparison of prognostic algorithms for estimating remaining useful life of batteries. Trans Inst Meas Control 31(3–4):293–308
51. Huang H-Z, Wang H-K, Li Y-F, Zhang L, Liu Z (2015) Support vector machine based estimation of remaining useful life: Current research status and future trends. J Mech Sci Technol 29(1):151–163
52. Hu J, Tse P (2013) A relevance vector machine-based approach with application to oil sand pump prognostics. Sensors 13(9):12663–12686

53. García Nieto PJ, García-Gonzalo E, Sánchez Lasheras F, de Cos Juez FJ (2015) Hybrid PSO–SVM-based method for forecasting of the remaining useful life for aircraft engines and evaluation of its reliability. Reliab Eng Syst Saf 138:219–231
54. Li X, Qian J, Wang G (2013) Fault prognostic based on hybrid method of state judgment and regression. Adv Mech Eng 5(0):149562–149562
55. Liao L, Kottig F (2014) Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. IEEE Trans Reliab 63(1):191–207
56. Wang T, Yu J, Siegel D, Lee J (2008) A similarity-based prognostics approach for remaining useful life estimation of engineered systems. IEEE, Denver
57. Xue F, Bonissone P, Varma A, Yan W, Eklund N, Goebel K (2008) An instance-based method for remaining useful life estimation for aircraft engines. J Fail Anal Prev 8(2):199–206
58. Baurle RA, Gaffney RL (2008) Extraction of one-dimensional flow properties from multidimensional data sets. J Propuls Power 24(4):704–714
59. Malinowski S, Chebel-Morello B, Zerhouni N (2015) Remaining useful life estimation based on discriminating shapelet extraction. Reliab Eng Syst Saf 142:279–288
60. Ramasso E, Gouriveau R (2014) Remaining useful life estimation by classification of predictions based on a neuro-fuzzy system and theory of belief functions. IEEE Trans Reliab 63 (2):555–566
61. Zhang Q, Tse PW-T, Wan X, Xu G (2015) Remaining useful life estimation for mechanical systems based on similarity of phase space trajectory. Expert Syst Appl 42(5):2353–2360
62. Mosallam A, Medjaher K, Zerhouni N (2014) Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction. J Intell Manuf 27 (5):1037–1048
63. Ramasso E, Rombaut M, Zerhouni N (2013) Joint prediction of continuous and discrete states in time-series based on belief functions. IEEE Trans Cybern 43(1):37–50
64. Wang P, Youn BD, Hu C (2012b) A generic probabilistic framework for structural health prognostics and uncertainty management. Mech Syst Signal Process 28:622–637
65. Hu C, Youn BD, Wang P, Taek Yoon J (2012) Ensemble of data-driven prognostic algorithms for robust prediction of remaining useful life. Reliab Eng Syst Saf 103:120–135
66. Ahmad R, Kamaruddin S, Azid IA, Almanar IP (2012) Failure analysis of machinery component by considering external factors and multiple failure modes–a case study in the processing industry. Eng Fail Anal 25:182–192
67. Huang W, Askin RG (2003) Reliability analysis of electronic devices with multiple competing failure modes involving performance aging degradation. Qual Reliab Eng Int 19(3):241–254
68. Bichon BJ, McFarland JM, Mahadevan S (2011) Efficient surrogate models for reliability analysis of systems with multiple failure modes. Reliab Eng Syst Saf 96(10):1386–1395

Study on the Poisson's Ratio of Solid Rocket Motor by the Visual Non-Contact Measurement Teleoperation

Yu-Biao Li, Hai-Bin Li, and Yang-tian Li

Abstract Measurement of Poisson's ratio play an important role in the parametric and structure analysis of viscoelastic material. In order to accurately determine the Poisson's ratio of a solid rocket motor grain, the integral expression of Poisson's ratio of viscoelastic material in time domain is firstly derived using Laplace transform, inverse Laplace transform and inheritance integration. The measure of the relaxation (creep) modulus of the viscoelastic material specimen is performed by using the dynamic mechanical thermal analyser, and 3D visual non-contact method is used to measure the strain. Experiment data obtained from these tests are then substituted into the theoretical formula derived from the integral expression. Time varying curve of a certain type viscoelastic material Poisson's ratio and the Prony series form to facilitate further processing are obtained using MATLAB software. The results are demonstrated consistent with results of traditional method. This method provides an effective way for high accuracy measurement of Poisson's ratio of viscoelastic materials.

Keywords Poisson's ratio • Viscoelastic • Vision metrology • Relaxation (Creep) • Grain • Non-contact measurement

1 Introduction

As one of the basic parameters of materials, the measurement of Poisson's ratio has a very important significance for the performance analysis of materials. Poisson's ratio is the ratio of the material to the axial strain and the transverse strain of the material under uniaxial tension or compression. In the case of the linear elastic material, the ratio of the transverse strain and the longitudinal strain is constant, so the Poisson's ratio is constant. But in the viscoelastic material, Poisson's ratio is a parameter related to the time, frequency and temperature [1, 2]. The measurement of Poisson's ratio of viscoelastic material cannot use simple mechanics of elasticity to characterize the relationship between modulus and strain [3]. The results

Y.-B. Li • H.-B. Li (✉) • Y.-t. Li
College of Science, Inner Mongolia University of Technology, Hohhot 010050, PR China
e-mail: lhbnm2002@163.com

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_12

obtained by this method are complex and require a large amount of computation. The influence of contact measurement on viscoelastic material is very large, for example, the data obtained by the contact method such as strain gauge are not consistent with the actual data [4]. Compared with the contact measurement, non-contact measurement has the advantages of high precision and little influence on the performance of the material [5–7].

In this paper, the method of non-contact measurement is used to analyse the displacement of the specimen after speckle processing [8–10]. The calculated Poisson's ratio of the transverse strain and relaxation modulus is obtained from the experimental measurements [11, 12].

2 The Theoretical Analysis

The mechanical behaviour of viscoelastic materials is between the elastic Hooke's law and the viscous Newton theory. At present, most of the mechanical behaviour of viscoelastic are obtained by uniaxial tension and shear tests. The creep, stress relaxation (hereafter referred to as relaxation) and the dynamic response of the sine change are common [13].

2.1 Basic Theory of Viscoelasticity

Under the action of constant load (or stress), the process of strain increasing with time is called creep [14]. And the ratio of strain to constant stress is called creep compliance, which is denoted as D(t) and given as follow.

$$D(t) = \frac{\varepsilon(t)}{\sigma_0} \tag{1}$$

The process of stress relaxation under constant strain is called stress relaxation [14]. And the ratio of stress and constant strain is called relaxation modulus, which is denoted as E(t) and given as follow.

$$E(t) = \frac{\sigma(t)}{\varepsilon_0} \tag{2}$$

The above Eqs. (1) and (2) can be expressed by the following step function H(t).

$$H(t) = \begin{cases} 1, & (t \geq 0) \\ 0, & (t < 0) \end{cases} \tag{3}$$

For example, creep, $\sigma(t) = \sigma_0 H(t)$, using Boltzmann superposition principle, we can obtain,

$$\varepsilon(t) = \sigma_0 D(t) + \int_0^t D(t-\tau)\frac{\partial \sigma(t)}{\partial \tau}d\tau \tag{4}$$

The same relaxation can be obtained as follow

$$\sigma(t) = \varepsilon_0 E(t) + \int_0^t E(t-\tau)\frac{\partial \varepsilon(t)}{\partial \tau}d\tau \tag{5}$$

2.2 Poisson's Ratio in Time Domain

Poisson's ratio of viscoelastic materials need to be applied in the solution of the corresponding theory, including differential and integral. The mechanical model and the loading history determine the Poisson's ratio. Different forms of differential expressions often lead to different type, which cannot be unified, and the inverse Laplace calculation is complex in this method. However, the integral type has a great advantage [15]. In this paper, the integral type is used to analyse and solve [16, 17].

Poisson's ratio in the time domain can be expressed as,

$$\mu(t) = -\frac{\varepsilon_x(t)}{\varepsilon_y(t)} \tag{6}$$

where, $\varepsilon_x(t)$ is the transverse strain, and $\varepsilon_x(t)$ is the longitudinal strain.

In the case of the viscoelastic material in the longitudinal direction, the response lags behind the longitudinal deformation history. Therefore, the ratio of transverse longitudinal strain cannot be directly used to calculate the Poisson's ratio of viscoelastic material, and the corresponding calculation formula is obtained in the Laplace domain [15]. Poisson's ratio expression in Laplace domain can be converted to the formula as follow.

$$\overline{\mu}(s) = -\frac{\overline{\varepsilon_x}(s)}{\overline{\varepsilon_y}(s)} \tag{7}$$

When the creep test is carried out, we can have

$$\overline{\sigma_y}(s) = \frac{\sigma_0}{s} \tag{8}$$

$$\overline{\varepsilon_y}(s) = \frac{\overline{\sigma}_y(s)}{\overline{E}(s)} \tag{9}$$

The Eqs. (8) and (9) are substituted into (7), and we can obtain

$$\overline{\mu}(s) = -\frac{s}{\sigma_0}\overline{\varepsilon_x}(s)\bar{E}(s) \tag{10}$$

By using the convolution theorem, the formula (10) is used to obtain the Poisson's ratio in time domain as follow [15],

$$\mu(t) = -\frac{1}{\sigma_0}\left[\int_0^t E(t-\tau)\frac{\partial \varepsilon_x(t)}{\partial \tau}d\tau + E(t)\varepsilon_x(0)\right] \tag{11}$$

where, E(t) is the relaxation modulus, and $\varepsilon_x(t)$ is the creep strain.

Using the formula (11) as the theoretical basis, the relevant experimental test is carried out in this paper.

3 Test

3.1 Test Samples and Equipment

Samples are made from a certain model of solid rocket motor grain. According to the special requirements of the test bench, the tensile sample is processed with height 50.00 mm, width 11.00 mm, and thickness 3.60 mm. In order to ensure the universality of the experiment, a total of 5 samples were prepared, and the specific size of the sample is shown in Table 1.

In order to meet the accuracy requirements, sample processing is completed in a high temperature environment (50 °C) for 24 h, eliminating the residual stress generated during processing.

The EPLEXOR dynamic mechanical thermal analyser and the VIC-3D visual test bench produced by the American Solutions Correlated company are used in the experiment. The dynamic mechanical thermal analysis instrument is used to measure the tensile relaxation modulus and shear creep compliance. VIC-3D is used to collect the transverse strain of the sample in the test, as shown in Fig. 1.

For the measurement error, measuring accuracy of dynamic mechanical thermal analyser GOBA is +0.1%; the accuracy of VIC-3D is consistent with the 0.5 level

Table 1 Sample size

Sample No.	Height (mm)	Width (mm)	Thickness (mm)
1	50.60	11.40	3.62
2	51.40	10.70	3.60
3	49.82	10.92	3.56
4	47.36	10.78	3.58
5	50.24	10.02	3.64

Fig. 1 Test equipment

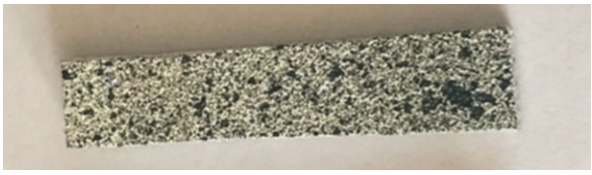


Fig. 2 Tensile test specimen

standard. Both kinds of test data obtained in the accuracy of more than six after the decimal point, fully meet the accuracy required by the test.

When using the VIC-3D, visual bench displacement of the sample is measured, and diagram speckle is sampled and shown in Fig. 2.

3.2 Test Procedure

In order to obtain the corresponding parameters in the formula (11), a total of two sets of tests have been done. Each group has been repeated five times under the same conditions using different samples. The results obtained from the test with abnormal data removed are as the final test result, and then follow-up calculation.

The two sets of tests are tensile relaxation and tensile creep, as shown in Fig. 3. During the test, the test process is kept at a constant temperature of 27 °C. In the process of the test, the temperature change is controlled in the range of 0.5 °C to meet the accuracy requirements.

In the relaxation test, each test time is set to 104 s, and the relaxation modulus is obtained. The test results are shown in Fig. 4.

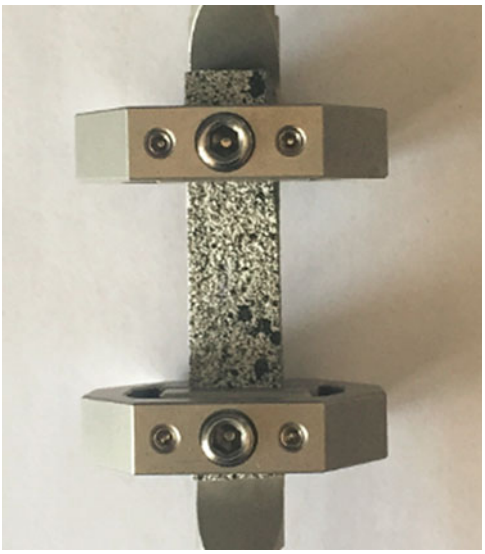


Fig. 3 Tensile test

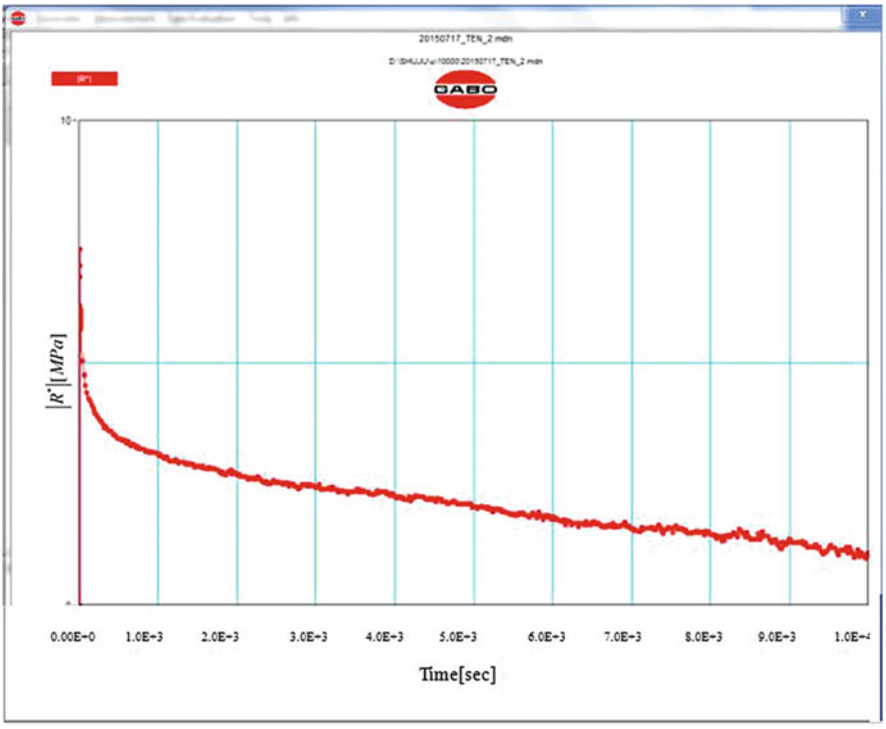


Fig. 4 E(t) test result

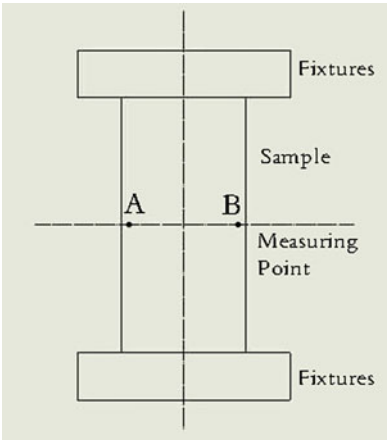


Fig. 5 Measurement chart

The creep test of each test time is 104 s. The VIC-3D measure displacements of Point A and Point B under two coordinates with time. The positions of Point A and Point B are shown in Fig. 5.

The Point A in the test process in the horizontal coordinates of the record is denoted as I_a(t). The Point B of the horizontal axis is denoted as I_b(t). The initial time point displacement are recorded as $I_a(0)$ and $I_b(0)$. The I_a(t) and I_b(t) get displacement data as shown in Fig. 6a, b.

In the experiment, the visual test bench is only on the displacement of two points. Therefore, the strain of the test process can be obtained by a simple calculation, and the formula is as follows,

$$l(t) = |\, l_a(t)| + I_b(t) \tag{12}$$

$$l(0) = |\, l_a(0)| + I_b(0) \tag{13}$$

$$\varepsilon_x(t) = \frac{|l(t) - l(0)|}{l(0)} \times 100\% \tag{14}$$

The transverse strain ε_x(t) is calculated by the formula (13), as shown in Fig. 7, which is further in the Eq. (11) for Poisson's ratio calculation.

3.3 Data Processing and Test Results

Through the above two experiments, the relaxation modulus and the transverse strain in creep are obtained respectively. In order to improve the accuracy of the test, the test abnormal data need to be excluded. The application of Grubbs criterion in the choice (set reliability $\alpha = 95\%$) can delete abnormal data [18].

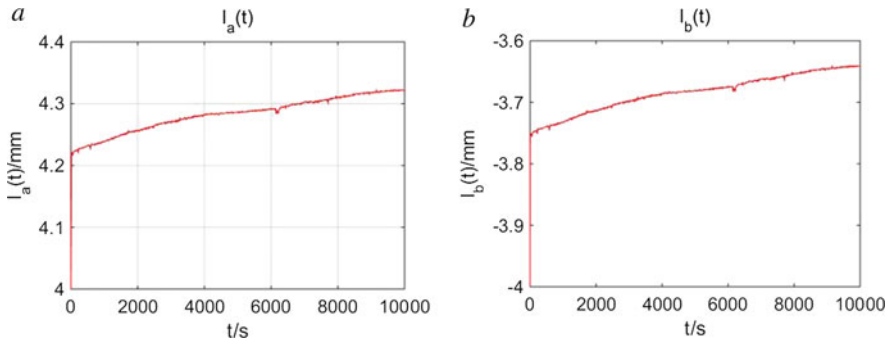


Fig. 6 (**a**) $I_a(t)$ displacement; (**b**) $I_b(t)$ displacement

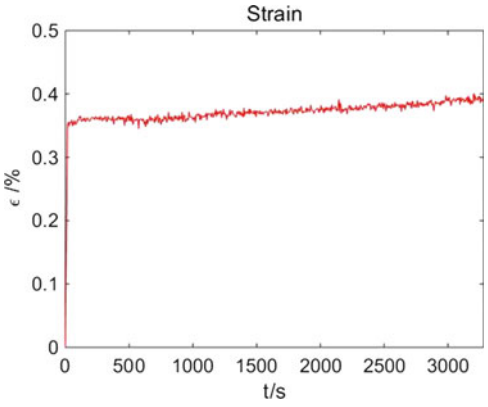


Fig. 7 $\varepsilon_x(t)$ calculation results

In order to obtain the stable relaxation modulus and lateral strain, the five sets of data are calculated using MATLAB, and the corresponding average data are obtained.

The relaxation modulus and transverse strain are measured by Prony series fitting using MATLAB platform. In order to ensure that there is sufficient fitting precision, the Prony series are set to the nine class number.

In the fitting of the data in the MATLAB, in order to ensure the parameters have good accuracy in time domain on the whole, the nine coefficients are in different number of basis coefficients, such as τ_i in Table 2. This coefficient setting can not only ensure the accuracy of the initial segment, but also can guarantee the accuracy of small amplitude change in the long time course.

It is found that the instability of the initial transient load has great influence on the acquisition of the experimental data and the result of the initial time. Therefore, there is an abnormal in the data measured at the initial time of the test. The [0,1]s interval segment approximation curve of the data is fitted by the least square fitting before the test time in 1s.

Table 2 E(t) fitting results

i	a_i	τ_i
0	0.558	
1	0.03183	10^{-3}
2	0.2769	10^{-2}
3	0.04617	10^{-1}
4	0.1154	10^{0}
5	0.5113	10^{1}
6	0.5404	10^{2}
7	0.1912	10^{3}
8	0.2681	10^{4}
9	$2.22e^{-14}$	10^{5}

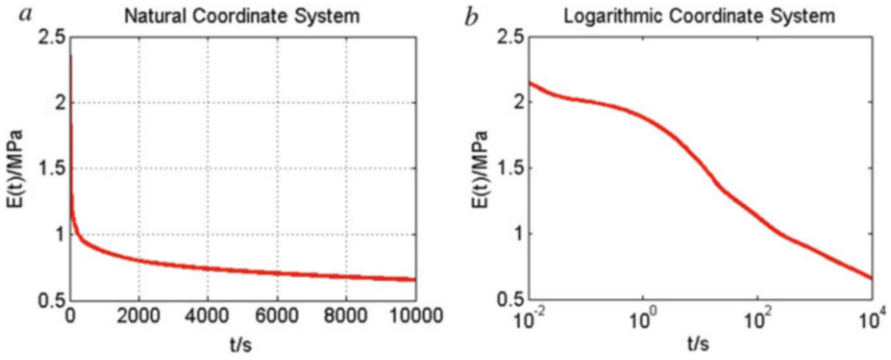


Fig. 8 (**a**) E(t) in natural coordinate system; (**b**) E(t) in logarithmic coordinate system

After the process of experimental data obtained, Prony series fitting. Prony series fitting formula for relaxation modulus is given as follow,

$$E(t) = a_0 + \sum_{i=1}^{9} a_i \exp\left(-\frac{t}{\tau_i}\right) \tag{15}$$

The coefficients in the formula are shown in Table 2. The fitting results are shown in Fig. 8a, b. Due to the selection of coordinate system, the date changes significantly in the front part in natural coordinates, and the date changes significantly in the rear part in logarithmic coordinates. Therefore, the obtained data are used in two coordinates expression, which are separately presented in figures of the natural coordinates system and after logarithmic coordinate system.

In the same way, the coefficients of the Prony series of $\varepsilon_x(t)$ are shown in Table 3. The fitting results are shown in Fig. 9a, b. The Prony series fitting formula for $\varepsilon_x(t)$ is given as,

$$\varepsilon_x(t) = b_0 + \sum_{i=1}^{9} b_i \exp\left(-\frac{t}{\varphi_i}\right) \tag{16}$$

Table 3 $\varepsilon_x(t)$ fitting results

i	b_i	φ_i
0	0.1284	
1	−0.0104	10^{-3}
2	−0.0104	10^{-2}
3	−0.008851	10^{-1}
4	−0.0116	10^{0}
5	−0.01375	10^{1}
6	−0.006363	10^{2}
7	−0.001113	10^{3}
8	$-2.406e^{-08}$	10^{4}
9	−0.0001876	10^{5}

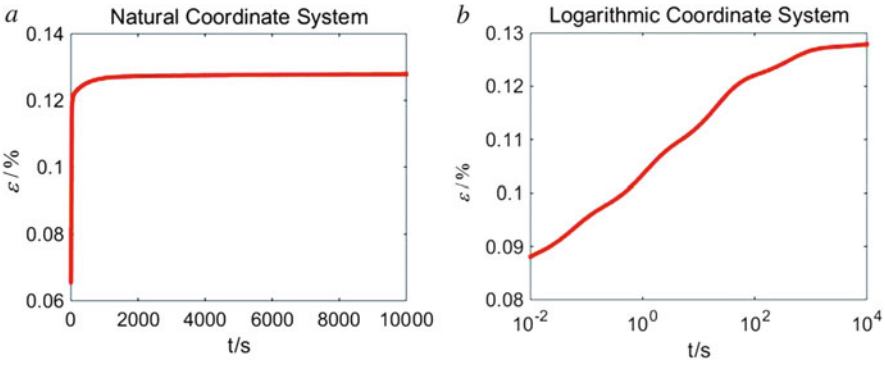


Fig. 9 (**a**) $\varepsilon_x(t)$ in natural coordinate system; (**b**) $\varepsilon_x(t)$ in logarithmic coordinate system

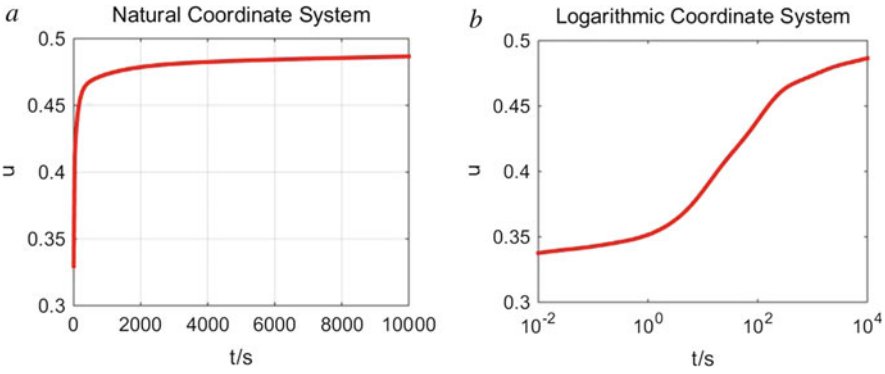


Fig. 10 (**a**) μ(t) in natural coordinate system; (**b**) μ(t) in logarithmic coordinate system

The above obtained two fitting results by MATLAB are substituted into the Eq. (11) with the Poisson like Fig. 10a, b. The calculated Poisson's ratio function is fitted to Prony series, and the results are shown in Table 4.

Table 4 μ(t) fitting results

i	c_i	ω_i
0	0.4997	
1	−0.004561	10^{-3}
2	−0.004561	10^{-2}
3	−0.004561	10^{-1}
4	−0.004592	10^{0}
5	−0.04872	10^{1}
6	−0.06418	10^{2}
7	−0.01812	10^{3}
8	−0.0105	10^{4}
9	−0.01022	10^{5}

$$\mu(t) = c_0 + \sum_{i=1}^{9} c_i \exp\left(-\frac{t}{\omega_i}\right) \tag{17}$$

4 Conclusion

In this paper, the non-contact measurement method is applied to the determination of Poisson's ratio of the viscoelastic material. Compared with contact measurement non-contact measurement method to obtain accurate and simple measurement of the accuracy of a good guarantee. When the contact test is used, the contact force generated by the con-tact measurement will have influence on the properties of the viscoelastic proper-ties [4], cause errors of the resulting Poisson's ratio data. The Poisson's ratio is assumed to be constant in the viscoelastic analysis of the grain, which is generally about 0.49 [19, 20]. From the results of this paper, we can see that the Poisson's ratio is the Variables with time, so the results obtained in this paper are more in line with the objective reality, avoiding the error caused by the Poisson's ratio in the structural analysis.

During the experiment, the transient load is applied to the ideal state, which cannot be achieved in practice. Therefore, the experimental data in the initial transient instability; and this instability has been to continue 1s, so the curve [0,1] s data with the least squares fitting.

Because the maximum time course of this paper is 104, it can be considered that the data in 104 is measured, and the data is fitted and extrapolated after 104. In fact, the range of the data after 104 is negligible, so the subsequent estimation data is available.

Due to the limitation of the working environment of the vision measuring instrument, the measurement under the condition of high and low temperature cannot be realized, so the law of Poisson's ratio with the temperature change cannot be obtained [21]. Whether the Poisson's ratio of viscoelastic material still exists the temperature equivalence principle needs to be further studied.

Acknowledgements This work was supported by the National Natural Science Foundation of China under the contract number 11262014.

References

1. Hilton HH, Yi S (1998) Significance of (an)isotropic viscoelastic Poisson ratio stress and time dependencies. Int J Solids Struct 35(23):3081–3095
2. Lee HS, Kim JJ (2009) Determination of viscoelastic poisson's ratio and creep compliance from the indirect tension test. J Mater Civil Eng 21(8):416–425
3. Lakes RS, Wineman A (2006) On Poisson's ratio in linearly viscoelastic solids. J Elast 85 (1):45–63
4. Zhang JB, Ju YT, Meng HL et al (2012) Research on measurement method of surface of double base propellant grain. J Ball 24(2):62–65
5. Aushev AA, Barinov SP, Vasin MG et al (2015) Alpha-spectrometry and fractal analysis of surface micro-images for characterisation of porous materials used in manufacture of targets for laser plasma experiments. Quantum Electron 45(6):533–539
6. Kabeer S, Attenburrow G, Picton P et al (2013) Development of an image analysis technique for measurement of Poisson's ratio for viscoelastic materials: application to leather. J Mater Sci 48(2):744–749
7. Zhou GD, Li HN, Ren L et al (2007) Study on influencing parameters of strain transfer of optic fiber Bragg grating sensors. Eng Mech 24(6):169–173
8. Edmundson K, FraserC S (1998) A practical evaluation of sequential estimation for vision metrology. ISPRS J Photogramm 53(5):272–285
9. Sun W, He XY, Xu M et al (2007) Study on the tension test of membrane materials using digital image correlation method. Eng Mech 24(2):34–38
10. Sun W, He XY, Huang YP et al (2008) Experimental study on identification of modal parameters of cable. Eng Mech 25(6):88–93
11. Kassem E, Grasley Z, Masad E (2013) Viscoelastic Poisson's ratio of asphalt mixtures. Int J Geomech 13(2):162–169
12. Keramat A, Kolahi AG, Ahmadi A (2013) Waterhammer modelling of viscoelastic pipes with a time-dependent Poisson's ratio. J Fluid Struct 43:164–178
13. Klompen ETJ, Govaert LE (1999) Nonlinear viscoelastic behaviour of thermorheologically complex materials. Mech Time Dep Mater 3(1):49–69
14. Yang TQ, Luo WB, Xu P et al (2004) Viscoelastic theory and application. Science Press, Hubei
15. Zhao BH (1995) An investigation on viscoelastic Poisson's ratio and dynamic complex Possion's ratio. J Prop Tech 3:1–7
16. Fu YM, Li PE, Zheng YF (2005) Nonlinear dynamic responses of viscoelastic symmetrically laminated plates. Eng Mech 22(4):24–30
17. Zheng JL, Lu ST, Tian XG (2008a) Viscoelastic damage characteristics of asphalt based on creep test. Eng Mech 25(2):193–196
18. Shalom BY, Li XR, Kirubarajan T (2001) Estimation with application to tracking and navigation. Wiley, New York
19. Tian SP, Lei YJ, Li DK et al (2007) Incremental viscoelastic three-dimensional fem based on herrmann variational principle and its application. Eng Mech 24(7):28–32
20. Zhang HL, Zhou JP (2001) Viscoelastic stochastic finite element simulation of solid propellant grain with random Poisson's ratio. J Prop Tech 22(3):245–249
21. Zheng JL, Guoping Q, Ronghua Y (2008b) Testing thermal viscoelastic constitutive relation of asphalt mixtures and its mechanical applications. Eng Mech 25(1):34–41

Reliability Allocation of Multi-Function Integrated Transmission System Based on the Improved AGREE Method

Qi-hai Liang, Hai-ping Dong, Xiao-jian Yi, Bin Qin, Xiao-yu Yang, and Peng Hou

Abstract To solve the reliability allocation problem of multi-function integrated transmission system, an improved AGREE allocation method is presented in this paper. Firstly, the reliability index of each function of an integrated transmission system is allocated to its subsystems based on the AGREE method. Then the reliability index allocated to the subsystems is adjusted considering that some subsystems may be included in multiple functions. Then, the reliability index of those non-reused subsystems in different functions is reallocated based on the improved AGREE method. Finally, reliability index of an integrated transmission system considering three functions: straight running, swerve and braking is allocated based on this proposed method, and the allocation results meet the function reliability requirements of the integrated transmission device. The application result shows that the improved AGREE method is feasible for the reliability allocation of a multi-function integrated transmission system and provides a guidance for reliability allocation of such an integrated transmission system and other similar multi-function systems.

Keywords Reliability allocation • Multi-function systems • AGREE method

1 Introduction

An integrated transmission system is a key device that decides the mobility of tanks and armored vehicles, and undertakes the task of transmitting mechanical energy to crawler traveling device, and makes tanks to generate movements [1]. Reliability is an important characteristic, which directly affects the design quality and operation

Q.-h. Liang • H.-p. Dong • B. Qin • X.-y. Yang • P. Hou
Beijing Institute of Technology, Beijing, China
e-mail: 490075514@qq.com; donghaipingphd@126.com; 614714864@qq.com; 15040097206@163.com; apolloar@163.com

X.-j. Yi (✉)
China North Industries Group Corporation, Beijing, China
e-mail: yixiaojianbit@sina.com

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_13

of the integrated transmission system. Reliability allocation contributes to the system reliability design and analysis work. For the integrated transmission system, reliability allocation is also a particularly important work. Due to the integrated transmission system with straight running, swerve, braking and other functions, the widely used reliability allocation methods which work well with the series and parallel systems cannot be directly used for the multi-function integrated transmission system. Guo [2] proposed an allocation method for an integrated transmission system on balance of reliability and maintainability index with the goal of availability, and conducted optimization allocation based on the improved genetic algorithm. However, he ignored the impact of multiple functions on the reliability allocation of the integrated transmission system. Yi et al. [3–5] proposed a method based on genetic algorithm, which used function importance factors to conduct reliability allocation of multi-function integrated transmission system, and solved the problem of multi-function system reliability allocation. But the evaluation of the function importance factors was especially complex. In order to solve the reliability allocation problem of the multi-function integrated transmission system, a reliability allocation method for multi-function integrated transmission system is proposed in this paper. Firstly, the reliability index of each function of an integrated transmission system is allocated to its subsystems based on the AGREE method. Then the reliability index allocated to the subsystems is adjusted considering that some subsystems may be included in multiple functions. Then, the reliability index of those non-reused subsystems in different functions is reallocated based on the improved AGREE method. Finally, reliability index of an integrated transmission system considering three functions: straight running, swerve and braking is allocated based on this proposed method. The result shows that the method proposed in this paper is more practical.

This paper consists of the following sections: In Sect. 2, the improved AGREE method for reliability allocation of a multi-function system is introduced. In Sect. 3, the reliability of a certain type of multi-function integrated transmission system is allocated based on the proposed method. In Sect. 4, the reliability allocation results are analyzed. In Sect. 5, the conclusions are discussed.

2 Reliability Allocation Method Considering Multi-Function States Based on Improved Agree Method

2.1 AGREE Method

Suppose that $R_s^*(t)$ is the reliability index of a system at operating time t. For a series system composed of the subsystems with exponentially distributed lifetimes, it is known that the greater number of basic component in a subsystem, the more complex the subsystem is. Suppose that each basic component in the series system

has the same impact on the reliability of the series system, so based on the AGREE method [6], the allocated reliability $R_i(t)$ of subsystem $i$at time t can be expressed as

$$R_i(t) = \left[\left(R_s^*(t)\right)^{1/N}\right]^{n_i} = \left[R_s^*(t)\right]^{n_i/N} \tag{1}$$

Where, n_i is the number of basic components in subsystem i; $N = \sum_{i=1}^{n} n_i$ is the total number of basic components in the system.

Moreover, the impact of different subsystem on the system reliability is different, and the different impact can be described by importance factor. The importance factor of subsystem i represents the probability that the system will fail if subsystem i fails, and Eq. (1) can be written as

$$\omega_i\left(1 - e^{-\lambda_i t_i}\right) = 1 - \left[R_s^*(t)\right]^{n_i/N} \tag{2}$$

Where, ω_i is the importance factor of subsystem i and λ_i is the failure rate of subsystem i.

When $x \to 0$, using Taylor formula expansion, $e^x \approx 1 + x$, so Eq. (2) can be changed as

$$\left[R_s^*(t)\right]^{n_i/N} = 1 - \omega_i\left(1 - e^{-\lambda_i t_i}\right) \approx 1 - \omega_i \lambda_i t_i \approx e^{-\omega_i \lambda_i t_i} \tag{3}$$

And the failure rate of subsystem i can be expressed as

$$\lambda_i = -\frac{n_i \ln\left[R_s^*(t)\right]}{N \omega_i t_i} \tag{4}$$

The main words in all headings (even run-in headings) begin with a capital letter. Articles, conjunctions and prepositions are the only words which should begin with a lower case letter.

2.2 *The Initial Reliability Allocation for Each Function of a System*

For a simple series and parallel system, it is easy to determine the importance factor of a subsystem, but for a multi-function complex system, the importance factor of a subsystem cannot be easily determined, so the traditional AGREE method can only be applied in the reliability allocation of a simple structure system (e.g., a series system or a series–parallel system), and cannot be directly applied in reliability allocation of a multi-function complex system. To solve this problem, a heuristic algorithm is adopted to calculate the subsystem importance factor in this paper.

(1) Suppose that the initial importance factor of each subsystem is 1.

$$\omega_{k,i} = 1, \quad i = 1, 2, \ldots, m_k \tag{5}$$

Where, m_k is the number of subsystems involved in function k.

(2) Calculate the failure rate coefficient $\alpha_{k,i}$ of subsystem i.

$$a_{k,i} = \frac{n_{k,i}}{N_k \omega_{k,i} t_{k,i}}, \quad i = 1, 2, \ldots, m_k \tag{6}$$

Where, $n_{k,i}$ is the number of basic components within the subsystem i for function k; $N_k = \sum_{i=1}^{n} n_{k,i}$ is the total number of basic components for function k; $t_{k,i}$ is the working time of subsystem i for function k.

(3) Calculate the failure rate and reliability of each subsystem.

$$\lambda_{k,i} = a_{k,i}\beta, \quad i = 1, 2, \ldots, m_k \tag{7}$$

$$R_{k,i}(t) = \exp(-\lambda_{k,i} t_{k,i}) = \exp(-a_{k,i}\beta t_{k,i}) \tag{8}$$

Where β is always positive;

(4) Use Matlab optimization function to solve the following optimization problem.

$$\begin{aligned} \max \quad & \beta \\ \text{s.t.} \quad & R_k^*(t) \le R_k(t) \\ & \beta > 0 \end{aligned} \tag{9}$$

Where, $R_k^*(t)$ is the reliability requirement of function k at time t, $R_k(t)$ is the calculated reliability of function k at time t;

(5) Calculate the new failure rate $\lambda'_{k,i}$ and reliability $R'_{k,i}$ of each subsystem i for function k with β.

$$\lambda'_{k,i} = a_{k,i}\beta, \quad i = 1, 2, \ldots, \mathrm{m}_k \tag{10}$$

$$R'_{k,i}(t) = \exp\left(-\lambda'_{k,i} t_{k,i}\right), \quad i = 1, 2, \ldots, m_k \tag{11}$$

(6) Calculate the new importance factor $\omega'_{k,i}$ of each subsystem i for function k.

$$\omega'_{k,i} = \frac{\partial R_k(t)}{\partial R'_{k,i}(t)}, \quad i = 1, 2, \ldots, m_k \tag{12}$$

Where, $R_k(t) = f(R_{k,1}(t), R_{k,2}(t), \cdots, R_{k,m_k}(t))$ is the calculated reliability of function k.

(7) Compare $R'_{k,i}(t)$ with $R_{k,i}(t)$

If any $\left|R'_{k,i}(t) - R_{k,i}(t)\right| \geq \varepsilon$, then let

$$\begin{cases} \omega_{k,i} = \omega'_{k,i} \\ \lambda_{k,i} = \lambda'_{k,i} \\ R_{k,i}(t) = R'_{k,i}(t) \end{cases}, \quad i = 1, 2, \ldots, m_k \tag{13}$$

And go to step (2); otherwise, stop and $R'_{k,i}(t)$ is the initial allocation result for function k.

In the above allocation process, the constraint $f(R_{k,1}(t), R_{k,2}(t), \cdots, R_{k,m_k}(t)) \geq R_k^*(t)$ will always be met. Under the constraint of system function reliability goal, we aim to determine the most reasonable and allowable reliability for subsystems through reliability allocation because a higher reliability level generally requires a more cost and more difficult design. The purpose of reliability optimization allocation is to find a reasonable allocation result that not only satisfies the requirement of function reliability goal, but also with the minimum cost and the simplest design.

2.3 *Adjust the Reliability Allocation Result for a Multi-Function System*

Because some subsystems not only exist in a function, but also exist in two or more functions, then the same subsystem in different functions may be allocated to different reliability value. So the reliability allocation values of such subsystems need to be adjusted. In order to meet reliability requirements of all functions, we select the strictest reliability allocation value, i.e., $\max\left\{R_{k_1,i_x}(t), R_{k_2,i_x}(t), \ldots, R_{k_j,i_x}(t)\right\}$, where k_1, $k_2, \ldots, k_j$ are functions 1 to j in which the subsystem i_x is involved. For other subsystems, if we keep the initial allocation values unchanged, the reliability of each function will be larger than its reliability requirement, and the more design resources will be needed than what is required. So, after the reliability of subsystems which are reused in multiple functions is determined, the reliability of other subsystems which are not reused will be reallocated in the following section.

2.4 Reliability Reallocation of Subsystems Which Are Not Reused in a Multi-Function System Based on the Improved AGREE Method

For function k, suppose that the number of subsystems reused in multiple functions is p_k, and their reliability values can be expressed as $R^*_{k,1}, R^*_{k,2}, \ldots, R^*_{k,p_k}$, then the reliability of other subsystems which are not reused can be reallocated as follows:

(1) Suppose the initial reliability importance factor of subsystems which are not reused is 1

$$\omega_{k,i} = 1, \quad i = p_{k+1}, p_{k+2}, \ldots, m_k \tag{14}$$

(2) Solve the following optimization problem

$$\begin{array}{ll} \max & \beta \\ R^*_k(t) \le R_k(t) & \\ s.t. & R_{k,1}(t) = R^*_{k,1}(t) \\ & R_{k,2}(t) = R^*_{k,2}(t) \\ & \cdots \\ & R_{k,p_k}(t) = R^*_{k,p_k}(t) \end{array} \tag{15}$$

Where, $R_{k,i}(t) = \exp(-n_{k,i}\beta / N_k \omega_{k,i}) \quad i = p_{k+1}, p_{k+2}, \ldots, m_k$.

(3) Calculate the new failure rate $\lambda'_{k,i}$ and reliability $R'_{k,i}(t)$ using the new β

$$\lambda'_{k,i} = a_{k,i}\beta, \quad i = p_{k+1}, p_{k+2}, \ldots, m_k \tag{16}$$

$$R'_{k,i}(t) = \exp\left(-\lambda'_{k,i} t_{k,i}\right) \tag{17}$$

(4) Calculate the new importance factor of subsystems which are not reused in function k

$$\omega'_{k,i} = \frac{\partial R_k(t)}{\partial R'_{k,i}(t)}, \quad i = p_{k+1}, p_{k+2}, \ldots, m_k \tag{18}$$

(5) Compare $R'_{k,i}(t)$ with $R_{k,i}(t)$

If any $\left|R'_{k,i}(t) - R_{k,i}(t)\right| \ge \varepsilon$, then let

$$\begin{cases} \omega_{k,i} = \omega'_{k,i} \\ \lambda_{k,i} = \lambda'_{k,i} \\ R_{k,i}(t) = R'_{k,i}(t) \end{cases} , i = p_{k+1}, p_{k+2}, \ldots, m_k \tag{19}$$

And go to step (2); otherwise, stop and $R'_{k,i}(t)$ is the final allocation result of each subsystem.

3 The Reliability Allocation of a Certain Type of Integrated Transmission Device

The reliability indexes which are often used in function reliability allocation are Mean Time Between Failures, reliability or failure rate. In this paper, reliability is selected to be allocated for a certain type of integrated transmission device based on the following assumption [7]:

1. There are only two states as failure and success for the system and each subsystem;
2. There are no common cause failures, which mean the failures of different subsystems are independent of each other;
3. Times to failures all follow exponential distributions.

3.1 The System Analysis of an Integrated Transmission Device

Take an integrated transmission device with three main functions as an example. The three main functions are straight running, swerve and braking. The reliability indexes for these functions are assumed as 0.9, 0.8 and 0.7 respectively. There are 19 subsystems in the integrated transmission device, and they are 1-body parts, 2-middle bracket, 3-torque converter assembly, 4-transmission assembly, 5-hydraulic torque converter, 6-planet before shift gear, 7-auxiliary drive, 8-hydraulic gear reducer, 9-hydraulic retarder control valves, 10-left side cover, 11-right side cover, 12-fan drive assembly, 13-liquid viscous clutch assembly, 14-oil pump group, 15-couplet of pump motor, 16-oil supply system, 17-hydraulic control system,18-manipulation of the electronic control system and 19-overall gearing. The system function block diagram of the integrated transmission device is shown as in Fig. 1:

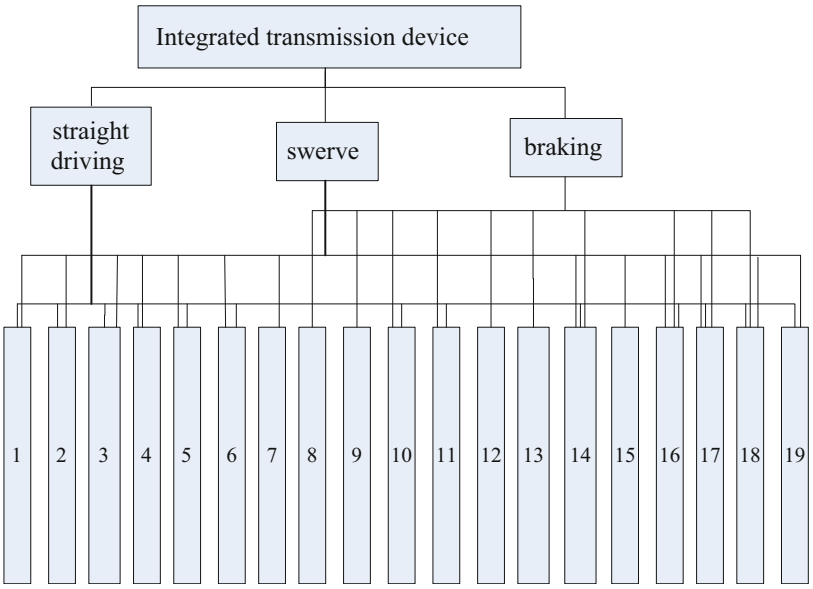


Fig. 1 System function block diagram of the integrated transmission device

3.2 *The Initial Allocation of Reliability for Each Function*

Assume that the function reliability requirements in this system for straight running, swerve and braking at time 1000 h are 0.9, 0.8, 0.7 respectively, and 0.0001 is the allocation accuracy requirement. According to the calculation process in Sect. 2.2, reliability indexes of the three functions can be allocated to each subsystem. The complexity of each subsystem, i.e. the number of basic components involved in each subsystem, and reliability values under different function states are shown in Table 1.

3.3 *The Reliability Adjustment of Reused Subsystems Because of Multiple Functions*

Some subsystems may not only exist in a single function, but in two or more functions, so the same subsystem involved different functions may be allocated different reliability values. For example, the allocated reliability of body parts in straight running function is 0.99623, while its allocated reliability in the steering task is 0.99167. The body parts are not involved in braking function, so the reliability of braking function is not allocated to the body parts. In order to meet the reliability requirements of all functions, we choose the maximum reliability allocation value for such reused subsystems, i.e., body parts' reliability value is

Table 1 Subsystem complexity and initial reliability under different functions

No.	Subsystems	n_i	R/Straight running	R/Swerve	R/Braking
1	Body parts	2	0.99623	0.99167	NA
2	Middle bracket	2	0.99623	0.99167	NA
3	Torque converter assembly	3	0.99436	0.98758	0.98176
4	Transmission assembly	5	0.99064	0.97956	NA
5	Hydraulic Torque converter	6	0.98880	0.97562	NA
6	Planet before shift gear	6	0.98880	0.97562	NA
7	Auxiliary drive	4	NA	0.98355	NA
8	Hydraulic gear Retarder	8	NA	NA	0.95344
9	Hydraulic retarder control valve	3	NA	NA	0.98176
10	Left side cover	3	0.99436	NA	0.98176
11	Right side cover	3	0.99436	NA	0.98176
12	Fan drive assembly	5	NA	NA	0.97014
13	Liquid viscous clutch assembly	8	NA	NA	0.95344
14	Oil pump Group	6	0.98880	0.97562	0.96448
15	Couplet of pump motor	4	NA	NA	0.97590
16	Oil Supply System	6	0.98880	0.97562	0.96448
17	Hydraulic control system	5	0.99064	0.97956	0.97014
18	Manipulation of the electronic control system	5	0.99064	0.97956	0.97014
19	Overall gearing	4	0.99249	0.98355	NA

Notes: NA means that the subsystem is not used in the corresponding function

determined to be maximum, 0.99623. For other not-reused subsystems, if we keep the initially allocated reliability unchanged,, the function reliability will be higher than the required, and more cost and better design are needed. Therefore, after the reliability indexes of the reused subsystems are determined based on the maximum principle, as shown in Table 2, the reliability of those not-reused subsystems should be reallocated to.

3.4 The Reliability Reallocation for Not-Reused Subsystems in the Multi-Function System

The reliability of non-reused subsystems are reallocated to according to the Sect. 2.4, as shown in Table 3.

Table 2 Reliability values of reused subsystems in different functions

No.	Subsystems	Subsystem reliability values
1	Body parts	0.99623
2	Middle bracket	0.99623
3	Torque converter assembly	0.99436
4	Transmission assembly	0.99064
5	Hydraulic Torque converter	0.98880
6	Planet before shift gear	0.98880
7	Left side cover	0.99436
8	Right side cover	0.99436
9	Oil pump group	0.98880
10	Oil supply system	0.98880
11	Hydraulic control system	0.99064
12	Manipulation of the electronic control system	0.99064
13	Overall gearing	0.99249

Table 3 The reliability reallocation values of non-reused subsystems

No.	Subsystems	Subsystems reliability value
1	Auxiliary drive	0.87887
2	Hydraulic gear reducer	0.91958
3	Hydraulic retarder control valve	0.96747
4	Fan drive assembly	0.94745
5	Liquid viscous clutch assembly	0.91958
6	Couplet of pump motor	0.95731

4 Result Comparison

In order to verify the rationality and accuracy of the reliability allocation method proposed in this paper for multi-function systems, take the same integrated transmission device whose reliability index is allocated by the traditional AGREE method as a comparison. The reliability requirement for the three functions: straight running, swerve, braking is still assumed as 0.9, 0.8, 0.7 respectively, and the reliability importance factor of each subsystem is still set to be 1. We choose the maximum allocation value for those subsystems which are reused in different functions, and the reliability allocation results are shown in Table 4.

According to the reliability allocation results of all subsystems in Tables 2 and 3, we can obtain the expected reliability level of three functions: straight running, swerve and braking, is 0.900035, 0.800009, 0.700032, respectively, a little larger than the required. However, we also can obtain the reliability of three functions based on AGREE method from Table 4: straight running, swerve, braking are 0.900091, 0.895405, 0.790525, obviously, larger than the required, especially the last two results. This result shows that the reliability allocation results based on

Table 4 Reliability allocation results of the integrated transmission device with traditional AGREE method

No.	Name	R/Straight running	R/Swerve	R/Braking	Rmax
1	Body parts	0.99624	0.99177	NA	0.99624
2	Middle bracket	0.99624	0.99177	NA	0.99624
3	Torque converter assembly	0.99437	0.98768	0.98203	0.99437
4	Transmission assembly	0.99064	0.97955	NA	0.99064
5	Hydraulic Torque converter	0.98881	0.97551	NA	0.98881
6	Planet before shift gear	0.98881	0.97551	NA	0.98881
7	Auxiliary drive	NA	0.98361	NA	0.98361
8	Hydraulic gear Retarder	NA	NA	0.95279	0.95279
9	Hydraulic retarder control valve	NA	NA	0.98203	0.98203
10	Left side cover	0.99437	NA	0.98203	0.99437
11	Right side cover	0.99437	NA	0.98203	0.99437
12	Fan drive assembly	NA	NA	0.97023	0.97023
13	Liquid viscous clutch assembly	NA	NA	0.95279	0.95279
14	Oil pump Group	0.98881	0.97551	0.96438	0.98881
15	Couplet of pump motor	NA	NA	0.97611	0.97611
16	Oil Supply System	0.98881	0.97551	0.96438	0.98881
17	Hydraulic control system	0.99063	0.97955	0.97023	0.99063
18	Manipulation of the electronic control system	0.99063	0.97955	0.97023	0.99063
19	Overall gearing	0.99250	0.98361	NA	0.99250

Notes: NA means that the subsystem is not used in the corresponding function

improved AGREE method is more accurate and can save more resource and cost under the condition of required reliability.

5 Conclusions

This paper proposes a reliability optimization allocation method based on improved AGREE method for a multi-function integrated transmission device. First, the AGREE method is used to allocate the reliability index of the three functions: straight running, swerve and braking, of an integrated Transmission device to their corresponding subsystems. Then reliability index of those reused subsystems in different functions are adjusted based on the maximum principle. Finally, the reliability of non-reused subsystems are reallocated based on the improved AGREE method. The result shows that the reliability allocation result based on the improved AGREE method can meet the requirements of multi-function integrated transmission device, and it provides an important guidance for reliability allocation for the integrated transmission device and other similar systems.

References

1. Zheng MQ, Feng CZ, Lan ZY (2003) Tank and armored vehicle. Beijing Institute of Technology Press, Beijing
2. Guo SW (2016) Study on the allocation method for the power-shift steering transmission on balance of reliability and maintainability index with the goal of availability. Beijing Institute of Technology, Beijing
3. Yi XJ, Lai YH, Dong HP et al (2015b) A reliability optimization allocation method considering differentiation of functions. Int J Comput Methods 13(4). doi:10.1142/S0219876216410206
4. Nabil N, Mustapha N (2005) Ant system for reliability optimization of a series system with multiple-choice and budget constraints. Reliab Eng Syst Saf 72:1–12
5. Yi XJ, Hou P, Dong HP et al (2015a) A reliability optimization allocation method for systems with differentiation of functions. In: Proceedings of ASME 2015 international mechanical engineering congress and exposition, IMECE 2015-52928
6. Ebeling CE (2009) An introduction to reliability and maintainability engineering. Waveland, Long Grove, IL
7. Li RY, Wang JF, Liao HT et al (2015) A new method for reliability allocation of avionics connected via an airborne network. J Netw Comput Appl 48:14–21

A Study of Sustained Attention Improving by Fuzzy Sets in Supervisory Tasks

Cheng-Li Liu, Ruei-Lung Lai, and Shiaw-Tsyr Uang

Abstract In supervisory task, the operator needs not do any thing over prolonged periods of time, but must keep a certain level of attention to detect more serious signals for confirming the problems of the system. When the activity is to be performed for a continuous period of time, the ability to maintain attention may be severely impaired. The purpose of this study is to develop a quantitative sustained attention measuring model by fuzzy sets, which integrates the concepts of hit, false alarm and response time. Then, an alarm system is developed for improving performance of the sustained attention in supervisory task. The experimental results showed that the effect of the system is good for improving performance of sustained attention.

Keywords Sustained attention • Fuzzy sets • Supervisory task • Signal detection theory • Sensitivity

1 Introduction

As systems are integrated and automated, the role of the human operator has changed from that of an active controller to a decision maker and manager, a shift from active to supervisory control. The supervisory task can be automatically started when the automation is turned on or the operation mode is set to auto mode. The supervisory responsibility includes monitoring the automated operating system, the start or shut off of the system, and restoring the order. This duty emphasizes judgment and solution [1]. Therefore, the efficiency carrying out and the back coupling do not mostly come directly from automation but is by the panel. As a result, automation monitoring work is mostly boring work, such as: the system-monitoring supervisor in a nuclear power plant. The task is much more difficult if

C.-L. Liu (✉) • R.-L. Lai
Vanung University, No. 1, Vannung Rd., Jhongli, Taoyuan City, Taiwan
e-mail: johnny@vnu.edu.tw

S.-T. Uang
Minghsin University of Science & Technology, No. 1, Xinxing Rd., Xinfeng, Hsinchu County, Taiwan

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_14

attention must be maintained on some source of information for the occurrence of infrequent, unpredictable events over long periods of time [2]. When the activity is to be performed for a continuous period of time, the ability to maintain attention may be severely impaired. Under such an irrevocable work environment, loss of attention may lead to accidents. Therefore evaluating the performance of sustained attention is often applied to improve supervisory tasks. Smit et al. [3] found that attention decreases due to hard mental work in performance of misses, false alarms and reaction time. The ability to maintain attention for such events typically declines as time goes, a phenomenon known as the attention decrement. A good measurement of sustained attention performance can help to understand observers' ability to remain alert for prolonged periods, further, to improve the monitoring environment, to combat the sustained attention decrement [4]. The theory of signal detection is a model of perceptual processing that is often used to characterize performance effectiveness in signal detection situations because it permits the derivation of independent measures of perceptual sensitivity and response bias [5, 6]. The most frequently used measures of perceptual sensitivity include the parametric index *d'* and its nonparametric analog *A'*. The most common bias measures are the parametric index β and the nonparametric indices B'_{H} and B'' [7, 8]. Although, in previous researches, these indices were used to measure sustained attention performance, there are some limitations. In some supervisory environment, the reaction time is a free response. For example, the supervisory operator of nuclear power plant, who must observe the warning level of cooling water for a long time. It can be defined as unlimited hold task. It should not only be emphasizes either "Yes" or "No" in the judgment concept, but the different evaluation of decision-making at different time points. A reliable detection measure of response bias is particularly critical in this field of study, because alterations in response bias are a predominant feature of attention performance [9, 10].

Some researchers studied the measures of sensitivity and bias of sustained attention obtained from reaction time during continuous and repetitive activity. According to the characteristics of unlimited task, the stimuli and responses form a continuous, mixed time series. Reaction time provides the principal measure in tasks with supra-threshold signals. Egan et al. [11] used a task without discrete events and without specified observation intervals. Signals were presented at times unknown to the observer, who was free to respond at any time, (i.e., free response). Because stimuli and response form a continuous, mixed time series, classifying the observer's response into hits and false alarms was difficult. The observer's response rate was plotted as a function of time following a signal. Response rate rose sharply immediately following a signal and then fell to a constant, low level a few seconds later. The probabilities of hits and false alarms were equated with the areas under the two segments of the distribution. By inducing observers to adopt different criteria for reporting targets they were able to generate pairs of hit and false alarm probabilities. Silverstein et al. [12] used sensitivity (*d'*) and reaction time to assess the ability of individuals with schizophrenia to sustained attention to visual stimuli. Szalma et al. [13] discussed the effects of sensory modality and

task duration on performance by percentages of correct detections and reaction time in sustained attention. However, if we can integrate the concept of hit, false alarm and response time, the measurement of sustained attention should be more efficient in unlimited hold tasks. Using membership function of the fuzzy sets is an extremely good method. In view of the domain of decision response, the different weighting values can be assigned, then fuzzy logics are used to evaluate the sustained attention performance. So, the purpose of this study was to propose a quantitative sustained attention measuring model by fuzzy sets, verified with other indices by a simulation experiment.

2 A Fuzzy Sustained Attention Measuring Model (FSAMM)

In sustained tasks, there will be more sensory or neural activity in the brain when a signal is present than when it is absent [14]. We refer to the quantity X as the evidence variable when the external stimulus generate human neural activity. If there is enough neural activity, X exceeds the critical threshold X_1 of preliminary attention, and the operator will decide "yes." If there is too little, the operator will decide "no," as shown in Fig. 1. After a period of time, the signal gradually strengthens until X exceeds criterion X_1, and the operator will say "yes." Sometimes, even when no signal is present, X will exceed the criterion X_1 because of random variations in the environment and the operator's own "baseline" level of neural firing, and the operator will say "yes" (generating a false alarm). When the operator detects an unusual or infrequent condition in the preliminary vigilance, then he/she must keep sustained attention to detect any resulting emergency signals.

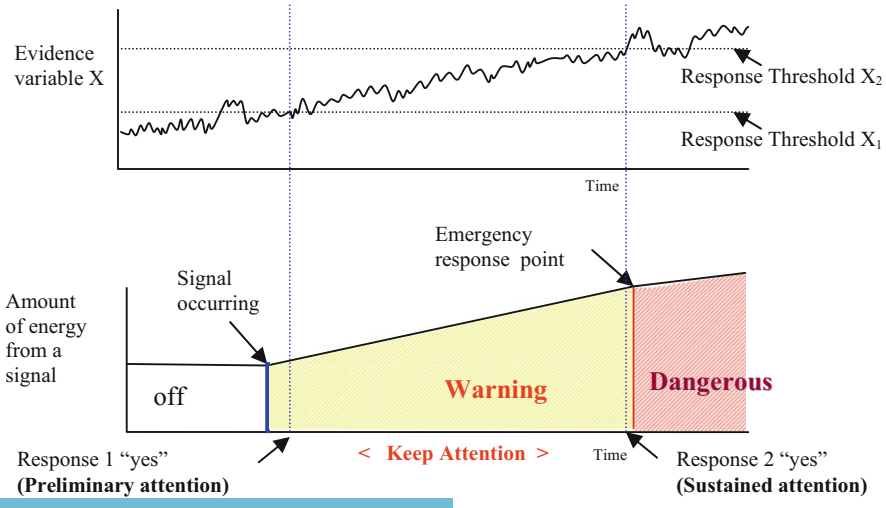


Fig. 1 The change in the evidence variable x caused by signal in supervisory task

If the system approaches a dangerous situation and there is enough neural activity, then X exceeds the critical threshold X_2 of the sustained attention, and the operator will decide "Yes," otherwise, decide "No."

When the human neural activity in preliminary and sustained attention is understood, a fuzzy logic inference can then be used to develop a fuzzy sustained attention measuring model (FSAMM). The fuzzy logic inference can be seen as a heuristic and modular way to define nonlinear, table-based measurement and control [15]. A fuzzy logic is a set of IF-THEN rules. The linguistic statements of the IF-part are obtained by fuzzyfication of numerical input values, the statements of the THEN-part are defuzzyficated to numerical output values [15, 16]. Assume the fuzzy rule consists of N rules as follows:

$$\begin{aligned} R_j(jth\ rule) : If\ x_1\ is\ S_{j1}\ and\ x_2\ is\ S_{j2}\ and \ldots and \ldots x_n\ is\ A_{jn} \\ Then\ \ y_1\ is\ O_{j1}\ and\ y_2\ is\ O_{j2}\ and \ldots and \ldots y_m\ is\ O_{jm}, \end{aligned} \tag{1}$$

where $j = 1, 2, \ldots, N$, x_i ($i = 1, 2, \ldots, n$) are the input variables to the fuzzy system, y_k ($k = 1, 2, \ldots, m$) are the output variables of the fuzzy system, and A_{ji} and O_{jk} are linguistic terms characterized by their corresponding fuzzy membership function $\mu_{\tilde{A}ji}(x_i)$ and $\mu_{\tilde{O}\ jk}(y_k)$.

Reminding the behavior of human decision-making, we use two factors to construct vigilance performance measuring model: one is the "ability of signal detecting" (Hit), another is the "ability of false alarm" (False alarm). We considered two variables as input fuzzy sets: one is the fuzzy set $\tilde{S}$, which represents the linguistic notion "*signal detection ability*" (Hit), and the other is the fuzzy set $\tilde{N}$, which represents the linguistic notion "*false alarm status*" (False alarm).

The fuzzy set $\tilde{S}$ includes $\tilde{S}1$ and $\tilde{S}2$ ($\tilde{S}1$ is the signal detection capability during preliminary attention, and $\tilde{S}2$ is the signal detection capability during sustained attention), are described by three attributes: "High signal detection ability" (HI), "Medium signal detection ability" (ME) and "Low signal detection ability" (Lo). The fuzzy set $\tilde{N}$ includes $\tilde{N}1$ and $\tilde{N}2$ ($\tilde{N}1$ is the status of false alarm during preliminary attention, and $\tilde{N}2$ is the status of false alarm during sustained attention), are described by three attributes: "Low false-alarm" (LO), "*Medium false-alarm*" (ME), and "*High false-alarm*" (HI). Second, the output fuzzy set $\tilde{V}$ can be described as the linguistic notion "*Performance of sustained attention*", and includes five attributes: "Excellent", "Fine", "Good", "Fair", and "Poor." Let y represent the sustained attention performance level. If the value of y is large, this indicates a high sustained attention performance. According to fuzzy logic theory, the concepts of "quantization" and "uniformity" are used to discretize the continuous domain [−25, 125] into six segments using five judgment criterions: 0, 25, 50, 75, and 100. Therefore, we can define the membership functions of $\mu_{V_{PR}}, \mu_{V_{FR}}, \mu_{V_{GD}}, \mu_{V_{FN}}, \mu_{V_{EX}}$ as triangular functions. Finally, the IF-THEN rules are defined to process fuzzy reasoning and the final sustained attention performance value y* is calculated using the Center-of-Area method (also referred to as the Center-of-Gravity method in the literature). The rule base was subsisted of 81 rules. In matrix notation, these relations can be shown in Table 1.

Table 1 Rule base for sustained attention performance

Sustained attention performance				Signal detection ability during preliminary attention								
				High			Medium			Low		
				Signal detection ability during sustained attention								
				High	Medium	Low	High	Medium	Low	High	Medium	Good
False alarm status during preliminary attention	High	False alarm status during sustained attention	High	Poor	Poor	Poor	Poor	Fair	Fair	Poor	Good	Good
			Medium	Poor	Poor	Poor	Fair	Fair	Fair	Good	Good	Good
			Low	Poor	Poor	Fair	Fair	Fair	Good	Good	Good	Fine
	Medium		High	Poor	Fair	Fair	Fair	Good	Good	Good	Fine	Fine
			Medium	Fair	Fair	Fair	Good	Good	Good	Fine	Fine	Fine
			Low	Fair	Fair	Good	Good	Good	Fine	Fine	Fine	Excellent
	Low		High	Fair	Good	Good	Good	Fine	Fine	Fine	Excellent	Excellent
			Medium	Good	Good	Good	Fine	Fine	Fine	Excellent	Excellent	Excellent
			Low	Good	Good	Fine	Fine	Fine	Excellent	Excellent	Excellent	Excellent

3 Experiment

3.1 Experimental Environment

A simulated automation supervisory system was designed to verify the effect of FSAMM as shown in Fig. 2. The automated system is an Auxiliary Feed-Water System (AFWS) of a generic Pressurized Water Reactor in a nuclear power plant. There are four steam generators (SG1, SG2, SG3, and SG4). If the water level of any SG falls below the normal water level of 6 m (defined as abnormal status), the alarm signal on the SG will become orange (e.g., SG3 in Fig. 2), and the participant must turn on the manual feed-water system (MFS) by pressing the MFS valve button in order to add water (the preliminary vigilance). However, if the MFS valve is turned, but the water level continues to decrease, the participant must pay increased attention to prevent the water level from dropping below the emergency level (which is set at 5 m in this experiment). If the water level of any SG falls below the emergency level, the alarm signal on the SG will become red (e.g., SG2 in Fig. 2), and the participant must turn on the emergency feed-water system (EFS) by pressing the EFS valve button (the sustained attention). In order to simulate a complex experimental environment as a real supervisory environment, there are four signals flashing on the top side of the screen designed as secondary tasks for

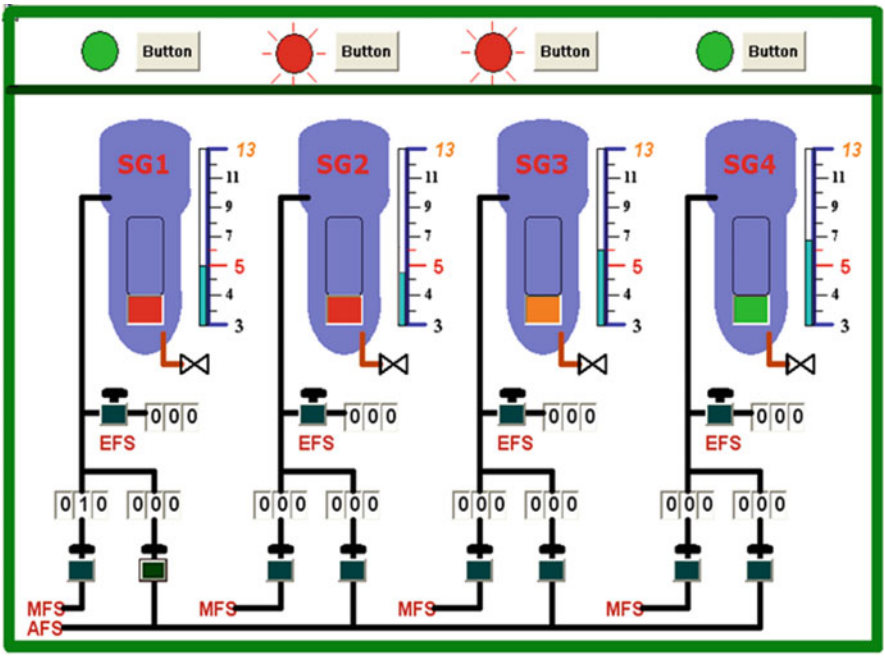


Fig. 2 A simulation of auxiliary feed-water system

subjects. When any of these signals flashed, the subject should press the switch to stop flashing.

3.2 Subjects

There are 16 Subjects requested to monitor and feed cool water satisfying the SG situations, and are requested to maintain safety: If he/she detects emergency status, it should be controlled immediately. Each subject must accept the training (at least three times) to learn how to control the simulated system.

3.3 Experimental Design

There are two independent variables "Event interval" and "Event rate" tested in a two-factor replicate experiment.

1. Event interval (factor A): refers to the total time of the emergence situation (i.e. event). There are three levels to be considered: 5 min. (A_1), 10 min. (A_2), and 20 min. (A_3).
2. Event rate (factor B): refers to the frequency of the unusual situation (i.e. event). There are two levels: 1×/per half hour (B_1) and 5×/per hour (B_2).

There are four dependent variables (performance variables) used to record the performance of the participants, including:

1. The water level at the time of the participant's response.
2. "Hit" rate. The "Hit" condition is defined as follows. If the participant detects the water level of any SG to be below the normal water level and presses the MFS valve button between 5.9 and 6 m, the response is defined as a "Hit" in the preliminary vigilance. If not, it is defined as a "Miss." In the extended vigilance, the "Hit" is defined as the interval between 4.9 and 5 m.
3. "False Alarm" rate. If the alarm signal appears when the water level is at 6.3 m (a noise in the preliminary vigilance), the participants must press the AFS (Automatic feed-water system) valve button to cancel the alarm before it reaches 6.1 m, otherwise it will be a "False Alarm". In the extended vigilance, the alarm signal appears at 5.3 m and must be cancelled before it reaches 5.1 m.
4. Fuzzy value y^* of sustained attention performance.

4 Results and Discussion

4.1 Effects of FSAMM

In experiment, we define "Miss" as a type I error, which may cause a system shutdown with very serious consequences. "False Alarm" can be defined as a type II error, which will result in the loss of benefits from the system and may increase the operator's workload. In general, the damage of a type I error is more serious than that of a type II error. The results showed that the fuzzy logic value can be attributed to this difference, but the d' and β value cannot be. Secondly, as can be seen from result, the hit rate of participant #1 was 100%, the Z (hit) value (a standard score of the normal distribution) was equal to $-\infty$, so the d' value was ∞. Since this value is infinite, it is impossible to explain the sustained attention performance level. However, the value of the fuzzy sustained attention measuring model is 71.76% (i.e., the value of the fuzzy sustained attention measuring model is between 0% and 100%; a value of ∞ cannot occur in the fuzzy model). The results imply that the sensitivity of the FSAMM is better than index d' and β in supervisory tasks.

The analysis of variance in the performance of the sustained attention via the fuzzy logic model shows that the F of 27.2473 ($p < 0.0004$) for event interval (factor A) was significant. This result indicated that there was a considerable difference among the three levels of event intervals during the extended vigilance and the vigilance performance in the 20 min interval was worse than the 10 and 5 min intervals. The correlation coefficient of the preliminary and extended vigilance performance is 0.8955. The result showed that if the participant has a good performance in preliminary attention, his/her performance will be better during sustained attention. The reason for this may be that the presence (and detection) of events in the preliminary attention act as "stimulants" that better arousal during sustained attention. It was also consistent with the other previous research when the time of supervisory task is longer, the performance of sustained attention would be poor [17].

4.2 Improvement of Sustained Attention

In the FSAMM, we defined $y = 50$ (%) as good status, and in the first experiment the average performance of preliminary vigilance is 47.31 (%). Therefore, the performance value 50 (%) of y^* was defined as the threshold for producing an alarm signal. If the value of y^* was smaller than 50 (%), the system would produce an alarm signal to warn the participant to maintain his/her sustained attention.

In the sustained attention experiment, we found that the preliminary attention performances of participants #4 and #11 to #18 were poor, so the fuzzy alarm activated. A differences of means test showed that these three participants' sustained attention performances were significantly better than in the preliminary

attention. However, we also found that the means of the sustained attention performance of some participants was poor than in the preliminary attention when the FSAMM not working. These findings indicated that the effect of an alarm for improving the sustained attention performance is significant when the performance of preliminary attention of participants is poor.

5 Conclusions

In this study, a fuzzy sustained attention measuring model (FSAMM) with relation to human decision-making in supervisory task was proposed and verified by conducting experiments. The main findings of this study are as follows:

1. In general, sensitivity parametric index *d'* and bias parametric index β are used to evaluated the performance of sustained attention. These indices only used observer's response "yes" or "no" to explain and evaluate sustained attention. However, for some supervisory tasks such as unlimited hold tasks (e.g., supervisors in nuclear plant), the reaction time is longer. These indices are unable to respond to this characteristic, but the FSAMM could.
2. The FSAMM could consider the characteristics of the difference between Type I error (miss), Type II error (false alarm) and reaction time in unlimited supervisory control task. Using fuzzy sets to evaluate the domain of decision response, the differences of performance could be evaluated. It could expand the traditional two values logic to continuous multiple-valued performance evaluation, then, clearly describe difference of the judgment in the monitoring "also this also other" work.
3. The fuzzy logic model and *d'* are more efficient in response sensitivity than β in this study. The correlation coefficient between d' value and the fuzzy model is 0.569. However, if the automation is in high reliability, the effect of d' value on zero miss will be over estimated. The fuzzy model will not.

By using fuzzy variables and fuzzy rules, the fuzzy logic measuring model combined the "*signal detection ability*" and "*false alarm status*" to evaluate sustained attention. According to the results, then, we can take one step ahead to analyze and design when and how to call operator's sustained attention at the right time to set an adapted attention performance alarm to reduce the probability of human decision-making error.

References

1. Szalma JL, Taylor GS (2011) Individual differences in response to automation: the big five factors of personality. J Exp Psychol Appl 17:71–96

2. Warm JS, Jerison H (1984) The psychophysics of vigilance. In: Warm JS (ed) Sustained attention in human performance. Wiley, England
3. Smit AS, Eling TM, Coenen ML (2004) Mental effort causes vigilance decrease due to resource depletion. Acta Psychol 115:35–42
4. Hancock PA (2013) In search of vigilance: the problem of iatrogenically created psychological phenomenon. Am Psychol 68:97–109
5. Kenneth RB, Lloyd K, James PT (1986) Handbook of perception and human performance, vol 2. Wiley, New York
6. Wickens CD (1992) Engineering psychology and human performance. HarperCollins, New York
7. Grier JB (1971) Nonparametric indexes for sensitivity and bias: computing formulas. Psychol Bull 75:424–429
8. Hodos W (1970) A nonparametric index of response bias for use in detection and recognition experiments. Psychol Bull 74:351–354
9. Parasuraman R (1979) Memory load and event rate control sensitivity decrements in sustained attention. Science 205:924–927
10. Parasuraman R, Davies DR (1977) A taxonomic analysis of vigilance performance. In: Mackie RR (ed) Vigilance: theory, operational performance, and physiological correlates. Plenum, New York
11. Egan JP, Greengerg GZ, Schulman AI (1961) Operating characteristics, signal detectability and the method of free response. J Acoust Soc Am 33:993–1007
12. Silverstein SM, Light G, Palumbo DR (1998) The sustained attention test: a measure of attentional disturbance. Comput Hum Behav 14:463–475
13. Szalma JL, Warm JS, Matthews G, Dember WN, Weiler EM, Meier A (2004) Effects of sensory modality and task duration on performance, workload, and stress in sustained attention. Hum Factors 46:219–233
14. Wickens CD, Lee JD, Liu Y, Becker SG (2004) An introduction to human factors engineering, 2nd edn. Person Education, New Jersey
15. Driankov D, Hellendoorn H, Reinfrank M (1996) An introduction to fuzzy control. Spring, New York
16. Zadeh LA (1973) Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans Syst Man Cybernet 3:28–44
17. Pattyn N, NeytX HD, Soeten E (2008) Psychophysiological investigation of vigilance decrement: boredom or cognitive fatigue? Physiol Behav 93:369–378

Waukesha 7044 Gas Engine Failure Investigation

Xiaofeng Liu and Adam Newbury

Abstract Three engine failures occurred from March to July in 2015. The failures were due to engine overloading and in each of the three instances caused severe scuffs and cracking on pistons and liners, particularly on #4 cylinders. Operating condition changes were identified as the root cause for the recent engine overload failures. Several maintenance issues, such as incorrect stepper motor settings, lubrication oil over filling and failed spark, also contributed to the failures. These factors combined and caused the engines to be overloaded and triggered the failure mechanisms which lead to the engine failures. To avoid failures in the future, the recommendations are provided, such as reduce the engine load by increasing the compressor clearance; monitor the engine load when operating condition changes and provide appropriate engine load awareness training to engineering and field personnel.

Keywords Engine failure investigation • Gas compressor • Performance modelling • Root cause analysis • Engine overloading • Corrective maintenance

1 Introduction

Since March 2015, there have been a number of engine cylinder failures on one APA compressor station. The compressor station consists of three Enerflex packages which comprise of a Waukesha 7044 GSI gas engine and Ariel JGK/4 single stage gas compressor [3]. The failure history is as follows:

- Severe scuffing was found on Unit 2 Right Bank (RB) cylinder 4 piston and liner on 10/3/15 at 636 h operating hours.
- Severe scuffing was found on Unit 2 Left bank (LB) cylinder 4 piston and liner on 14/04/2015.

X. Liu (✉)
APA Group, Level 3, 121 Wharf Street, Spring Hill, QLD 4000, Australia
e-mail: xiaofeng.liu@apa.com.au

A. Newbury
APA Group, Suite 1, 60-62 Gladstone Street, Fyshwick, ACT 2620, Australia

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_15

- Split liner and piston damage were found on Unit 1 RB cylinder 4 and in addition severe scuffing on LB cylinder 4 piston and liner on 05/07/2015.

All failures occurred on left and right bank #4 cylinders. An initial investigation found that several factors might have contributed to the failures:

- According to the ESM (OEM engine system management computer)/SCADA data:
  - Engines were found to be overloaded prior to failures.
  - Cylinder temperature deviations were observed, these were higher on the bank that had the failure.
  - Manifold banks appear to have been imbalanced, the manifold air pressure was higher on the bank that failed.
  - The RB has far more alarms and Emergency Shut-downs (ESDs than the LB, this varies from a factor of three times up to 127 times that of the LB, refer to Table 2 for more details.
  - LB and RB gas/air ratios and (Inlet Manifold Absolute Pressure) IMAPs were found to be too high on unit 3.
- Oil level controller issues caused oil overfilling the frame and sump on unit 2 for the second failure which can increase engine load substantially.
- A couple of cylinder spark failures were also spotted which may have overloaded other cylinders.
- It appears that #4 cylinders, either left bank or right bank, are the most vulnerable cylinders.

2 Root Cause Analysis

2.1 Piston Failure Mechanism

Figure 1 shows a failed piston from Unit 2, on which scuffs and cracks are evident. This is a typical engine failure at the compressor station.

Research by Waukesha has found that the scuffs and cracks are caused by the so called wear load. Wear load is defined as a measure of frictional power due to asperity contact per unit area with units of MW/m^2. This frictional energy flux will dissipate heat into the piston and liner. Apparently, this frictional power is related to pressure of metal to metal contact and sliding velocity. The heat generated from this process can cause the oil viscosity to drop leading to further asperity contact and more frictional heating. This thermal instability will eventually lead to scuffing and even seizure [1].

Fig. 1 Failed unit 2 piston

2.2 Root Cause Analysis

Scuffing can occur under abusive operating conditions, typically overloading. Under overloading conditions, metal to metal contact pressure can occur and the piston sliding velocity is also higher as the engine normally operates at higher speed. Engine overloading was observed on the historical data of all the three engines prior to the failures. In one case; the engine was overloaded (exceeded 100% load) for a continuous period of 12 h. Of the 12 h, more than 100 min the engine load was at about 110%. There are two circumstances when engine can be overloaded:

- The actual engine power output is greater than nominated horse power, which is 100%;
- The actual engine power output is less than nominated horse power, but some cylinders have exceeded their capacity and are overloaded, due to engine tuning issues.

2.2.1 Operating Condition Changes and Performance Modelling

Engine overloading can be caused by compressor demanding more power due to changes of operating conditions, e.g. the suction pressure and discharge pressure changes. As the suction pressure increases, engine load will increase and then

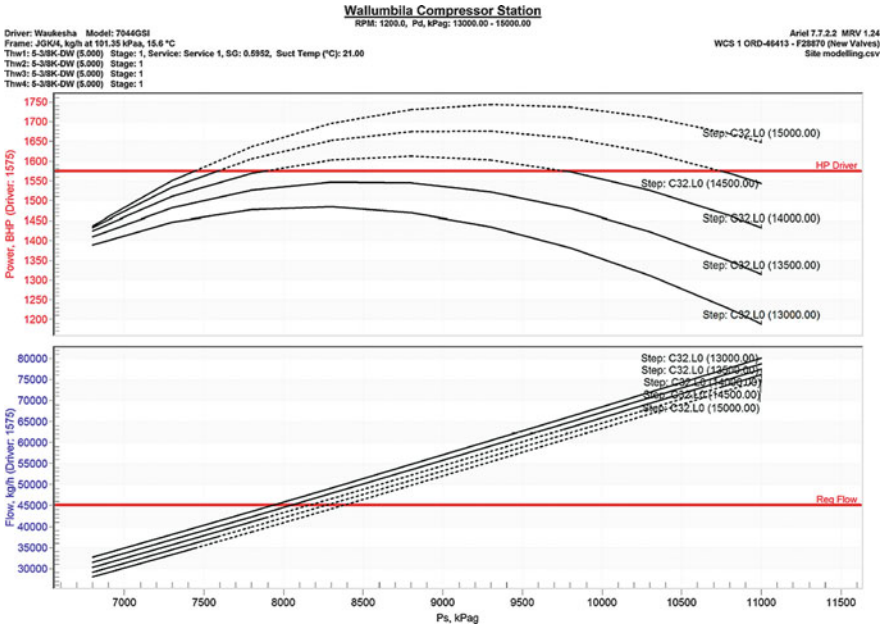


Fig. 2 Compressor performance under current configuration at 1200 rpm

decrease. It reaches maximum load at 8500 kPa suction pressure when discharge pressure is 13,000 kPa. As the discharge pressure increases, the engine load always rises, see Fig. 2. Figure 2 is acquired using the Ariel performance software.

From January 2015, the suction pressure at the compressor station has been operating in the range from around 7800 to 10,000 kPa. The discharge pressure has been fluctuating mostly within the range from 10,800 up to 14,800 kPa.

Under certain operating conditions, especially at the high discharge pressures (up to 14,800 kPa); it is likely that the compressor demands more power than the engine can generate. The engine will be overloaded when this occurs. Please refer to the dotted section of the engine power curves. At 1200 rpm the power required is 1680 hp, while the driver can only provide a maximum of 1575 hp which equates to 106.7% of sustainable available power before considering power deterioration due to tuning issues or engine condition. Figure 3 shows an engine overloading situation possibly caused by decreased suction pressure and elevated discharge pressure.

2.2.2 Engine Tuning Issues

Engine tune imbalance issues were observed in the historical data at compressor station and contributed to engine failures. The identified issues included cylinder and bank imbalance issues, typically cylinder temperature deviation and uneven LB and RB gas/air ratio.

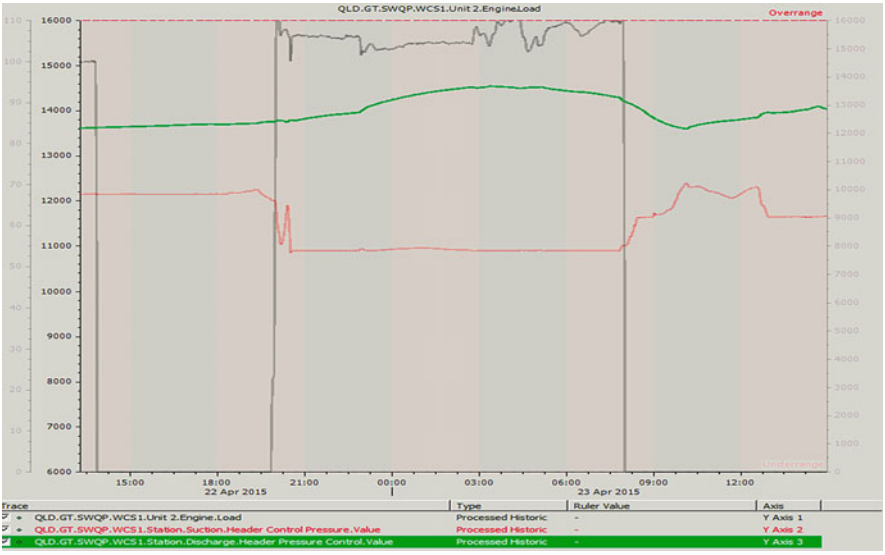


Fig. 3 Engine overloading possibly caused by changed operating conditions

Fuel pressure regulator stepper motor malfunction was found to be a major contributor. For the Unit 1 failure, one stepper motor setting was found to be at 400 and the other at 4000, which exceeded the acceptable 500 LB and RB stepper motor tolerance by a factor of 7.2.

Cylinder Temperature Deviation

Stepper motor malfunctioning can cause serious engine imbalance issues. Cylinder temperature deviations were observed and manifold banks Gas/Air ratios were also found to be out of range. The cylinder temperatures were also found uneven when Unit 1 engine failed [2]. In this instance high cylinder temperature was found on the failed LB #4 cylinder.

Gas/Air Ratio and Engine Intake Manifold Absolute Pressure (IMAP) Issues

On Unit 3, the LB and RB gas/air ratios and engine intake manifold absolute pressures (IMAP) were found to be too high, see Table 1. It can be seen from the table that Unit 3 was burning rich gas fuel which might have contributed to the temperature abnormality, while the IMAP is a direct reading of engine load. The IMAP is above the specification which shows the engine is in overloading condition.

Table 1 Unit 3 engine bank Gas/Air ratio and IMAP

Gas/Air ratio			IMAP		
LB (H2O")	RB (H2O")	Spec (H2O")	LB (HG")	RB (HG")	Spec (HG")
13–16	12–15	4.9	11.2	10.5	8

Table 2 Unit 1 alarms and shutdowns since reset

Faults (alarms and ESDs)	Number of faults, LB	Number of faults, RB	Fault ratio (LB: RB)
OXYGEN LOW (alarms)	77	231	1:3
OXYGEN OC (Alarms)	9	37	1:4.1
EXH TEMP OC (Alarms)	23	255	1:11.08
EXH TEMP HIGH (Alarms)	3	255	1:85
HIGH EXH TEMP (Alarms)	2	255	1:127.5
KNOCK CYL (ESD)	1	4	1:4

ESD Emergency shut down, OC Open circuit, *EXH* Exhaust, *TEMP*, Temperature, *CYL* Cylinder

Engine Bank Alarms and Shutdowns

Table 2 shows statistics of alarms and shutdowns on Unit 1 since reset, from this data it is apparent that the engine is in a severe bank imbalance situation. The alarm and shutdown count is substantially higher on the right bank than the left bank, the exhaust temperature high alarm shows the greatest ratio of 1:127.5.

2.2.3 Other Maintenance Issues

There are several maintenance issues that also contributed to engine overloading failures, these are described below.

Oil Overfilling

When Unit 2 failed on 14th of April, both the compressor frame and engine sump were found to be overfull with lubricating oil. Upon further investigation it was learned that the oil level controllers were malfunctioning which caused the issue. The high oil level creates parasitic loads on the engine and compressor crankshafts as they work to displace the oil during each rotation. This causes extra friction within the engine and compressor thus increasing engine load. A field test was conducted and proved the assumption, see Table 3.

It is recommended that the Kenco engine lube oil level controllers be replaced. The vent tubing from Kenco to engine crankcase should be minimum ½" tube, continuously sloping towards Kenco and should be as short as possible.

Table 3 Engine load vs oil level test

	Oil overfilling	Normal oil level
Engine speed	750 RPM	750 RPM
Compressor oil temperature	29 °C	29 °C
Compressor oil pressure	451 Kpa	446 Kpa
Engine oil temperature	55 °C	58 °C
Recycle valve	100% open	100% open
Engine load (start)	**37%**	**25%**
Engine load (running)	**25.3%**	**16.8%**

Ignition/Spark Failure

There were also several cases when failed ignition was found. If one cylinder has a failed spark, the remaining cylinders will need to work harder to compensate for the power loss, which causes overloading. For a 12 cylinder engine, each cylinder contributes 8.33% power. Therefore, 131.25 hp was lost for each spark failure. One known issues is that the engine wiring harness was damaged by rodents shortly after commissioning and repaired locally; the harness integrity has been compromised. Therefore upgrading of the engine wiring harness is recommended to reduce ignition failures and other reliability issues.

2.2.4 Potential Design Weak Point: #4 Cylinder

The fact that all the failures are on #4 cylinder, either left bank or right bank, indicates #4 cylinder is likely to be in the most vulnerable location. Although no evidence was found to support this hypothesis, the vulnerability may be caused by insufficient lubrication and/or heat transfer. However, historic operation of these engines indicates that the #4 cylinders would be less vulnerable if the engine is tuned and loaded properly.

2.2.5 Vibration

Concerns were raised by field personnel that the failures might be triggered by vibration issues as high vibration was reported during commissioning. Enerflex was contacted and the original commissioning document was reviewed. The vibration was found not to be an issue.

It is understood that over time vibration levels may change, there still lack of evidence that there is a correlation between engine failure and vibration in these failures. Further investigation may be required if more vibration failure evidence becomes available.

2.2.6 Lubrication

The engine lubrication oil was changed from Castrol Duratec L-MG to Castrol Duratec L. Although the Kinematic Viscosity of Duratec L is slightly lower than that of Duratec L-MG, both of them meet Waukesha engine lubrication requirements. The reason for the change was that Duratec L-MG40 became obsolete and no longer available from the vendor. Duratec L was the vendor replacement product. The change was reviewed and approved under management of change procedures. It is therefore unlikely that the oil change contributed to the engine failures.

However, upon review of the oil analysis reports, it appears that the oil sampling is conducted on a run hour basis. Considering the actual frequent starts/stops which could contribute to accelerated wear, It is recommended that fluid sampling be conducted at a minimum monthly interval. The oil sampling and analysis program should be reviewed for alignment to OEM requirements for each piece of equipment.

3 Corrective Maintenance Actions

3.1 Apply Valve Spacers to Reduce Engine Load

Valve clearance spacers can be applied to reduce the compressor power requirement and therefore, the engine load. However as a result, the gas flow rate will be reduced. To minimize flow rate losses, the clearance used need to be optimized.

Based on Ariel performance modeling, two valve spacers at suction side are recommended for 14,500 kPa discharge pressure, and four spacers are recommended if discharge pressure of 15,000 kPa is anticipated. The Variable Volume Clearance Pocket (VVCP) also needs adjustment to maximize gas flow rate, depending on the suction pressure. Figure 4 shows the VVCP settings under 14,500 kPa discharge pressure.

3.2 Review Engine Operating and Maintenance Practices

As some maintenance issues also contributed to the engine failures, it is recommended the following actions be considered:

- Review daily read sheets to ensure all essential engine parameters are logged and compared from one day to the next.
- Introduce criteria for assessment of deviation in each parameter from one day to another. Introduce requirement for engineering notification if deviation exceeds limits noted.

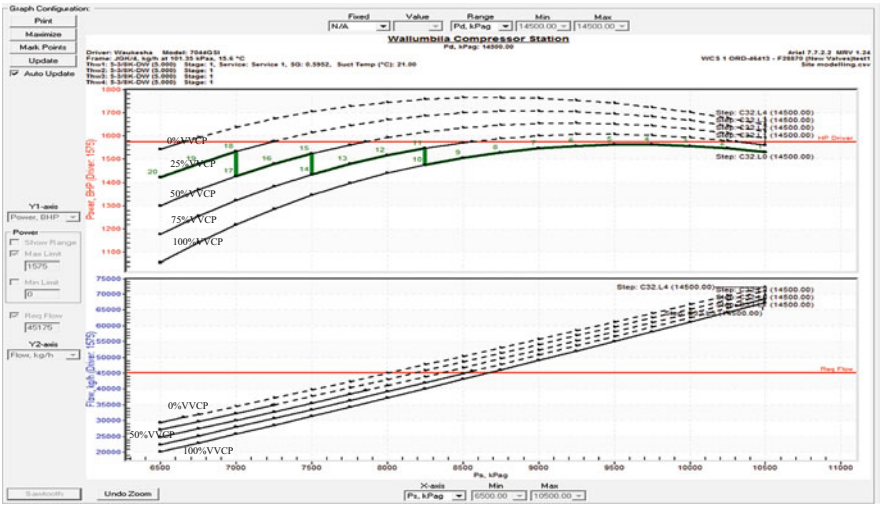


Fig. 4 Recommended VVCP operating envelope (*green line*) at 1200 rpm, under 14, 500 kPa

- Introduce weekly ESM fault logging data capture, reporting to Engineering and clearing.
- Revise maintenance documentation to record essential tuning parameters every 2000 h (Gas/air pressure left and right bank and pressure difference, stepper position left and right and difference, IMAP left and right and difference, throttle position) as well as log basic operating parameters for a period of no less than 1 h following every service.
- Recommend replacing Kenco engine lube oil level controllers, confirm that vent tubing from Kenco to engine crankcase is minimum ½" tube, continuously sloping towards Kenco and as short as possible.

4 Conclusion

Operating condition changes were found to be the root cause for the failures, in particular the elevated suction and discharge pressures created exceeded available engine power. In addition, several maintenance issues such as incorrect stepper motor settings, lubrication oil compartment over filling and failed spark contributed to the failures. These factors combined and caused the engines to be overloaded and triggered the failure mechanisms which led to the engine failures.

5 Recommendations

- Reduce the engine load by increasing the compressor clearance (install valve spacers). Increasing the compressor clearance will reduce the power requirement. As a result the capacity might be affected. VVCP can be adjusted to optimize compressor performance under changing operating conditions.
- Monitor the engine load when operating condition (especially the suction and discharge pressure) changes; implement engine load alarms and trips to protect the engine from being overloaded for prolonged period of time.
- Review the operating and maintenance practices.
- Provide appropriate engine load awareness training to engineering, field personnel, engineers and controllers; make sure engine tuning procedure is in place and comply to the OEM requirements and provide field personnel with engine tuning training to ensure any tuning issues can be identified and rectified properly.

References

1. Donahue RJ, Draege WA, Schaefer NP (2009) Robust design of a large industrial natural gas engine piston. In: Proceedings of the ASME internal combustion engine division 2009 spring technical conference ICES2009, Milwaukee, Wisconsin, USA
2. Liu X, Newbury A (2016) Waukesha 7044 gas engine failure investigation, Technical report, February 2016, APA Group, Australia
3. MPA Consulting Engineers (2008) Wallumbilla compressor station project design basis manual, May 14, 2008, APA Group, Australia

Construction of Index System of Comprehensive Evaluation of Equipment Maintenance Material Suppliers

Xuyang Liu and Jing Liang

Abstract To evaluate and choose equipment maintenance material suppliers (EMMS) through index system of comprehensive evaluation is one of key approaches for equipment maintenance support. Index system of comprehensive evaluation of EMMS is established based on analysis on basic conditions of being EMMS and research of particularities of EMMS' comprehensive evaluation.

Keywords Excellent suppliers • Equipment maintenance material supply chain • Index system of comprehensive evaluation

1 Introduction

Equipment maintenance support is the general designation of technique measures and correlative service activities to maintain and renew technique status of equipment. The support quality of equipment maintenance material, which is the basis of normal work of equipment maintenance support system, directly relate to realization of maintenance support purpose [1]. As the "headstream" of the whole equipment maintenance material supply chain, equipment maintenance material suppliers (EMMS) is extraordinary important. Accordingly, evaluation and choice of excellent suppliers becomes one of the key approaches to equipment maintenance support. The key step of comprehensive evaluation of EMMS is to confirm the evaluation index system which is the gist of comprehensive evaluation of suppliers. Therefore, this article researches the construction of index system of comprehensive evaluation of EMMS based on supply chain management combining the practice need of military equipment purchase for evaluating EMMS scientifically and objectively.

X. Liu (✉)
Academy of Armored Forces Engineering, Beijing, 100072 Beijing, China

Equipment Academy, Beijing, 101416 Beijing, China
e-mail: liuxuyang7733@163.com

J. Liang
Equipment Academy, Beijing, 101416 Beijing, China
e-mail: jellyjing1216@163.com

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_16

2 Equipment Maintenance Material Supplier (EMMS)

2.1 Definition

The EMMS is the general designation of partners (corporations) who lawfully establish supply needed productions (or service) to the army and locate the top of the equipment maintenance supply chain [2]. The basic mission of EMMS is to supply material according to purchasing and maintenance need to the army and ensure the army's equipment maintenance material need either in peacetime or in wartime.

2.2 Basic Conditions of Equipment Maintenance Material Supplier

The military's demands to equipment maintenance material suppliers commonly include the following aspects:

2.2.1 Keeping Secret

The production purchasing activities of the army, especially the special-purpose equipment maintenance materials purchasing, have a high demand of keeping secret. Accordingly, most contents of production purchasing, such as, the variety, index, demand, quantity, transport, service place and so on, should be required for keeping secret.

2.2.2 Quality and Service

The military consumers have higher demand of quality and after service than general consumers. They demand not only higher cost performance, but also service based on whole life circle management provided by suppliers. That service based on whole life circle management, which consist acquirement service, exercise service and disposition service, is the service including the whole process of production exploitation, design, manufacture, sell, logistics, training, using, maintenance, rejecting and recovery.

2.2.3 Time Limit

The suppliers' ability of accomplishing the content of contract in formulary limited time, is required, such as, service production, transport, training, using and maintenance, and so on.

2.2.4 Military Special Demands

Military equipment purchasing has a bright character of order. Therefore, productions supplied by suppliers may be special in standard, life and appearance for adapting to the military's special demands of fighting, training, being on duty, maintenance. Some departments of equipment maintenance material purchasing have special producing technical requirement, for example, asking the suppliers to produce non-standard productions, to produce according to design blueprint and special technical requirement confirmed by higher-up and, to buy raw and processed materials and outsourcing parts according to special requirement established by higher-up.

3 Differences Analysis of Comprehensive Evaluation to Equipment Maintenance Material Supplier

Comprehensive evaluation and choice of EMMS are different from other suppliers in purpose and process of evaluation, choice of indexes and indexes weigh value coefficient and evaluation method design.

3.1 Stand of Evaluation

In the past time, the war industry scientific research task of the military was absolutely occupied by ten war industry group corporations which belonged to the national defence science industry committee. The participation chance of private enterprises was few. Along with the continual deepening of marketization reform of military industry, the military industry market opened to private enterprises gradually and some enterprise groups having advanced technology and abundant strength became the key cooperation partners. Evaluation of EMMS must be especially careful because that most cooperation items belong to advanced military core technic field, the productions and technic are proprietary, the demand of security and safety is extraordinary strict and items relate to the creation of military battle effectiveness and national military superiority. Therefore, the evaluation to EMMS should stand on the core competitive power and future development capability of enterprises, including technic standard, development potential, strategic cooperation capability, commerce credit, etc., to guarantee the success of strategic cooperation.

3.2 Index System of Evaluation

When traditional military equipment supplier's evaluation indexes are chosen, more attention should pay to those short-term indexes and indexes relating to

purchasing, such as price, quality, delivery date and history performance, and so on [3]. Evaluation index system under the condition of supply chain should add indexes relating to strategy, sustainable development, complementarity, cooperation compatibility, and so on. Evaluation indexes such as employee quality, management standard, history performance, enterprise credit, innovation capability of productions and technics, and so on, are emphases of evaluation indexes. Furthermore, evaluation and examination should be done to construction level of information system, consciousness of circumstance protection and so on, to confirm suppliers chosen can carry out of responsibility of not only supply raw and processed materials, accessories and parts to the army with high-quality, duly, well and truly but also take part in the military's activities such as new production exploitation, stock control, quality improvement, service betterment, and so on with active attitude, well diathesis and enough ability.

3.3 Enactment of Index Weigh Value

The EMMS evaluation based on supply chain management differs from traditional military equipment supplier's evaluation in index system. Even if the index was same, the enactment of index weigh value is different. An example is the index of price. Price is an very important index with a biggish weigh value in the evaluation of traditional military equipment suppliers, but its enactment of weigh value descend more in evaluation of EMMS based on supply chain management because that the military customers pay more attention to other indexes, including quality management ability, delivery ability, reaction flexibility, innovation ability, development potential, diathesis and credit, whose enactment of indexes weigh value are more bigger than the index of price's.

4 Construction of Index System of Comprehensive Evaluation to Equipment Maintenance Material Suppliers

4.1 Index System of Comprehensive Evaluation to Equipment Maintenance Material Suppliers

The military supply chain management, whose core is to satisfy the arm's need, is not only an operation integration process which begins from the army (the end of the chain) and ends at military equipment suppliers (the beginning of the chain), but also a process of realizing the best join of every parts and steps on the chain. In order to control the risk of military supply chain effectively, enterprises whose competitive power of production is powerful, capability of core operation is

extraordinary and management credit is well must be chosen to develop the strategic cooperation relationship. Therefore, the comprehensive evaluation of EMMS under the circumstance of supply chain management should follow some standards as performance of supplier, manufacture ability and management level of supplier, informationization degree, factors of person, environment of enterprise, cooperation ability of enterprise and characters of supplier (Table 1), etc.

4.2 Analysis on Relational Indexes

Explanations only about relational indexes which are different from common suppliers are given in the following subsections.

4.2.1 Military Purchasing Advanced Period Discount

Shorter the goods order advanced period is, better the supply chain's respond ability to the military customer's requirement is, and fewer stock in trade is needed. Under the condition of cooperation of supply and demand, the military customer wants the order batch output and purchasing advanced period as smaller as possible, while the supplier wants the military customer to increase order batch output. In the procession of real support, the military customer usually allows to increase order batch output to satisfy the need of speedy reaction support and shorten the purchasing advanced period. So the supplier uses the purchasing advanced period discount as an instrument to attract the military customer increase order batch output. By this way, the supplier's total cost is reduced and the military customer's support cost is reduced at the same time, that is to say, both sides of supply and demand win the income by the cooperation decision-making.

4.2.2 Military Purchasing Cost

Together with unit production quote price, the military purchasing cost which is the index should be consulted when the military analyses the cost is the gross of the price. The index of military purchasing cost has a macroscopical point of view, so the military can choose the supplier form the aspect of total price easily. It involves review cost, palaver cost, goods order cost and goods delivery cost of EMMS. In the period of T, assuming the production current rate of the supplier by P, the stock quantity by N, goods order cost spent by $C0$, the purchasing cost of unit production is:

Table 1 Index system of comprehensive evaluation to equipment maintenance material suppliers

One level indexes	Two level indexes	Three level indexes
Performance of supplier	Cost analysis	Quoted price of supplier's single production
		Military advanced purchasing discount
		Military purchasing cost
		Cost expense availability
	Quality performance	Pass percent of production
		Percent of damaged
		Percent of repair and return
	Delivery and transport guarantee performance	Delivery cost
		Percent of batch output abidance
		Loading efficiency
		Percent of delivery on time
		Average delivery time
		Advanced delivery time
		Delivery flexibility
		Batch output flexibility
		Variety flexibility
	Ability of feedback needed information to military	***Ability of smart reaction***
		Satisfaction degree of information feedback and maintenance
		Feedback degree of military requirement
		Percent of user satisfaction
	Future development foreground of supplier	
Manufacture ability and management level of supplier	Status of manufacture technique	Technique equipment level per person
		Devotion percentage of new technique innovation
		Personnel percentage of technique empolder
		Mission success rate of new production empolder
		Eligible empolder ability of new production
		Combination flexibility
		Patent level
		Empolder circle of new production
	Financial affairs status	Yield rate
		Balance rate
		Stock turnover rate
		Capital turnover rate

(continued)

Table 1 (continued)

One level indexes	Two level indexes	Three level indexes
	Equipment status	Facility advanced level
		Facility using
		Facility time using rate
		Facility disrepair rate
	Status of manufacture and production	Total throughput ability & total output
		Productivity of entire personnel labour
	Ability of market control	Market share
		Market overlap
		Market strain capacity
Quality system	Quality system satisfaction percentage	
	Quality engineer percentage	
Informationization degree	Outfit of information equipment	Integrative information equipment scale
		Improvement level of integrative information equipment scale
	Outfit of Information personnel	Full-time information personnel rate
		Improvement level of full-time information personnel rate
	Using of information equipment	Integrative using scale of information equipment
		Improvement level of integrative using scale of information equipment
Factors of person	Integrated diathesis exponential of high-grade manager	
	Training expense per person	
	Training time per person	
	Universities & colleges personnel proportion	
	Staff demission rate	
Environment of enterprise	Geography position/km	
	Economy technique level	
Cooperation ability of enterprise	Amalgamation of enterprise strategic target	
	Compatibility of enterprise culture	
	Information share degree	
	Consciousness of safeguarding the army	
Characters of supplier	Credit	Credit of repaying loan
		Keeping appointment scale
		Plan realization scale
		Bad credit record scale

(continued)

Table 1 (continued)

One level indexes	Two level indexes	Three level indexes
	Kind of enterprise	National property percentage
		Private property percentage
		Foreign property percentage
	Security	Security accident record
		Security accident record in a year
		Record of economy expense by security accident
		Difficulty economy expense by difficulty security accident
	Secrecy	Blow-the-gaff affair record
		Blow-the-gaff affair record in recent 5 years
		Carrying out of secrecy system
		Related reports of national secrecy department

$$C = \frac{P \times N + C0}{N} \tag{1}$$

4.2.3 Delivery Flexibility

The index of delivery flexibility, which is the ability to adjust and plan the delivery plan by the supplier when external condition changes, can reflect the supplier's reaction ability changing with the military's require. Apparently, the faster the reflection speed is, the more advantageous the military gains [4].

4.2.4 Batch Output Flexibility

The supplier's production output in a period of a program usually is planned carefully in advance. The production output can't be changed easily or the enterprise's economy benefit of the whole program period will be influenced. But just as two requests above, as the military's strategic cooperation partner, the supplier should have flexibility in order to make the production output in one planning period change flexibly in the stated range. The value of batch output flexibility must be discreet and right. When consulting the index of batch output flexibility, the military must do from a practical point of view instead of pursuing extreme. The index of batch output flexibility reflects the amount range of production supplied by EMMS in stated time limit. Assuming the average demand of a product supplied by

EMMS by Q, complete capacity by $Q\max$, the batch output flexibility of this EMMS is:

$$FOQ = \frac{Q\min - Q\max}{Q} \tag{2}$$

4.2.5 Variety Flexibility

The index of variety flexibility, which can be expressed by the production's variety amount synchronously produced by EMMS, reflects the ability of changing the variety of production by EMMS.

4.2.6 Ability of Smart Reaction

In order to complete support mission, the supplier should fleetly design and produce new type equipment fitting the campaign demand or ameliorate existing equipment. The index of smart reaction reflects the reaction ability of EMMS when variety of production needed by the military changes. The stronger the reaction ability is, the shorter time needed is.

4.2.7 Satisfaction Degree of Information Feedback and Maintenance

Information maintenance, which is a index reflecting the satisfaction degree of user about information maintenance done by the supplier, is one of the important content of after service of the supplier. With incessant advance of informationalized degree of current military equipment, the military user's information systems, acting as campaign flats, whose status and effect have no different from traditional equipment, also need maintenance and improve that is the meaning of information maintenance. The information feedback and satisfaction degree will be in a low level and the military will be dissatisfied if the supplier pays few attentions to information maintenance in after service, which makes military's information system upgrade lag, and the supplier doesn't answer the military user's production information reflection adequately, doesn't reworks exposed problems in time.

4.2.8 Feedback Degree of Military Demand

The index of feedback degree of military demand reflects the recognition degree of EMMS to military customer's demand. Military's needs will change when campaign circumstance changes. Military customer's problems will be settled in time if EMMS pay enough attendance to their demand. The index of feedback degree of

military demand expressed by the time quantum starting from new demand is put forward by the military customer and ending in settle program is sent to purchasing decision-making department of the military.

5 Tag

Comprehensive evaluation of equipment maintenance material suppliers is a decision-making issue of multi-object complex system and the important method and measure of military equipment supplier's management. EMMS's comprehensive evaluation, which importantly infects the whole maintenance material supply chain, must be scientific and objective. Construction of index system of comprehensive evaluation of EMMS just contributes to this.

References

1. Mao Z (2007) Suppliers evaluation and choice in the management of suppliers chain. Master Thesis of Hefei Industry University
2. Peng Y, Wu Q (2007) Establish process of military material suppliers partnership. China Logistics and Purchasing, vol 24, p 40
3. Wang J (2005) Proposition of military suppliers evaluation index system construction. National Defence Technology Base 5:1–4
4. Yu F (2007) Research on suppliers evaluation and choice based on supply chain management. Master Thesis of Wuhan Institute of Technology

Addressing Missing Data for Diagnostic and Prognostic Purposes

Panagiotis Loukopoulos, George Zolkiewski, Ian Bennett, Suresh Sampath, Pericles Pilidis, Fang Duan, and David Mba

Abstract One of the major targets in industry is minimising the downtime of a machine while maximising its availability, with maintenance considered as a key aspect towards achieving this objective. Condition based maintenance and prognostics and health management, which relies on the concepts of diagnostics and prognostics, is a policy that has been gaining ground over several years. The successful implementation of this methodology is heavily dependent on the quality of data used which can be undermined in scenarios where there is missing data. This issue may compromise the information contained within a data set, thus having a significant effect on the conclusions that can be drawn, hence it is important to find suitable techniques to address this matter. To date a number of methods to recover such data, called imputation techniques, have been proposed. This paper reviews the most widely used methodologies and presents a case study using actual industrial centrifugal compressor data, in order to identify the most suitable technique.

Keywords Missing data • Imputation techniques • Centrifugal compressor • Condition monitoring data

1 Introduction

One of the major targets in industry is the minimisation of downtime of a machine and the maximisation of its availability. Maintenance is considered a key aspect towards achieving this objectives, leading to various maintenance schemes being

P. Loukopoulos (✉) • S. Sampath • P. Pilidis
School of Aerospace, Transport and Manufacturing, Cranfield University, Cranfield, Bedfordshire MK43 0AL, UK
e-mail: p.loukopoulos@cranfield.ac.uk; s.sampath@cranfield.ac.uk; p.pilidis@cranfield.ac.uk

G. Zolkiewski • I. Bennett
Shell Global Solutions, Rijswijk, Netherlands
e-mail: george.zolkiewski@hotmail.com; Ian.Bennett@shell.com

F. Duan • D. Mba
School of Engineering, London South Bank University, 103 Borough Road, London SE1 0AA, UK
e-mail: duanf@lsbu.ac.uk; mbad@lsbu.ac.uk

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_17

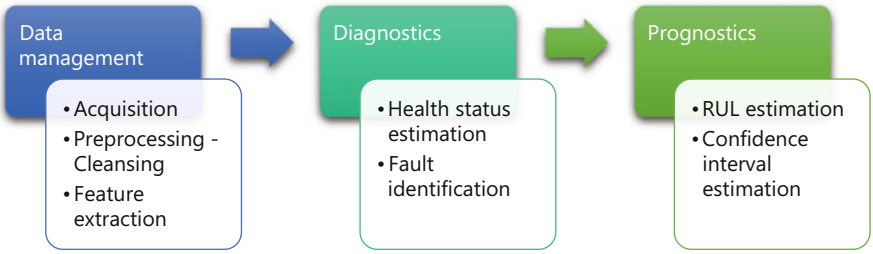


Fig. 1 CBM/PHM process

proposed over the years [1, 2]. Condition Based Maintenance and Prognostics and Health Management (CBM/PHM) [3], which is founded on the diagnostics and prognostics principles, has been increasingly popular over the recent years. It consists of three steps [3–6], and is presented in Fig. 1.

The success of CBM/PHM is heavily dependent on the quality of information used due to its sequential structure (Fig. 1), and missing data is one of the major issues that affect it. It is a frequent phenomenon in industry that can manifest in various ways with the most dominant being sensor failure [7]. While the amount of measurements recorded increases, so is the probability of missing data occurring. The presence of missing values may compromise the information contained within a set, introducing bias or misleading results [8, 9]. Consequently, as noted in [9]), it is important to address this matter. To deal with this problem, various methodologies to recover such data, called imputation techniques, have been developed. To the authors' knowledge, none of them has yet been applied to centrifugal compressor data. The purpose of this paper is to apply the most widely used imputation techniques in order to identify the most suitable technique.

2 Literature Review

Despite the lack of available literature on addressing centrifugal compressor missing data, a variety of imputation techniques have been developed and employed successfully in other fields like biological studies. [7, 9–27]. In [28]), they compared Bayesian Principal Component Analysis (BPCA) with Singular Value Decomposition (SVD) and K-Nearest Neighbours (KNN) imputation, on DNA microarray data where BPCA outperformed the other methods. In [18]), they studied the imputation performance of various Principal Components Analysis (PCA) methods, using artificial was well as actual data from Netflix. For the artificial information, BPCA outperformed each method though it was the most time consuming. In the case of high dimensional and sparse Netflix data, the best method for these data was a variation of BPCA, called BPCAd, which was created in order to deal with cases of high dimensional sparse data. In [26]), they presented a software package containing several PCA variations, applied to microarray data. BPCA was the best method while Probabilistic PCA (PPCA) was the fastest and is recommended when dealing with big data sets. In [29]), they applied

PPCA, BPCA, cubic spline interpolation and historical imputation to reconstruct missing values on traffic data, with BPCA being superior. In [12]) they compared the internal imputation offered by the classifiers C4.5 and CN2 with that of KNN and mean imputation, with KNN being the best. In [27]), they applied mean imputation, normal ratio method, normal ratio with correlation method, Multi-Layer Perceptron Neural Network (MLPNN) and Multiple Imputation (MI) to meteorological time series data sets. MI, although being the most computationally demanding, was more robust and outperformed the rest. In [19]), nearest/linear/cubic interpolation methods, regression based imputation, KNN, Self-Organising Maps (SOM), MLPNN, hybrid methods and MI, were compared on air quality data. MI despite being the slowest, offered the best results. In [30]), they compared SOM with linear regression and back propagation neural network when applied to water treatment time series data, with SOM being superior. In [31]), they applied KNN, SVD, and mean substitution on DNA microarray data, with KNN being the best.

3 Review of Imputation Techniques

As mentioned in the introduction, several imputation methods have been developed, though none of them has yet been implemented on centrifugal compressor data. The techniques employed in this work and reviewed below can be divided into two groups: univariate (ad hoc, interpolation, and time series methods) which are applied on a single variable and utilise information before and after the missing data for estimation, and multivariate (SOM, KNN, BPCA, and MI) which are applied on the complete set and utilise the variable interrelations to estimate the missing values.

Ad Hoc Method
Missing values are replaced with a fixed value [17]. Common fixed values are the mean (method 2), the median (method 3) ([12]; [17]), and the previous measured value which is carried forward (method 1) [17].

Interpolation Method
A curve is fitted along the missing data in an attempt to estimate the missing value. The methods used are: [19]: (i) nearest neighbour (method 4), (ii) linear (method 5), (iii) cubic (method 6).

Time Series Method
Observed information is used to train a model to predict missing values [32]. This method (method 7) can be enhanced with the combination of forward and backward prediction [33], where data before and after the missing measurements are used to train two separate models and then predict the missing values by averaging the two predictions through an iterative procedure. The model used was an autoregressive one [32], and was selected for its simple structure.

Self-Organising Map
SOM (method 8) is used to project multidimensional data into a two-dimensional structure ([34]; [19]; [30]) in a way so that data with similar patterns are associated

with the same neurons (Best Matching Unit—BMU) or their neighbours [30]. The map is constructed with the available information. Then the data with the missing values are fed to the map to calculate their BMUs. The missing measurements are estimated as their corresponding BMU values of the respective weight vectors.

K-Nearest Neighbours
Assuming a variable within a set contains a missing value, KNN imputation (method 9) uses the K other variables that don't have a missing value at the same time stamp and are most similar to that variable, to estimate the missing sample ([20]; [31]) The number of neighbours (K) used affects strongly the performance of this method.

Bayesian Principle Components Analysis
PCA is the linear projection (scores) of data where the retained variance is maximum (principal components) [35]. PPCA [36], is an extension where the principal components are assumed to have a prior distribution. BPCA (method 10) [35], is an extension where the optimum number of principal components is selected automatically.

Multiple Imputation
MI (method 11) ([14]; [37]; [24]) takes into account the uncertainty caused by the existence of missing data, by creating m complete sets ([17]; [23]; [24]; [25]). Usually, a small number of sets is adequate $m = 3 - 5$. The backbone of MI is data augmentation algorithm ([24]), a two-step iterative procedure where missing values are simulated.

4 Application of Imputation Techniques

There are three mechanisms of missing data [10–12, 14, 16, 21, 23–25, 29]: (i) missing completely at random, (ii) missing at random (MR), (iii) missing not at random. Most methods, can perform only under the assumption of the first two types [11, 14, 16, 21]. In this work MR type was assumed.

The information employed for this study was taken from an operational industrial centrifugal compressor. After a preliminary examination, it was observed that in 92% of the sets, the missing data had the form which can be found in Fig. 2. For a specific variable within a set there is a single group of missing values with observed data before and after it. This type of missing data is referred to in this paper as continuous missingness and was the focus of the project.

For analysis, a complete set of 474 samples containing 25 variables was selected. Five percentages of missing data were simulated: 1, 5, 15, 25 and 50%. At any time, only one variable within the set contained missing values, since as stated above, at any time only one variable presented missing data. For each percentage and the chosen variable, a sliding window with span equal to the percentage of missingness was translated into samples; started from the beginning of the data set, the window

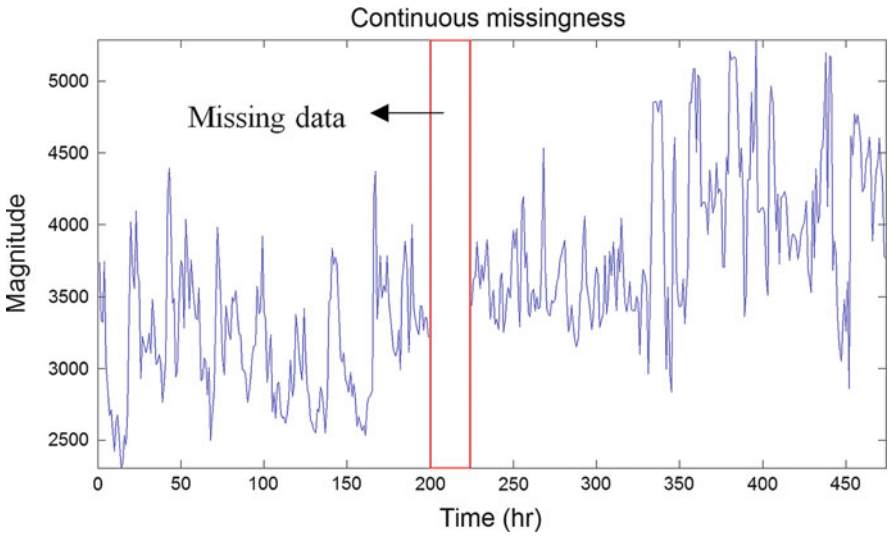


Fig. 2 Continuous missingness

slid across the signal and removed the respective samples, creating new sets. This way, the effect of the position of missing information on the quality of imputation was also considered. Hence, for each variable the new sets created were: 470 for 1%, 451 for 5%, 404 for 15%, 356 for 25% and 238 for 50%. In total, the number of sets created and analysed was $(470+451+404+356+238)^{*}25=47975$. In order to benchmark the performance of the imputation techniques the normalised root mean square error (NRMSE) between the predicted and actual values was employed, as given in [38]).

The results regarding the NRMSE for 50% of missingness are presented in Fig. 4. In the x-axis, ranging from 1–25, are the variables within the set while in the y, ranging from 1–11, are the methods used for imputation. The graph is separated in a number of boxes, which are the combination of each variable and each method. For example, box [1,9] corresponds to the results of applying method 1 to variable 9. Within each box there are several lines, indicating the location of the missing values. For 50%, as stated previously, there are 238 sets for each variable. From top to bottom, the 1st line (top) corresponds to a set where the missing data are found in the beginning, while the 238th line (bottom) corresponds to a set where the missing data are found in the end. These locations can be seen in Fig. 3. The lines are colour coded based on their NRMSE value. Although NRMSE ranges from $-\infty$ (no fit) to 1 (perfect fit), for scaling reasons the results range from -1, dark blue indicating poor performance, to 1, dark red yielding perfect estimation. For box [1,9], it can be observed that method 1 performs inadequately as it is filled with a variety of blue lines corresponding to low NRMSE values.

Going through the results, it is evident that multivariate methods were superior to univariate techniques employed as indicated by the high NRMSE values (red lines in the colour map of Fig. 4). It can be seen that regardless the percentage or the

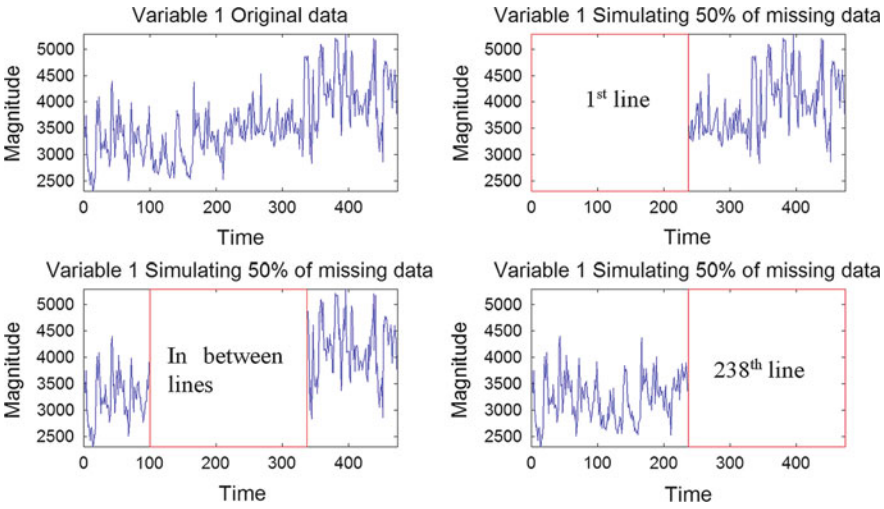


Fig. 3 Missing data location for 50% missingness

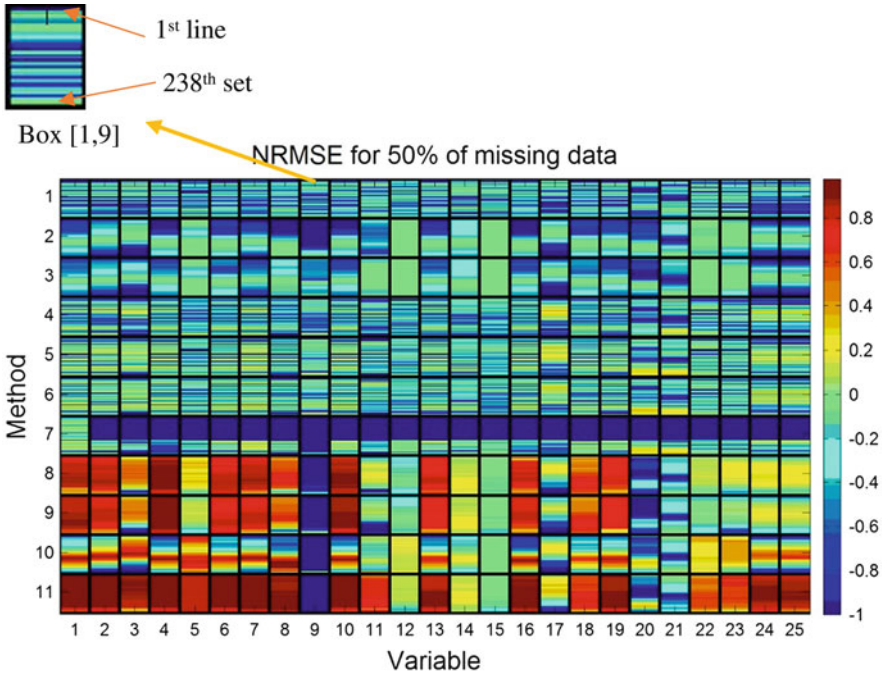


Fig. 4 NRMSE results for 50% missingness

location of missing information, MI was superior, followed by self-organising maps and k-nearest neighbours. It can be noted though that their performance was highly related to the selected variable they were applied to. Despite having a robust performance for most variables, it can be seen that for some others (9, 17, 20 and 21) their boxes are blue (Fig. 4) indicating poor relative performance. This observation was also noted for other techniques, as they were also affected either by position or percentage of missingness.

5 Conclusions

Missing data is an important issue that needs to be resolved in order to apply CBM/PHM successfully, with imputation being a common solution. There are various imputation techniques available but none of them has yet been applied to centrifugal compressor data. According to the results, it has been shown that the best and most robust method is multiple imputation followed by self-organising maps and k-nearest neighbours, but are highly dependent on the variables for which they are applied.

References

1. Kothamasu R, Huang SH, VerDuin WH (2006) System health monitoring and prognostics—a review of current paradigms and practices. Int J Adv Manuf Technol 28(9–10):1012–1024. doi:10.1007/s00170-004-2131-6
2. Lee J, Wu F, Zhao W, Ghaffari M, Liao L, Siegel D (2014) Prognostics and health management design for rotary machinery systems—reviews, methodology and applications. Mech Syst Signal Process 42(1–2):314–334. doi:10.1016/j.ymssp.2013.06.004
3. Vachtsevanos G, Lewis F, Roemer M, Hess A, Wu B (2006) Intelligent fault diagnosis and prognosis for engineering systems. Wiley, Hoboken, NJ. doi:10.1002/9780470117842
4. Jardine AKS, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech Syst Signal Process 20(7):1483–1510. doi:10.1016/j.ymssp.2005.09.012
5. Peng Y, Dong M, Zuo MJ (2010) Current status of machine prognostics in condition-based maintenance: a review. Int J Adv Manuf Technol 50(1–4):297–313. doi:10.1007/s00170-009-2482-0
6. Sikorska JZ, Hodkiewicz M, Ma L (2011) Prognostic modelling options for remaining useful life estimation by industry. Mech Syst Signal Process 25(5):1803–1836. doi:10.1016/j.ymssp.2010.11.018
7. Brown ML, Kros JF (2003) Data mining and the impact of missing data. Ind Manag Data Syst 103(8):611–621. doi:10.1108/02635570310497657
8. Pantanowitz A, Marwala T (2009) Evaluating the impact of missing data imputation. In: Advanced data mining and applications. Lecture Notes in Computer Science, vol 5678, pp 577–586. doi:10.1007/978
9. McKnight PE, McKnight KM, Souraya Sidani AJF (2007) Missing data: a gentle introduction. The Guilford Press, New York

10. Acock AC (2005) Working with missing values. J Marriage Fam 67(4):1012–1028. doi:10.1111/j.1741-3737.2005.00191.x
11. Baraldi AN, Enders CK (2010) An introduction to modern missing data analyses. J Sch Psychol 48(1):5–37. doi:10.1016/j.jsp.2009.10.001
12. Batista GEAPA, Monard MC (2003) An analysis of four missing data treatment methods for supervised learning. Appl Artif Intell. doi:10.1080/713827181
13. Donders a RT, van der Heijden GJMG, Stijnen T, Moons KGM (2006) Review: a gentle introduction to imputation of missing values. J Clin Epidemiol 59(10):1087–1091. doi:10.1016/j.jclinepi.2006.01.014
14. Enders CK (2001) A primer on maximum likelihood algorithms available for use with missing data. Struct Equ Model Multidiscip J 8(1):128–141. doi:10.1207/S15328007SEM0801_7
15. Enders CK (2010) Applied missing data analysis. The Guilford Press, New York
16. Graham JW (2009) Missing data analysis: making it work in the real world. Annu Rev Psychol 60:549–576. doi:10.1146/annurev.psych.58.110405.085530
17. Horton NJ, Kleinman KP (2007) Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models. Am Stat. doi:10.1198/000313007X172556
18. Ilin A, Raiko T (2010) Practical approaches to principal component analysis in the presence of missing values. J Mach Learn Res 11:1957–2000
19. Junninen H, Niska H, Tuppurainen K, Ruuskanen J, Kolehmainen M (2004) Methods for imputation of missing values in air quality data sets. Atmos Environ 38(18):2895–2907. doi:10.1016/j.atmosenv.2004.02.026
20. Li L, Li Y, Li Z (2014) Missing traffic data: comparison of imputation methods. IET Intell Transp Syst 8(1):51–57. doi:10.1049/iet-its.2013.0052
21. Little RJA, Rubin DB (2002) Statistical analysis with missing data, 2nd edn. Wiley, Hoboken, NJ. doi:10.1002/9781119013563
22. Myrtveit I, Stensrud E, Olsson UH (2001) Analyzing data sets with missing data: An empirical evaluation of imputation methods and likelihood-based methods. IEEE Trans Softw Eng 27(11):999–1013. doi:10.1109/32.965340
23. Pigott TD (2001) A review of methods for missing data. Educ Res Eval 7(4):353–383. doi:10.1076/edre.7.4.353.8937
24. Schafer JL (2000) Analysis of incomplete multivariate data. Chapman & Hall/CRC, London. doi:10.1201/9781439821862
25. Schafer JL, Graham JW (2002) Missing data: our view of the state of the art. Psychol Methods 7(2):147–177. doi:10.1037/1082-989X.7.2.147
26. Stacklies W, Redestig H, Scholz M, Walther D, Selbig J (2007) pcaMethods – a bioconductor package providing PCA methods for incomplete data. Bioinformatics 23(9):1164–1167. doi:10.1093/bioinformatics/btm069
27. Yozgatligil C, Aslan S, Iyigun C, Batmaz I (2013) Comparison of missing value imputation methods in time series: the case of Turkish meteorological data. Theor Appl Climatol 112(1–2):143–167. doi:10.1007/s00704-012-0723-x
28. Oba S, Sato M, Takemasa I, Monden M, Matsubara K, Ishii S (2003) A Bayesian missing value estimation method for gene expression profile data. Bioinformatics 19(16):2088–2096. doi:10.1093/bioinformatics/btg287
29. Qu L, Li L, Zhang Y, Hu J (2009) PPCA-based missing data imputation for traffic flow volume: a systematical approach. IEEE Trans Intell Transp Syst 10(3):512–522. doi:10.1109/TITS.2009.2026312
30. Rustum R, Adeloye AJ (2007) Replacing outliers and missing values from activated sludge data using Kohonen self-organizing map. J Environ Eng. http://doi.org/10.1061/(ASCE)0733-9372(2007)133:9(909)
31. Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB (2001) Missing value estimation methods for DNA microarrays. Bioinformatics 17(6):520–525. doi:10.1093/bioinformatics/17.6.520

32. Ljung L (1999) System identification: theory for the user, 2nd edn. Prentice Hall, Englewood Cliffs
33. Moahmed TA, Gayar NE, Atiya AF (2014) Forward and backward forecasting ensembles for the estimation of time series missing data. Lect Notes Comput Sci 8774:93–104. doi:10.1007/978-3-642-12159-3
34. Folguera L, Zupan J, Cicerone D, Magallanes JF (2015) Self-organizing maps for imputation of missing data in incomplete data matrices. Chemom Intel Lab Syst 143:146–151. doi:10.1016/j.chemolab.2015.03.002
35. Bishop CM (1999) Variational principal components. In: 9th International conference on artificial neural networks ICANN 99, No. 470, pp 509–514. doi:10.1049/cp:19991160
36. Tipping ME, Bishop CM (1999) Probabilistic principal component analysis. J R Stat Soc Series B Stat Methodology 61(3):611–622. doi:10.1111/1467-9868.00196
37. Harel O, Zhou X-H (2007) Multiple imputation: review of theory, implementation and software. Stat Med 26(16):3057–3077. doi:10.1002/sim.2787
38. Ljung L (2015) System identification toolbox ™ user β€™ s guide. The MathWorks, Inc

Imperfect Coverage Analysis for Cloud-RAID 5

Lavanya Mandava, Liudong Xing, and Zhusheng Pan

Abstract Existing works on reliability modeling of cloud systems have assumed perfect fault detection and recovery mechanisms. In this paper, we relax this assumption by presenting combinatorial approaches based on binary decision diagrams for reliability analysis of a cloud-RAID 5 system subject to imperfect fault coverage. Both element level coverage and fault level coverage are addressed. Numerical example results are provided to illustrate effects of those two types of imperfect fault coverage on the reliability performance of cloud storage systems.

Keywords Cloud storage system • Cloud computing • Reliability • Cloud-RAID • Imperfect fault coverage • Element level coverage • Fault level coverage

Acronyms

BDD	Binary decision diagram
ELC	Element level coverage
FLC	Fault level coverage
IFC	Imperfect fault coverage
ite	if-then-else
RAID	Redundant array of independent disks

L. Mandava • L. Xing (✉)
Department of Electrical and Computer Engineering, University of Massachusetts Dartmouth, 285 Old Westport Road, North Dartmouth, MA 02747, USA
e-mail: lxing@umassd.edu

Z. Pan
College of Mathematics, Physics and Information Engineering, Zhejiang Normal University, Jinhua 321004, China

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_18

1 Introduction

A cloud storage system is a network consisting of remote servers accessible through the Internet, used to store, manage and process data [1, 2]. Many companies provide such cloud services, including for example Google Drive, Dropbox, and Amazon EC2. Cloud storage enables its users to enjoy on-demand high quality applications and services from a shared pool of configurable computing resources, without the burden of local data storage and maintenance [3].

Cloud users expect to be able to access their files anytime anywhere from the cloud. Any interruption in service or system failure can impact the reputation of cloud service providers in a serious and negative way. For example, in a reported breakdown of cloud computing, 8 h of unexpected outage of simple storage service on the Amazon Cloud storage (Amazon S3) affected services of numerous users [4]. Occurrences and impacts of such failures reflect urgency and significance of addressing reliability issues for the cloud storage systems.

Cloud-RAIDs (Redundant Array of Independent Disks) based on various data redundancy management techniques [5, 6] are one of the solutions to achieve high data reliability [7–10]. The redundancy management mechanism of a system is responsible for fault detection, fault isolation and system reconfiguration in the event of component fault happening. In real-world systems, the redundancy management task can seldom be done perfectly, and such systems are referred to as systems with imperfect fault coverage [11].

In [8, 9] a hierarchical modeling methodology was developed, which integrates a Markov model at the lower level for considering physical failure behaviors of individual disks, and a multi-valued decision diagram model at the higher level for evaluating state probabilities of the entire cloud-RAID storage system. The method is applicable to heterogeneous disks from different cloud storage providers. However, the method of Liu and Xing [8, 9] has assumed perfect fault coverage for the considered cloud-RAID system.

In this paper, we extend the work of Liu and Xing [8, 9] by incorporating effects of imperfect fault coverage in the reliability analysis of a cloud-RAID 5 system with heterogeneous disks. Two types of imperfect coverage models, element level coverage and fault level coverage (explained in Sect. 3) are considered. Combinatorial approaches based on binary decision diagrams are presented.

The remainder of the paper is organized as follows. Section 2 describes the cloud-RAID 5 system to be modeled in this work. Section 3 presents some preliminary concepts and methods. Section 4 presents the proposed combinatorial methods for reliability analysis of cloud-RAID 5 system subject to element level coverage and fault level coverage. Section 5 presents numerical analysis results and discussions. Lastly, Sect. 6 concludes the paper.

2 Cloud-RAID 5 System

There are seven levels in the traditional RAID [12] architecture. While the proposed methodology is theoretically applicable to all the levels, we use the cost-effective level 5 to illustrate the cloud-RAID architecture and the proposed reliability evaluation methods.

In RAID 5, data are divided into blocks and are striped across multiple disks that form an array [13]. Distributed parity is used, where the parity stripes are distributed across all the disks. In general a RAID architecture can be implemented using disk drives with different sizes. The total usable storage space, however, is limited by the disk with the smallest size.

Figure 1 shows the general structure of an example cloud-RAID 5 system, where the data are split into blocks or stripes with parity stripes being distributed across five disk drives from different cloud providers. The blocks on the same row of the array form a virtual disk drive with one of them storing the parity information, which provides data redundancy and thus fault tolerance. If any disk drive (or cloud storage provider) fails or is not available, the entire storage system can still work due to the use of data redundancy. Particularly, the failed or unavailable stripe can be recovered using the parity stripe and remaining data stripes through, for example, the exclusive OR operation.

In the context of reliability analysis, the example cloud-RAID 5 system can be modeled using a 4-out-of-5 model, where the system functions or is reliable if at least four disks are functioning correctly.

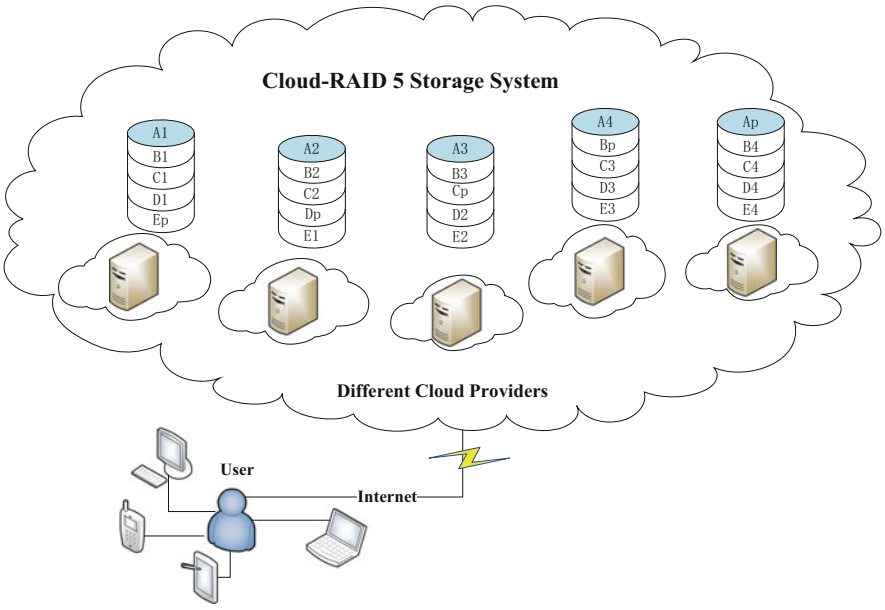


Fig. 1 Architecture of an example cloud-RAID 5 (adapted from Liu and Xing [8, 9])

3 Preliminary Concepts and Methods

In this section we explain the imperfect fault coverage, and binary decision diagram (BDD) models.

3.1 Imperfect Fault Coverage (IFC)

Models that consider effects of imperfect fault coverage are known as imperfect fault coverage models or simply coverage models [11]. There are three types of coverage models depending on fault tolerant techniques adopted [14]: element level coverage (ELC) or single-fault model, fault level coverage (FLC) or multi-fault model, and performance dependent coverage. This paper mainly concentrates on reliability analysis of the cloud-RAID 5 system with ELC and FLC, which are explained below.

With ELC, the fault coverage probability of an element is independent of states of other system elements. The ELC model is applicable or appropriate when the selection among the redundant elements is made based on self-diagnostic capability of individual elements. Effectiveness of the system recovery mechanism relies on the occurrence of individual element faults. Under ELC, a multi-element system can tolerate multiple co-existing single element faults. However, for any given element fault, the success or failure of the recovery mechanism (i.e., the coverage probability) is not dependent on faults in other elements.

Under FLC, the fault coverage probability is dependent on the number of failed elements belonging to a particular group. The FLC model is applicable or appropriate when the selection among available system elements varies between initial and subsequent element failures. Effectiveness of the system recovery mechanism relies on the occurrence of multiple element faults within a certain recovery window. The FLC model is typically applied in modeling, for example, computer control systems used in aircrafts applications [15, 16], and multi-processor systems in a load-sharing environment [17].

3.2 Binary Decision Diagram (BDD)

BDD is the state-of-the-art data structure for Boolean logical function representation and manipulation. A BDD is a rooted, directed acyclic graph model based on Shannon decomposition rule of (1) [11, 18].

$$f = xf_{x=1} + \bar{x}f_{x=0} = xF_1 + \bar{x}F_0 \tag{1}$$

In (1), f represents a Boolean logic function on a set of Boolean variables with x being one of them. Equation (1) can also be represented using the compact *ite* (if-then-else) format as $f = ite\ (x, F_1, F_0)$.

A BDD has two sink nodes '1' and '0' representing failure and function of the considered system, respectively. There also exists a set of non-sink nodes, each corresponding to a system component. Each non-sink node in a BDD model encodes an *ite* construct, as illustrated in Fig. 2. Each non-sink node has two outgoing edges: '0'-edge/else-edge leading to child node $f_{x=0}$ and '1'-edge/then-edge leading to child node $f_{x=1}$ [18, 19]. The probability associated with the then edge is $q\ (x)$ (component unreliability); the probability associated with the else edge is $p(x)$ (component reliability); and $p(x) + q(x) = 1$.

Equation (2) gives the manipulation rules for BDD generation [16]:

$$g \diamondsuit h = ite(x, G_1, G_0) \diamondsuit ite(y, H_1, H_0)$$
$$= \begin{cases} ite(x, G_1 \diamondsuit H_1, G_0 \diamondsuit H_0) & index(x) = index(y) \\ ite(x, G_1 \diamondsuit h, G_0 \diamondsuit h) & index(x) < index(y) \\ ite(y, g \diamondsuit H_1, g \diamondsuit H_0) & index(x) > index(y) \end{cases} \quad (2)$$

In (2), g and h represent two Boolean functions, and $\diamondsuit$ represents a logical operation ('AND' or 'OR'). The *index* represents the order of the Boolean variable in the input ordering list. The rules are used for combining two BDD models represented by g and h into one BDD model. In applying the rules, the indices of two root nodes (i.e., x for g and y for h) are compared. If x and y have the same index meaning that they belong to the same component, then the logic operation is applied to their children nodes; otherwise, the node with a smaller index variable becomes the root node of the newly combined BDD model and the logic operation is applied to each child of the smaller index node and the other BDD model as a whole. The rules are recursively applied for operations between sub-expressions until at least one of them becomes a constant '0' or '1'.

Once the system BDD model is generated, the system unreliability (reliability) can be evaluated as the sum of probabilities of all disjoint paths from the root to sink node '1' ('0').

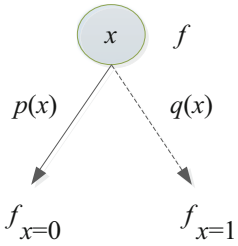


Fig. 2 A non-sink node in the BDD model

4 Combinatorial Methods

In this section we present combinatorial approaches for evaluating reliability of the cloud-RAID 5 systems considering ELC and FLC.

4.1 Cloud-RAID 5 with ELC

We use an ELC model representing the behavior of the system in response to the occurrence of a component fault, as shown in Fig. 3.

The entry point to the ELC model signifies the occurrence of a component fault, and the three exits correspond to three possible outcomes [20]. Transient restoration exit (R) is taken when the fault is transient and can be treated without discarding the component. Permanent coverage exit (C) is taken when the fault nature is permanent, and the faulty component must be discarded. Single-point failure exit (S) is taken when the component fault (by itself) causes the entire system to fail. The three exits form a partition of the event space, thus the three exit probabilities (r, c, and s) sum to one.

Let $q_d(t)$ denote the failure probability of disk d ($d = 1, 2, 3, 4, 5$) at time t. In the case of disk d following the exponential distribution with parameter λ_d, we have $q_d(\mathrm{t}) = 1-\exp(-\lambda_d{*}t)$. Based on the ELC model, we have the probability that disk d does not fail (denoted by $n(d)$), fails covered ($c(d)$), and fails uncovered ($u(d)$) as

$$n[d] = 1 - q_d(t) + q_d(t){*}r_d \tag{3}$$

$$c[d] = q_d(t){*}c_d \tag{4}$$

$$u[d] = q_d(t){*}s_d \tag{5}$$

Based on the simple and efficient algorithm in [20, 21], we define two events for the considered cloud-RAID 5 system:

E_1: no disks experience a single-point of failure
E_2: at least one disk experiences a single-point of failure.

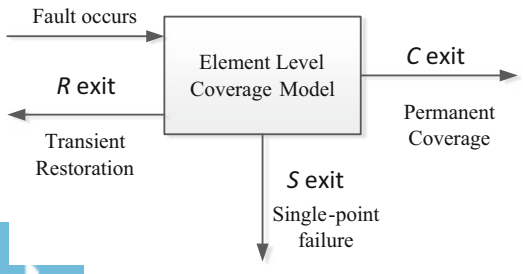


Fig. 3 General structure of an ELC model [20]

Define E as an event that the system fails. Based on the total probability law, the unreliability of the cloud-RAID 5 system with ELC can be obtained as.

$$U_{ELC} = \Pr(E|E_1)*\Pr(E_1) + \Pr(E|E_2)*\Pr(E_2) \tag{6}$$

where $\Pr(E_1) + \Pr(E_2) = 1$ and $\Pr(E|E_2) = 1$. Thus (6) can be simplified as

$$U_{ELC} = 1 - \Pr(E_1) + \Pr(E|E_1)*\Pr(E_1) \tag{7}$$

In (7), $\Pr(E_1)$ for the example cloud-RAID 5 system can be calculated as

$$\Pr(E_1) = (1 - u[1])(1 - u[2])(1 - u[3])(1 - u[4])(1 - u[5]) \tag{8}$$

To evaluate $\Pr(E|E_1)$ in (7), we evaluate a conditional failure probability for each disk d as

$$\tilde{q_d} = c[d]/(1 - u[d]) \tag{9}$$

Then, $\Pr(E|E_1)$ can be evaluated using the BDD method without considering the ELC.

Figure 4 illustrates the BDD model for evaluating $\Pr(E|E_1)$ of the example cloud-RAID 5 system, which has a well-defined 4-out-of-5 lattice structure. The solid edge represents that the corresponding disk is working and the dashed edge represents the disk is failed. Sink nodes '0' and '1' represent the system working and failed states, respectively.

Based on the generated BDD model, $\Pr(E|E_1)$ can be evaluated as

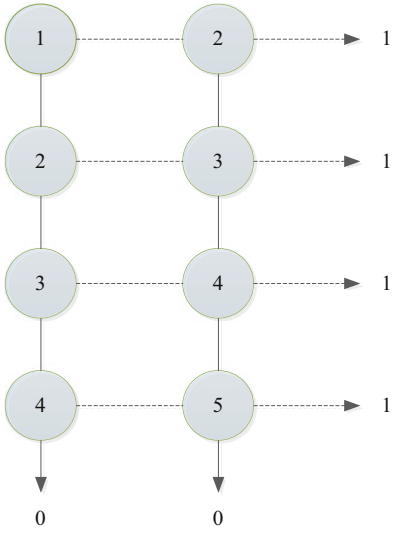


Fig. 4 Lattice structure BDD of the example cloud-RAID 5 with ELC

$$\begin{aligned}\Pr(E|E_1) = \; & \tilde{q_1}\tilde{q_2} + \tilde{q_1}\left(1-\tilde{q_2}\right)\tilde{q_3} + \tilde{q_1}\left(1-\tilde{q_2}\right)\left(1-\tilde{q_3}\right)\tilde{q_4} \\ & + \tilde{q_1}\left(1-\tilde{q_2}\right)\left(1-\tilde{q_3}\right)\left(1-\tilde{q_4}\right)\tilde{q_5} + \left(1-\tilde{q_1}\right)\tilde{q_2}\,\tilde{q_3} \\ & + \left(1-\tilde{q_1}\right)\tilde{q_2}\left(1-\tilde{q_3}\right)\tilde{q_4} + \left(1-\tilde{q_1}\right)\tilde{q_2}\left(1-\tilde{q_3}\right)\left(1-\tilde{q_4}\right)\tilde{q_5} \\ & + \left(1-\tilde{q_1}\right)\left(1-\tilde{q_2}\right)\tilde{q_3}\,\tilde{q_4} + \left(1-\tilde{q_1}\right)\left(1-\tilde{q_2}\right)\tilde{q_3}\left(1-\tilde{q_4}\right)\tilde{q_5} \\ & + \left(1-\tilde{q_1}\right)\left(1-\tilde{q_2}\right)\left(1-\tilde{q_3}\right)\tilde{q_4}\,\tilde{q_5}\end{aligned} \tag{10}$$

With $\Pr(E_1)$ and $\Pr(E|E_1)$ the unreliability of the cloud-RAID 5 system considering effects of ELC can be given by (7). The reliability of the cloud-RAID 5 system can thus be obtained as

$$R_{\mathrm{ELC}} = 1 - U_{\mathrm{ELC}} \tag{11}$$

4.2 Cloud-RAID 5 with FLC

Under FLC, the fault coverage probability c_i $c_i$depends on the number of failed elements i that belong to a specific group. Thus, calculation of c_i depends on a set of disks that already failed. In general for a k-out-of-n system, the system fails after $(n-k+1)$ failures (irrespective of fault coverage), thus c_i for $i \geq (n-k+1)$ are not applicable and can be considered as zero [17]. For the example Cloud-RAID 5 system, $n = 5$ and $k = 4$. Thus only c_1 is used. By definition, c_0 is 1.

According to [17], Eq. (12) gives a method for evaluating c_i for systems with n identical components following the exponential time-to-failure distribution with constant failure rate λ.

$$c_i = \exp[-(n\text{-}i)*\lambda*\tau] \tag{12}$$

where i represents the failure number, τ represents the recovery window time.

Applying the BDD method with the consideration of FLC, we have the BDD model as shown in Fig. 5 where only paths to sink node '0' (representing the function of the example cloud-RAID 5 system) are shown. Coverage probabilities $c_0 = 1$ and c_1 are added to the corresponding paths of the BDD model. Based on the BDD model, the example Cloud-RAID 5 system remains working if at least four disks are working. If more than two disks fail, then the entire system fails. Whenever a single disk is in a failed state it can be covered with a coverage factor c_1 and the system is still functioning.

It is assumed that all the five disks of the example cloud-RAID 5 system have the same failure rate and recovery window time so that they have the same coverage factor c_1 based on (12). In the case of disks with non-identical failure rates (in general, failure time distributions) or recovery window time, Eq. (12) needs to

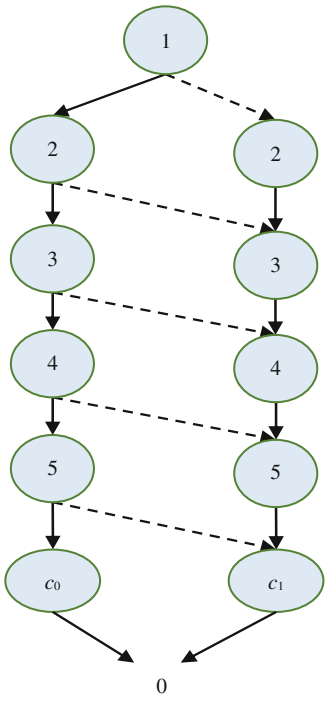


Fig. 5 BDD of the example Cloud-RAID 5 with FLC

be modified to consider a different reliability evaluation for a different component based on its time-to-failure distribution function. The BDD in Fig. 5 should also be expanded to associate a different coverage factor c_1 for paths involving a different single disk failure.

Based on the BDD generated in Fig. 5, the reliability of the example cloud-RAID 5 system with FLC can be evaluated as the sum of probabilities of all disjoint paths from the root to sink node '0' as

$$\begin{aligned} R_{FLC} &= (1-p_1)p_2p_3p_4p_5c_1 + p_1(1-p_2)p_3p_4p_5c_1 + p_1p_2(1-p_3)p_4p_5c_1 \\ &\quad + p_1p_2p_3(1-p_4)p_5c_1 + p_1p_2p_3p_4(1-p_5)c_1 + p_1p_2p_3p_4p_5 \end{aligned} \tag{13}$$

5 Reliability Analysis Results

This section presents numerical evaluation results for the reliability of the example cloud-RAID 5 system to demonstrate effects of ELC and FLC.

5.1 Results Considering ELC

Table 1 lists different combinations of r, c and s values and corresponding system reliability R_{ELC} for $t = 1000$ h calculated using the method presented in Sect. 4.1.

Note that while the method in Sect. 4.1 is applicable to non-identical disks with arbitrary types of time-to-failure distributions, we assume that the five disks follow the same exponential distribution with $\lambda_d = 0.0001$/h for this example illustration. The failure probability of each disk is thus $q_d(t) = 1-\exp(-\lambda_d * t)$.

In Table 1, when $r = 1$ (all faults are transient and can be covered without discarding any disk), based on (3), the disk reliability is 1, and thus the entire system reliability is 1. The system reliability is lowest when $s = 1$ as any disk fault causes the entire system to fail. In this case, the system is reduced to a purely series system. The case when $c = 1$ corresponds to the system having perfect fault coverage.

Table 2 presents system reliability values evaluated for different time t (in hours) under three different coverage factor combinations. As time proceeds, the system reliability decreases. The smaller the value of s is, the more reliable the disks and thus the entire system are. For the same value of s the system reliability is higher when the value of factor r is higher. Figure 6 illustrates the results of Table 2 graphically.

Table 1 Reliability results considering ELC

c	r	s	R_{ELC}
0	1	0	1
1	0	0	0.925477591
0	0	1	0.60653066
0	0.7	0.3	0.865177073
0	0.5	0.5	0.783681491
0	0.3	0.7	0.708446177
0.7	0	0.3	0.829793512
0.5	0	0.5	0.766004126
0.3	0	0.7	0.702214739
0.7	0.3	0	0.961247737
0.5	0.5	0	0.97943876
0.3	0.7	0	0.99230515
0.5	0.3	0.2	0.88901872
0.5	0.2	0.3	0.846380955
0.3	0.5	0.2	0.901135853
0.3	0.2	0.5	0.777052814
0.2	0.5	0.3	0.861984399
0.2	0.3	0.5	0.780675194

Table 2 Reliability results for different mission times

t	$R(c = 0.5, r = 0.3, s = 0.2)$	$R(c = 0.3, r = 0.5, s = 0.2)$	$R(c = 0.2, r = 0.3, s = 0.5)$
0	1	1	1
1000	0.88901872	0.901135853	0.780675194
2000	0.77075372	0.807797596	0.612740189
3000	0.658444637	0.722740086	0.483892622
5000	0.470841798	0.580140931	0.308164461
10,000	0.206537374	0.357127704	0.114932349
15,000	0.103873003	0.248557192	0.053403759
20,000	0.062065381	0.19363208	0.030576284

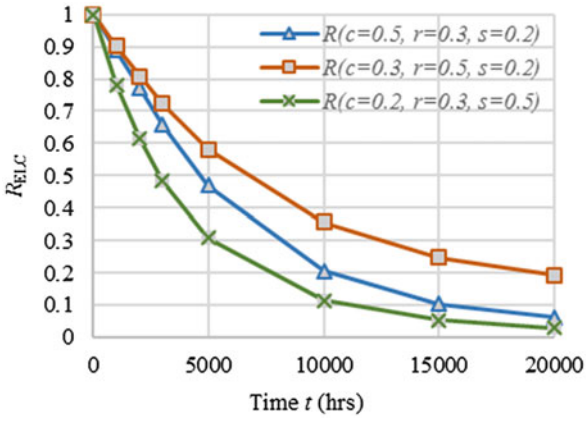


Fig. 6 Reliability of Cloud-RAID 5 system with ELC

5.2 *Results Considering FLC*

The reliability of the example Cloud-RAID 5 system with FLC is given by (13). Equation (12) is used to compute c_1, where λ of all the disks is assumed to be 0.0001/h. Table 3 lists values of c_1 calculated for various values of τ (in hours) and corresponding system reliability R_{FLC} evaluated for $t = 1000$ h using (13). Figure 7 plots the coverage probabilities as the recovery window time increases. Figure 8 illustrates the system reliability trend as the coverage probability increases.

From Table 3 we can observe that when the value of c_1 is 1, the system reliability is the highest and is actually equal to the reliability of system with perfect fault coverage. As shown in Table 3 and Fig. 8, as the coverage probability decreases, the system reliability gets worse.

Table 3 Reliability of Cloud-RAID 5 with FLC

τ	c_1	R_{FLC}
0	1	0.925477591
1000	0.670320046	0.820327182
2000	0.449328964	0.749842754
3000	0.301194212	0.702595629
5000	0.135335283	0.649695433
10,000	0.018315639	0.612372377
15,000	0.002478752	0.60732125
20,000	0.000335463	0.606637654

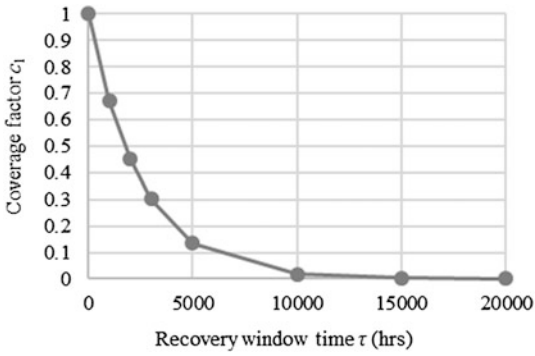


Fig. 7 Coverage factor vs. recovery window time

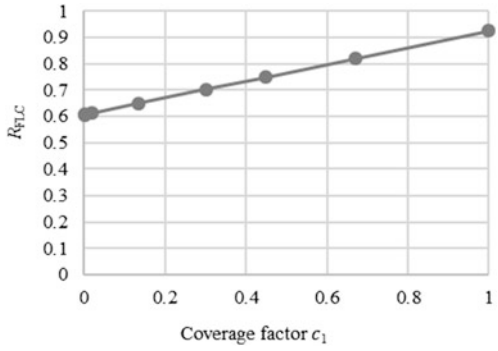


Fig. 8 Reliability of Cloud-RAID 5 with FLC

6 Conclusion

Cloud-RAID 5 based on single-bit parity code is one of the effective solutions for enhancing data reliability in the cloud storage. The existing reliability modeling methods for the cloud-RAID systems assumed that the system fault detection and recovery mechanism is perfect, which is often not true in practice. This paper makes new contributions by relaxing this assumption through BDD-based combinatorial

methods for the reliability analysis of cloud-RAID 5 subject to ELC or FLC. The methods are applicable to heterogeneous disks with arbitrary types of time-to-failure distributions. As demonstrated through numerical results, failure to consider the imperfect fault coverage may lead to inaccurate (overestimated) system reliability results, which can mislead the system design and optimization activities.

References

1. Deng J, Huang S C H, Han Y S, Deng J H (2010) Fault tolerant and reliable computation in cloud computing. In: Proceedings of IEEE Globecom workshops, Miami, FL, pp 1601–1605
2. Erl T, Puttini R, Mahmood Z (2013) Cloud computing concepts, technology & architecture, the Prentice Hall service technology series. Prentice Hall, Upper Saddle River, NJ
3. Wang C, Xing L, Wang H, Dai Y, Zhang Z (2014) Performance analysis of media cloud-based multimedia systems with retrying fault-tolerance technique. IEEE Syst J Spl Iss Recent Adv Cloud-Based Multimedia Syst 8(1):313–321
4. Robinson G, Narin A, Elleman C (2013) Using Amazon web services for disaster recovery. Amazon web services
5. Bausch F (2014) Cloud-RAID concept. http://blog.fbausch.de/cloudraid-3-concept/. Accessed May 2016
6. Jin T, Yu Y, Xing L (2009) Reliability analysis of RAID systems using repairable k-out-of-n modeling techniques. In: The international conference on the interface between statistics and engineering, Beijing, China
7. Fitch D, Xu H (2013) A RAID-based secure and fault-tolerant model for cloud information storage. Int J Softw Eng Knowl Eng 23(5):627–654
8. Liu Q, Xing L (2015a) Hierarchical reliability analysis of multi-state Cloud-RAID storage system. In: Proceedings of international conference on quality, reliability, risk, maintenance, and safety engineering, Beijing, China
9. Liu Q, Xing L (2015b) Reliability modeling of cloud-RAID-6 storage system. Int J Future Comput Commun 4(6):415–420
10. Zhang R, Lin C, Meng K, Zhu L (2013) A modeling reliability analysis technique for cloud storage system. In: Proceedings of 15th IEEE international conference on communication technology, Guilin, China, pp 32–36
11. Myers A (2010) Complex system reliability, 2nd edn. Springer series in reliability engineering
12. Jin T, Xing L, Yu Y (2011) A hierarchical Markov reliability model for data storage systems with media self-recovery. Int J Reliab Qual Saf Eng 18(1):25–41
13. Patterson D A, Chen P, Gibson G, Katz R H (1989) Introduction to Redundant Arrays of Inexpensive Disks (RAID). In: Proceedings of thirty-fourth IEEE computer society international conference: intellectual leverage, Digest of papers, San Francisco, CA, USA, pp 112–117
14. Amari S V, Myers A and Rauzy A (2007) An efficient algorithm to analyze new imperfect fault coverage models. In: Proceedings of annual reliability and maintainability symposium
15. Myers A (2007) k-out-of-n: G system reliability with imperfect fault coverage. IEEE Trans Reliab 56:464–473
16. Myers A, Rauzy A (2008) Assessment of redundant systems with imperfect coverage by means of binary decision diagrams. Reliab Eng Syst Saf 93(7):1025–1035
17. Amari SV, Myers A, Rauzy A, Trivedi K (2008) Imperfect coverage models: status and trends. In: Misra KB (ed) Handbook of performability engineering. Springer, Berlin
18. Xing L and Amari S V (2015) Binary decision diagrams and extensions for system reliability analysis, Wiley-Scrivener, MA, isbn:978-1-118-54937-7

19. Xing L, Wang H, Wang C, Wang Y (2012) BDD-based two-party trust sensitivity analysis for social networks. Int J Secur Netw 7(4):242–251
20. Amari SV, Dugan JB, Misra RB (1999) A separable method for incorporating imperfect fault-coverage into combinatorial models. IEEE Trans Reliab 48:267–274
21. Xing L, Dugan JB (2002) Analysis of generalized phased mission system reliability, performance and sensitivity. IEEE Trans Reliab 51(2):199–211

Research on Data Analysis of Material System Based on Agent Simulation

Ying Shen, JunHai Cao, HaiDong Du, and FuSheng Liu

Abstract There are two sides of system reliability parameters, such as basic reliability and mission reliability, for reliability requirement in the standard, with fine relevance and coordination. It reflects in the determination of the parameters and index and also in the prediction and allocation for basic reliability and mission reliability. In the paper, it sets up a fine mission reliability simulation model of materiel system, and analyzes from data collection and process, to lay the good foundation for analyzing dynamically mission reliability and making a perfect mission reliability simulation concept of materiel system.

Keywords System reliability • Mission reliability • Data collection • Simulation

1 Introduction

There are two sides of system reliability parameters, such as basic reliability and mission reliability, for reliability requirement in the standard, with fine relevance and coordination. It reflects in the determination of the parameters and index and also in the prediction and allocation for basic reliability and mission reliability. So it emphasizes the research on the analysis of mission reliability of materiel system to better the analysis of reliability simulation and settle the complexity of analysis of reliability simulation [1].

2 Data Collection and Processing of Adaptive Agent

Adaptive Agent can gather relative reliability data of itself with independent module of data gathering and processing, and computers statistically and shows in the chart. The metadata main includes:

- The failure times of a component: The failure times in the mission time for the working unit, as UF_M.

Y. Shen (✉) • J. Cao • H. Du • F. Liu
Academy of Armored Force Engineering, Beijing 100072, China
e-mail: sheny312@sina.com.cn

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_19

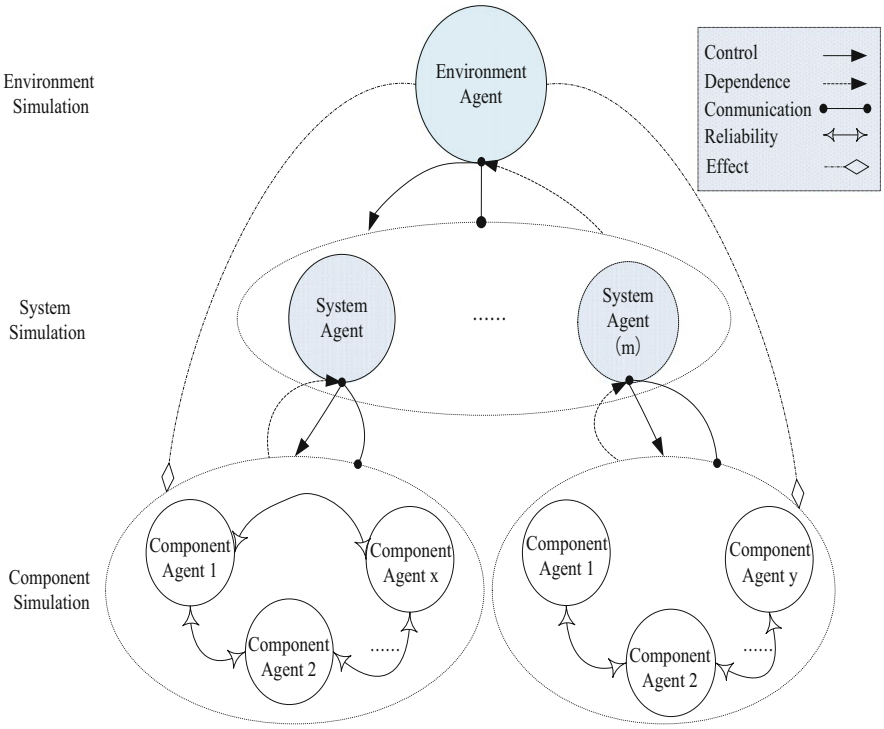


Fig. 1 The model of system structure and relationship of materiel system simulation

- Single accidental shutdown of a component: The time is when the failure unit is from into the state of failure to leave the state of failure in the time of mission.
- Standby time of a component: The time is when the working unit is in the state of standby in the time of mission.
- Single-time-to-repair of a component: The time is when the working unit is in the state of maintenance shown in Fig. 1.

On the basis of metadata, Adaptive Agent processes data statically and outputs in some forms. So there is main output as follows:

- Failure rate of a component, as λ_M: It is the basic parameter on the product reliability. It is the ratio of the total of failure time to the total of mission time in the simulation course of mission reliability for the working unit.

$$\lambda_M = \frac{\sum_{i=1}^{UF_M} UT_{Mi}}{MT} \tag{1}$$

Thereinto:

λ_M—Failure rate of a component in mission profile.
UT_{Mi}—unit downtime of a component.
UF_M—Number of unit failure.
MT—Mission time.

3 Data Collection and Processing of System Simulation Agent

The data of data collecting module of system simulation Agent is mainly debugging and analytical. Analytical data includes mainly the following metadata:

- Single standby time of a system: The time is when the system in the state of standby in the time of mission.
- Single-time-to-repair of a system: The time is when the system is from the start state of maintenance to repair faults.
- The number of maintenance: There are the times for a system in the state of maintenance.
- Single-time-to-repair of a critical failure of a system: The time is when the system is from the start state of maintenance to repair a critical failure.
- The number of critical failures of a component: There are the times for which a component takes occur critical failures.
- The number of critical failures of a system, as SF_M: There are the times for which a system takes occur critical failures, as the summation of critical failures of every components.

On the basis of metadata, System Simulation Agent processes data statically and outputs in some forms. The main output is as follows:

- Mean Time Between Critical Failures of a system as $MTBCF$: It is the parameter relative to mission reliability. It is the ratio of the total of mission time to the total of failure in the series of mission profiles.
  In the simulation course of mission reliability for the working unit.

$$\boldsymbol{MTBCF} = \frac{\boldsymbol{MT}}{\boldsymbol{SF_M}} \tag{2}$$

Thereinto:
$MTBCF$—Mean Time Between Critical Failures of a system.
MT—Mission Time.
SF_M—Number of critical failure of a system.

Mission time to mean recovery function of a system, as $MTTRF$: In the simulation course of the mission profile, it is the mean time for excluding critical failures. It is a kind of maintenance parameter relative to mission success. In the given mission profile and maintenance condition, it is calculated as follows:

$$MTTRF = \frac{\sum_{i=1}^{SF_M} STTR_{Mi}}{SF_M} \tag{3}$$

Thereinto:
MTTRF—Mission time to mean recovery function of a system.
$STTR_M$—The time of repairing a critical failure.
SF_M—Number of critical failure of a system.

Operational Availability of a system, as A_o: It is a kind of availability parameter relative to up time and down time in the simulation course of a system. It is calculated as follows:

$$A_o = \frac{\sum_{i=1}^{N} WTBF_i + \sum_{k=1}^{U} ST_k}{\sum_{i=1}^{N} WTBF_i + \sum_{k=1}^{U} ST_k + \sum_{j=1}^{M} TTR_j} \tag{4}$$

Thereinto:
Ao—Operational Availability of a system.
WTBF—Working time for no failure.
N—The number of samples of working time for no failure.
ST_k—Standby time of a system.
U—The number of samples of standby time.
TTR—The time of repairing a failure of system.
M—The number of samples of repairing time.

4 Data Collection and Processing of Environment Agent

The objects of Data Gathering and Processing of Environment Agent mainly include analytical data and debugging data. Analytical metadata is as follows [1]:

- Work Time Between Failure: Work Time Between Failure, Abbreviated as WTBF, is the time from the last failure to the next failure. In the simulation course, the system state is in randomness. Because there are more discontinuous work time between the last failure and the next, the sampling for WTBF is based on principle of effective work. The metadata is in accordance with WTBF which is gathered from Adaptive Environment Agent, so the principle is the same, shown as Fig. 2. Being worth explaining, the concept of the two metadata is the same, but the connotation is different, the main difference is different gathering gradation.
- Mission Time, *MT*: It is the experience of time from the start of a mission to exit or finish. In the simulation course, because of different mission mode (in or

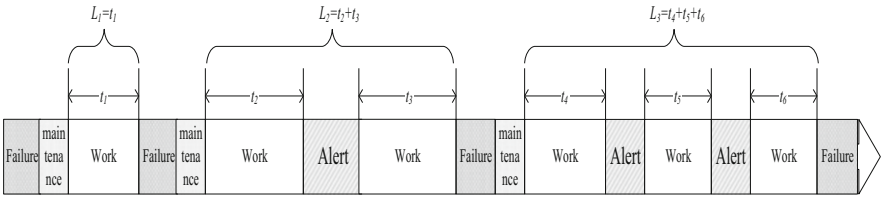


Fig. 2 Examples of part's WTBF

without consideration of maintenance) and random failure, it is random for the time of the same mission. There are the samples of effective mission time for *MT*1, *MT*2 and *MT*3.

- Mission Capable Time, as *MCT*: It is the time to execute mission except non mission Capable time, for example failure and maintenance in the course of mission execution. *MCT* includes Mission Capable Time and Alert time.

(1) Non Mission Capable Time, as *MNT*: It is the time not to execute mission because of failure and maintenance.

The relation of *MT*, *MCT* and *MNT* is as follows:

$$MT = MCT + MNT \tag{5}$$

(2) The times of mission success: There are the times of mission capable successfully. If a mission is successful, the submission is successful.

Environment Agent processes statistically the collected Metadata. The main contains:

(1) Mean Time Between Failure: Mean Time Between Failure, Abbreviated as *MWTBF*, is the mean work time between failure [2], as follows:

$$MWTBF = \frac{\sum_{i=1}^{N} L_i}{N} \tag{6}$$

Thereinto:

MWTBF is Mean Time Between Failure.
L_i is an effective sample of work time Between Failure.
N is the sample for the mean work time Between Failure.

(2) Mean Mission Capable Time is the mean time to execute a mission as is the mean of mission time, as follows:

$$MMT = \frac{\sum_{i=1}^{N} MT_i}{N} \tag{7}$$

Thereinto:

MMT is Mean Mission Capable Time.
MT is mission time.
N is the times of mission capable.

(3) Dependability: It is the rate of mission success [3], as follows:

$$R_{MS} = \frac{M}{N} \times 100\% \tag{8}$$

Thereinto:

R_{ms} is dependability.
M are the times of mission success.
N is the total times of mission capable.

5 Conclusion

It devises three types of Agent, as Component Agent, System Agent and Environment Agent to express reliability relations dynamically of materiel system in the mission with the change of mission profiles. Component Agent shows the basic working unit in the reliability model and simulates the reliability situation of a component. System Agent expresses a system object. Environment Agent simulates the change of mission and working condition. Because of limited length, it does not specify the simulation courses. But in the study, with an example of a communication product in the reference, it is effective for the simulation analysis results in comparison with the means and results of the reference. So in the paper, it analyzes data collection and process, defines the key elements for the analyses of mission reliability based on Agent simulation, and lays the good foundation for setting up a fine mission reliability simulation model of materiel system, analyzing dynamically mission reliability and making a perfect mission reliability simulation concept of materiel system.

References

1. Shen Y (2012) Research on reliability analysis methods for materiel system based on adaptive agent simulation. PhD Thesis, Academy of Armored Force Engineering
2. Chang XZ (2002) Supportability engineering, vol 59. Weapon Industry Press, Beijing, pp 65–69
3. GJB 451A-2005 (2005) Reliability, maintainability and supportability terms

A Relationship Construction Method between Lifecycle Cost and Indexes of RMSST Based on BP Neural Network

Ying Shen, ChenMing He, JunHai Cao, and Bo Zhang

Abstract In the paper, it divides LCC that affects the materiel effectiveness of the performance objectives of RMSST into the cost of development and the cost of support of operation and maintenance, and evaluates the cost of development, or the cost of support of operation and maintenance on BP Neural Network. There is the main limitation and insufficient of selecting BP Neural Network in addition to the method itself. But it provides important supporting to study the relationship between the performance objectives of RMSST and life cycle cost based on the theory of BP Neural Network.

Keywords LCC • RMSST • BP Neural Network

1 Introduction

BP Neural Network is the core of feed forward neural, with broad adaptability and effectiveness [1, 2]. As we know, Life cycle cost is abbreviated to LCC, and RMSST is in an abbreviated form of reliability, maintainability, supportability, security and testability. For cost evaluation on stages of life cycle, it is hard to express the relationship between cost and the indexes of RMSST in the general linear relationship and model [3]. And if in a model, it does not set up the model to the phase, in low precision and too much deviation. BP Neural Network sets up the relationship between input and output on the study of the samples, in case of ambiguity model, inadequacy information and unclear system, which takes advantages on the modelling, prediction, decision. So in the paper, it divides LCC that affects the materiel effectiveness of the performance objectives of RMSST into the cost of development and the cost of support of operation and maintenance, and evaluates the cost of development, or the cost of support of operation and maintenance on BP Neural Network shown in Fig. 1.

Y. Shen (✉) • C. He • J. Cao • B. Zhang
Academy of Armored Force Engineering, Beijing 100072, China
e-mail: sheny312@sina.com.cn

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_20

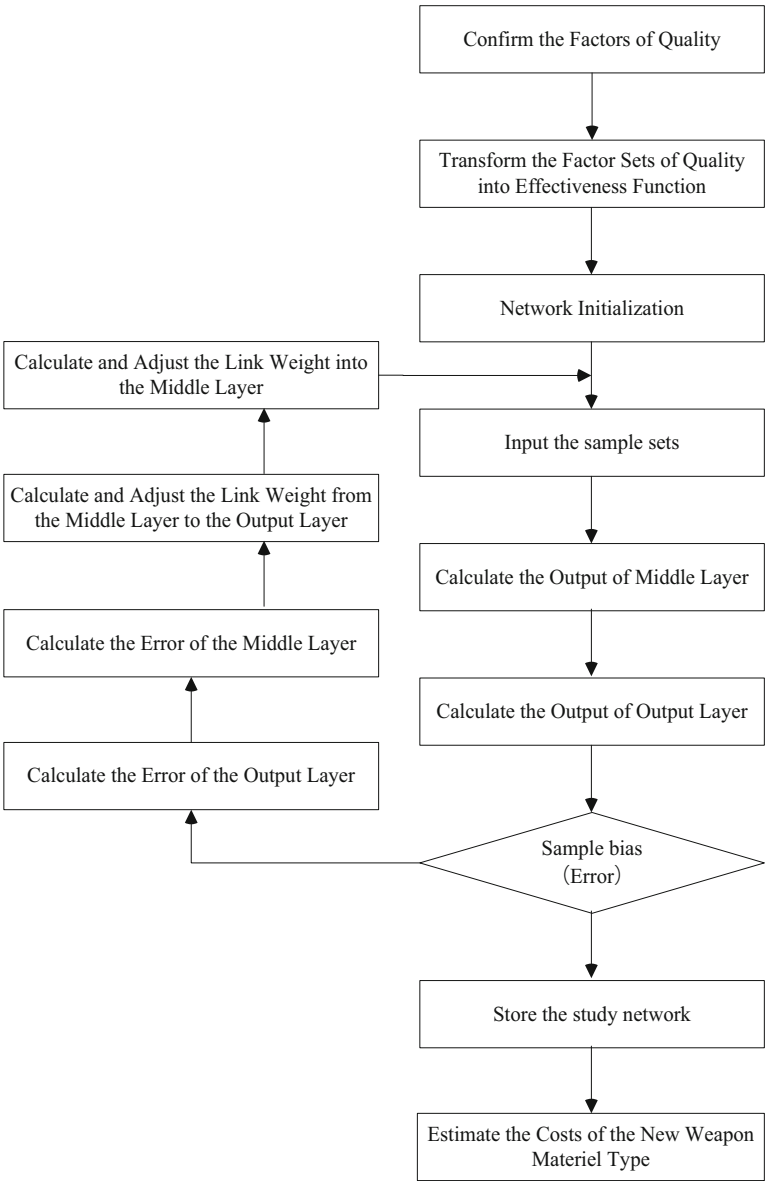


Fig. 1 The estimation procedures of BP Neural Network

2 The Structure Parameters of BP Neural Network

For the problem of the modelling of the LCC and the indexes of RMSST, it is the nonlinear mapping between the critical indexes of RMSST that affects the effectiveness of materiel and the cost of development, or operation and maintenance support. Because the three-layer of BP Network is arbitrary precision, it sets up the input layer, hidden layer and output layer. When there are many nodes in a hidden layer, it can add a hidden layer.

2.1 Input-Layer

1. Input Variable

   In general, there are the following basic rules when selecting input variables:

   - It can make a big difference for output, be detected or extracted.
   - The variables are mutually exclusive or in quite small correlation.

   So it takes the critical indexes of RMSST that affects the effectiveness of materiel, such as the cost of development or the one of operation and maintenance support, as the input variables, with mutual exclusiveness, no iteration and on overlap.

2. Transfer Function

   BP Neural Network is with the implicit layer of sigmoid and linear Output Layer, with very good mapping capability. If the output layer is tan–sigmoid, the range of output is from -1 to $+1$. While if the transfer function of output layer is purelin, the output is any value. So the implicit layer is the function of sigmoid, the output layer is linear function and broadens value field.

   So for the transfer function, the input layer is the function of tansig(), the middle layer is the function of logsig(), and the output layer is the linear function of purelin() or the function of logsig().

3. Pre-processing of the input

   It presents mainly scale variation, also named as normalization or standardization, which limit the input in the interval of [0,1] or [−1,1] by the Change of Address.

   When the dimension of every component is different, it is for the input transferred separately in the value range. When in the same physical significance and the dimension, it defines the maximum as xmax and the minimum as xmin in the total data domain, and transfers unifiedly. For example, if the input is transferred into the range of [0,1], by the follows:

$$\bar{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

Thereinto:
x_i—Input Data;
x_{min}—the minimum of data change.
x_{max}—the maximum of data change.
x_{mid}—the median of data change range.

If processed with nondimensionalization for primary data, it is by the follows:

$$x_{ij}^{*} = \frac{x_{ij}}{s_{ij}} \tag{2}$$

$$s_{ij} = \sqrt{\frac{\sum \left(X_{ij} - \bar{X}_j\right)^2}{n-1}} \tag{3}$$

2.2 Output-Layer

By and large, the output represents the functional goals of the system, as the desired output, which is defined easier, for example the performance indexes, the categories of classification problems or the functional value of nonlinear function. The output of the network sketches out the cost of development, or operation and maintenance support. So the transfer function of the output layer is the function of purelin() or logsig().

Since the input is multidimensional, the model is multi-input and single-output.

2.3 Hidden-Layer

It makes a difference for the number of nerve cells in the hidden layer. For less, the net is in poor acquisition capability, which can not summarize and embody the sample rule for the training set. For more, it remembers non-regularity as noise, which not only lowers generalization but also increases the training time. In general, the common method of ensuring the best number of hidden-layer nodes is cut-and-trial. It first sets fewer hidden-layer nodes, and then increases. It trains in the same sample sets, and determines the number of hidden-layer nodes at the time of minimum error of the net. For the initial value, it determines by the empirical formula:

$$m = \sqrt{nl} \tag{4}$$

Thereinto:

m—the number of hidden-layer nodes;
n—the number of input-layer nodes;
l—the number of output-layer nodes.

Since the net is nonlinear, it is for initial weight if the study is in the minimum in the local, in the convergence and in much influence with the training speed. It takes the initial weight as the random number in the range from −1 to 1.

2.4 Training and Testing of the net

For the middle and small size of the BP net, it takes best the algorithm as Levenberg–Marquardt, which is in the best rate of convergence and lower size of memory. But if it runs short of storage space, it selects other types of rapid algorithm. For large-scale net, it selects best conjugate gradient algorithm of Scaled or elastic gradient algorithm.

So for the point of the BP net, it trains in the algorithm of Levenberg–Marquardt.

3 Example Analysis

The cost of operation and maintenance is related to the service life of the materiel. The service life is longer, and the cost of maintenance support is more. But it is uncertain that the analysis of the cost of operation and maintenance. As the above principle, it sets a model of BP neural network with five indexes of RMSST as the independent variables and the cost of operation and maintenance as the dependent variable. The original data as inputs is shown in Table 1 and corresponding pre-processing data is given in Table 2.

Table 1 The table of original data (The cost of operation and maintenance Versus the related performance of RMSST for some type of tank)

Material	X_1	X_2	X_3	X_4	X_5	The cost of operation and maintenance support in some year (ten thousand yuan)
A	0.60	8.00	60.00	4.00	0.900	2129.00
B	0.50	6.00	40.00	4.50	0.700	1000.00
C	0.45	5.00	30.00	4.80	0.600	800.00
D	0.40	4.50	26.00	4.90	0.500	700.00
E	0.45	4.30	25.00	5.00	0.450	520.00
F	0.30	4.00	20.00	5.00	0.400	300.00

Table 2 The table of pre-processing data

Material	X_1	X_2	X_3	X_4	X_5	C
A	1	1	1	0	1	1
B	0.6667	0.5	0.5	0.5	0.6	0.3827
C	0.5	0.25	0.25	0.8	0.4	0.2734
D	0.3333	0.125	0.15	0.9	0.2	0.2187
E	0.5	0.075	0.125	1	0.1	0.1203
F	0	0	0	1	0	0

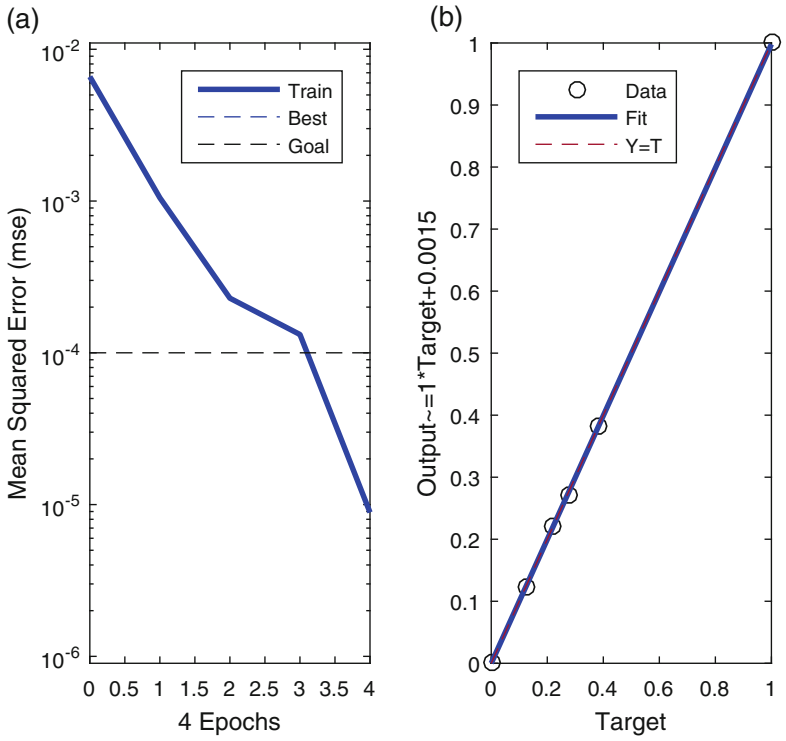


Fig. 2 The training and test results: (**a**) Training error as a function of training epoch and (**b**) Training output versus the target value

The test results are provided in Fig. 2. From Fig. 2, we can know that the relationship between the lifecycle cost and indexes of RMSST is generally linear, and the computational accuracy of the proposed method is acceptable.

4 Summary

There is the some presentation of the study in the task by the above analysis. It sets up the relationship model between the performance objectives of RMSST and development cost. It shows the veracity of BP Neural Network compared with regression analysis. There is the main limitation and insufficient of selecting BP Neural Network in addition to the method itself.

- It introduces the initial BP Neural Network model based on the performance objectives of RMSST to evaluate the cost of development. To the concrete issue, it fits the model according to the relative data, when it adjusts probably the relative function and method, to improve accuracy.
- The generalization of BP Neural Network is associated with prediction of the testing data. It need set and train the net based on the data. In view of the complexity and specialty on materiel construction, and with the new peculiarity of the newly developed materiel, it is different between current data and the relationship of some performance objectives of RMSST and development cost. So it changes for the performance objectives of RMSST that affect materiel effectiveness.

In view of the complexity and specialty of materiel construction and with the new characteristics of new developed materiel, it is different between the relationship of the performance objectives of RMSST and development cost and current data, so the objectives of RMSST affecting materiel effectiveness change, as the input-layer of network probably changes. If the trained network predicts, it is biased by the benchmark. It need more study and extendedly analyze on the basis of modelling theory if with quite strong generalization and veracity. But it provides important supporting to study the relationship between the performance objectives of RMSST affecting materiel effectiveness and life cycle cost based on the theory of BP Neural Network, and settles effectively the problems of setting up the cost appraisal models by stages in the whole life cycle and improving the accuracy of the models.

References

1. Shi Y, Han LQ, Lian XQ (2009) Design methods and example analysis of netual network. The Press of Beijing University of Posts and Telecommunications, Beijing
2. Zhu K, Wang ZL (2010) Mastering MATLAB netual network. Publishing House of Electronics Industry, Beijing
3. Shenying HC et al (2015) Relative RMSST trade-off analysis study

Asset Performance Measurement: A Focus on Mining Concentrators

Antoine Snyman and Joe Amadi-Echendu

Abstract The output from platinum mining is typically pre-processed in concentration plants before further refinement downstream. Asset performance measurement is a concept which has not enjoyed as much attention in the mining industry as in, for example, the automobile industry and this gap leaves room for improvement in the mining industry. This article discusses the measurement of the performance of equipment and processes that constitute a mining concentrator asset by using overall equipment effectiveness. A new conceptual model is suggested and applied to measure the performance and quality of a continuous processing system of seven concentrators of a case study mining company.

Keywords Asset performance measurement • Asset effectiveness • Mining concentrator performance

1 Introduction

In the last 40 years, the demand for platinum group metals (PGMs) has grown rapidly [1] but in recent times, the price of the commodity has manifested high volatility, somewhat in response to stricter standards for diesel-powered vehicles, the general slowdown in the global economy, and the consequential sales of stockpiles by car manufacturers [2]. With decreasing reserves of rich ore bodies coupled with high volatility in price of the commodity, platinum sector firms have to find ways of optimising the performance of engineered assets throughout the mining value chain. Most firms are challenged to produce more with existing natural resources and engineered assets [3], and

A. Snyman (✉)
Lonmin Plc, Marikana, South Africa
e-mail: antoine.snyman@gmail.com

J. Amadi-Echendu
Department of Engineering and Technology Management, University of Pretoria, Pretoria, South Africa
e-mail: joe.amadi-echendu@up.ac.za

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_21

naturally, these organisations tend to respond by implementing cost reduction initiatives.

In an attempt to define the objectives of managing engineered assets, Too [4] undertook a study and identified four main goals, namely cost efficiency, capacity matching, meeting customer needs and market leadership. The general view today furthers this and advocates the use of asset management strategies in order to unlock business value [5]. Amadi-Echendu [6] also found that effective asset management is ultimately accountable for the triple bottom line of a business in terms of economic, environmental and social performance measures.

With focus on the economic aspect of the triple bottom line, Henderson et al. [7] argue that the reactive approach to maintenance is extremely expensive both in terms repair cost as well as the cost of lost production. They further argue that the indirect costs of maintenance can be up to seven times the direct costs and these hidden costs often have the most severe financial impact. For example, Henderson et al. indicate that a 1% increase in the overall equipment effectiveness (OEE) will add 4–8% to the overall profit of an organisation and vice versa. Henderson et al. [7] conclude that the combination of leading edge proactive maintenance strategies amongst others can lead to increased reliability and sustainability.

One strategy which is predominantly present in literature is total productive maintenance (TPM). It is defined by Ahuja and Khamba [8] as a production driven improvement methodology that is designed to optimise equipment reliability towards efficient management of plant assets. Furthermore, TPM as a practice is aimed at improving the maintenance function and productivity of an organisation from the viewpoint of quality management. TPM involves cooperation between the departments in an organisation to address product quality, operation efficiency and production capacity [9]. In short, it is the aim of TPM to increase production reliability by focussing on equipment effectiveness and maintenance.

The effectiveness of equipment may be improved by applying the concept of overall equipment effectiveness (OEE). Kumar et al. [10] defined OEE as a simple and practical way to measure the effectiveness of an item of equipment on the basis of its availability, efficiency and throughput. This article discusses the application of the OEE model to examine the effectiveness of items of equipment installed in a mining concentrator plant. A review on OEE is briefly presented in Sect. 2, with a particular focus. A case study application of OEE on a platinum mining concentrator plant is described in Sects. 3 and 4. Following from the case study, Sect. 5 reiterates the value doctrine for managing engineered assets.

2 Measuring Equipment Effectiveness

Overall equipment effectiveness (OEE) is a metric typically used to measure the effectiveness of equipment installed in a manufacturing plant and it is expressed as:

$$OEE = Availability \times Performance \times Quality \tag{1}$$

The first component of OEE, availability, is represented as illustrated in Fig. 1 and Eq. (2).The total time in Fig. 1 is 24 h.

$$Availability = \frac{Actual\ Running\ Time}{Time\ Available\ for\ Running} \times 100 \tag{2}$$

Performance[1] is normally calculated by relating the actual performance to nameplate performance. This metric cannot be applied in a concentrator due to the varying nameplate performance of different items of equipment in the process chain. Hence, a new metric for performance is expressed in Eq. (3).

$$Performance = \frac{Actual\ Hourly\ Mill\ Feed\ Rate}{Average\ of\ Top\ 6\ Hourly\ Mill\ Feed\ Rates} \times 100 \tag{3}$$

Quality is the third and perhaps the most difficult component to measure since concentrators have no distinct product which can be inspected in real time. A plausible basis for determining the quality of the output is to use the *recovery* after the run-of-mine has passed through the concentrator. However, the assay results for *recovery* are only available three to five days after production. This

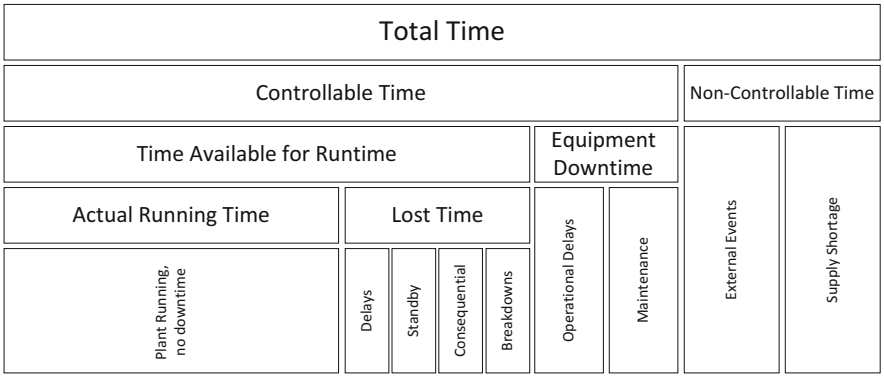


Fig. 1 OEE time allocation model

[1]Performance is also referred to as "Efficiency" in literature. Performance of the asset should not be confused with the nameplate performance of a specific piece of equipment.

delay means that OEE may only be determined retrospectively. An alternative approach is to determine the quality parameter in OEE from the metallurgical expression in Eq. (4) since these metallurgical components is managed on a daily basis to ensure a good quality product.

$$Quality = CrusherGrind \times PrimaryMillGrind \times FloatSG \times MassPull \times ReagentDosages \quad (4)$$

The *reagent dosages* is a combination of reagents which should all be in specific range and it is of extreme importance that this metric should be tailored for every plant. The *reagent dosages* calculation is seen in Eq. (5).

$$Reagent\ Dosage = Xanthate \times CopperSulphate \times Depressant \quad (5)$$

To establish the ranges for the quality metrics, the average of the top 5% of recovery figures was taken and this figure represents 100% quality.

3 Case Study

Technology has moved long since the days of manual data capturing and it should be avoided for data of this importance [11–15]. The first data set extracted was an indication of whether or not the concentrator was running and, as suggested by Andersson and Bellgran [16], a simplistic solution was opted for. In light of this suggestion, the mill at a concentrator was the best asset to monitor runtime, but the electrical load on a mill was also taken into account since mill running status can be misleading. It is safe to assume that a mill is running contributively when it is consuming a significant amount of electricity, typically 1.5 MW. The concentrator was subsequently deemed to be running when the mill was both on and it consumed more than a specified amount of electricity. This measure can easily be configured on a plant historian and integrated to a system.

Components of the OEE time allocation picture in Fig. 1 are based on operator selected reason recorded in the plant historian, e.g. when the operator selected a reason which is categorised as an “affects lost time” reason, the downtime is interpreted as lost time. All non-productive times are documented along with the associated reasons for delay.

The actual mill feed rate used in the performance calculation (Eq. 3) was sourced from the supervisory control and data acquisition (SCADA) system and averaged over the timeframe for which the metric was calculated. The average of top 6 hourly mill feed rates looked at history trends for the specific concentrator as well as the specific ore type milled. The top 6 feed rates provided a measure of the nameplate performance of the concentrator.

OEE is a difficult concept to apply in a mining concentrator environment due to the absence of a quality parameter on a real time basis. Metallurgists normally

Table 1 Deviance penalties

Deviance from min or max (%)	Score (%)	Deviance from min or max (%)	Score (%)
0	100	51–60	40
1–10	90	61–70	30
11–20	80	71–80	20
21–30	70	81–90	10
31–40	60	90+	0
41–50	50		

manage a concentrator by looking at various metrics such as those mentioned above. If all of these values are in range, metallurgists are somewhat convinced that the quality of the product will be good. The heterogeneity of individual concentrators and ore being treated dictates that there cannot be a single set of metrics which can be used as a standard and each concentrator should set up its own metrics. Furthermore, due to the nature of variability in a single concentrator, these metrics cannot be a distinct value but should rather be a range of values. A metric may be scored 100% when it was in the specified range. Table 1 lists the penalties associated with the deviations and can be read as follows: If the mill grind was 8% below the minimum set range, then the quality of mill grind scores 90%.

To source the data for the quality metrics, different processes were followed based on the dictation of the metric. The crusher grid and mill grind was obtained from a laboratory system. Float specific gravity were taken manually by the operators and fed into an online log sheet. Mass pull was a metric calculated by the SCADA systems and finally, the reagent dosages were monitored by operators. A sample of the flow and concentration of the reagents were taken and reported and the data for these were either found on a log sheet or in a database, depending on the monitoring method.

4 Analysis of Case Study Data

Data was collected by means of a mixture of methods as discussed earlier and proposed by Ljungberg [14]. Data was collected over a time period of two years and it involved seven concentrators, but data for the first half of 2014 were excluded from the results due to five month strike in the platinum sector in 2014. It is important to note that the figures used in this report are shrouded to maintain confidentiality.

4.1 Availability

No two concentrators are alike and this would have an impact on the results of metrics being measured. Figure 2 shows the boxplots of the availability metric for

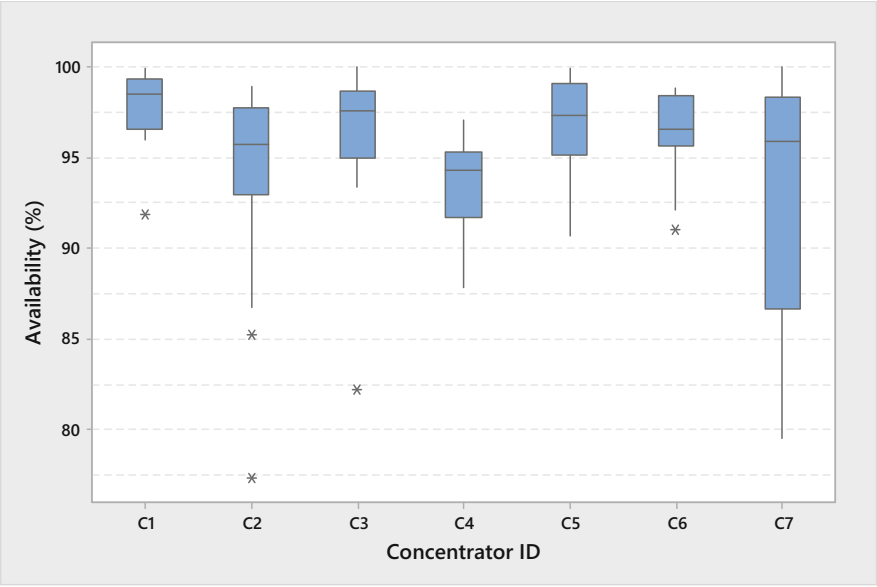


Fig. 2 Boxplot of the availability metric

the different concentrators and confirms the heterogeneity of concentrators. C7 shows a wide spread in terms of availability and may indicate that the concentrator is not under control with erratic breakdown patterns; a suspicion which was confirmed upon investigation.

4.2 Performance (Efficiency)

Similar to availability, the performance metric shows variances across different concentrators with strong indications that concentrators should run at a performance level of around 80–90%. The performance metric of a concentrator essentially compares the concentrator against itself and due to this fact, it is possible to compare the concentrators' performance figures to each other (Fig. 3).

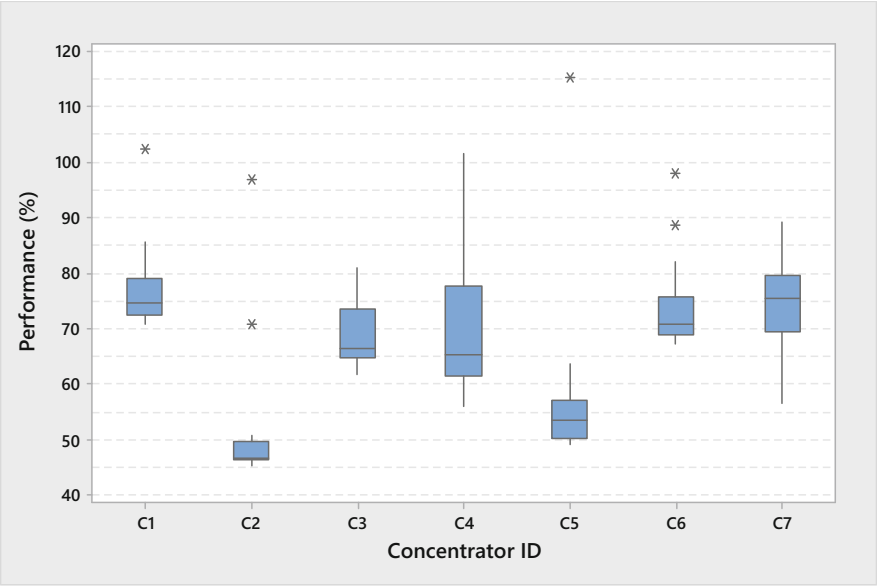


Fig. 3 Boxplot of the performance metric

It is noted that some of the concentrators have a fourth quarter performing above 100%. Although this is not impossible, it is not recommended as equipment should not be pushed beyond the limits.

4.3 Quality (Throughput or Yield)

Firstly, the acceptable range for each of the metrics contributing to the quality metric needed to be established. The Recovery figure for each concentrator was recorded for the last two and a half years and the quality metrics' data for corresponding timespan was matched to it. After the erroneous data points were brushed out, 500 data points per concentrator remained. The 25 values with the Recovery figure in the top 5% of the so-called clean data was separated and used to establish the ranges for the quality metrics.

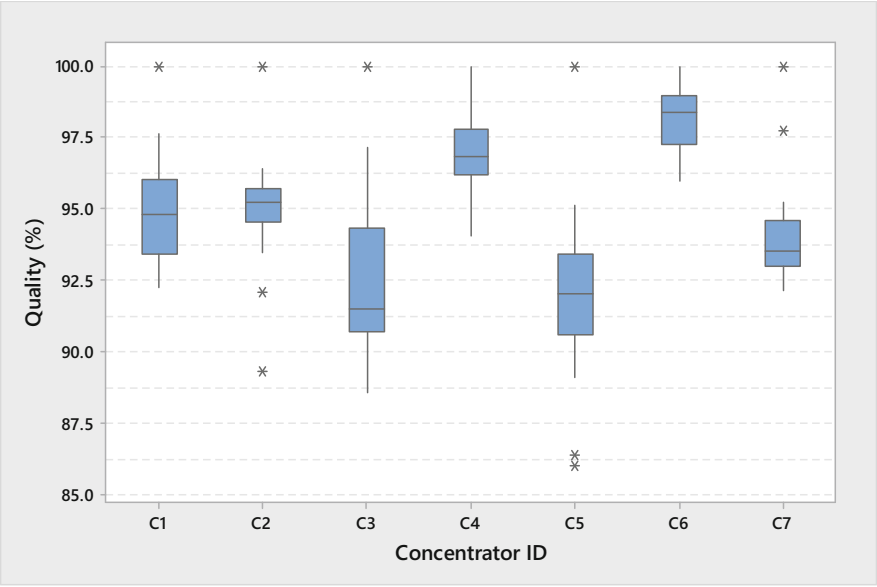


Fig. 4 Boxplot of the quality metric

When the abovementioned metrics are in range, it can be assumed that the recovery will be high and subsequently, the quality of the product produced will be 100%. Upon calculation, the boxplot in Fig. 4 confirms the validity of the quality metric as well as the heterogeneity in concentrators.

4.4 OEE

The OEE measure calculated shows the unique differences between each concentrator. It is noticed that most concentrators operate in a narrow band of OEE when compared to itself, but the OEE range has wide averages lying between 45% and 80% when looking at different concentrators. Figure 5 clearly shows a spike in August 2014s data which is attributed to the performance of each concentrator post the strike (Fig. 6).

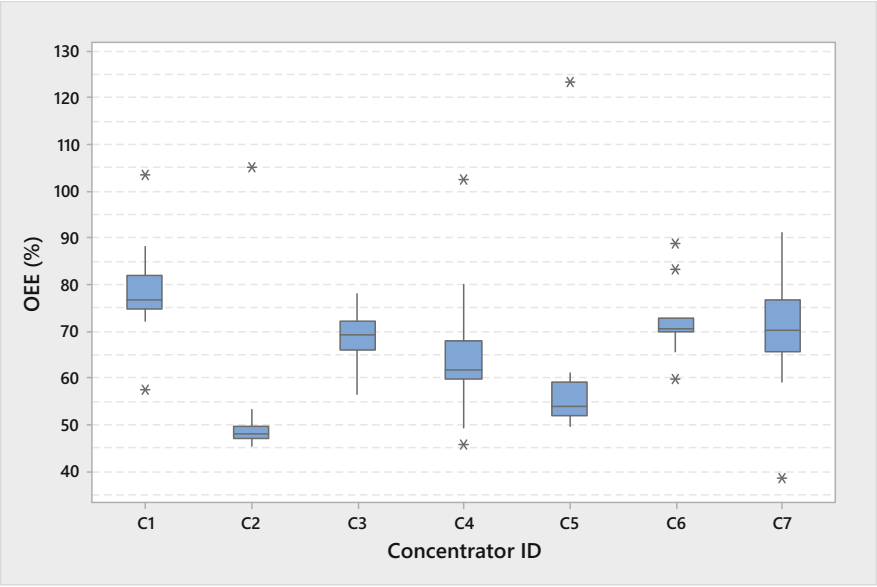


Fig. 5 Boxplot of OEE metric

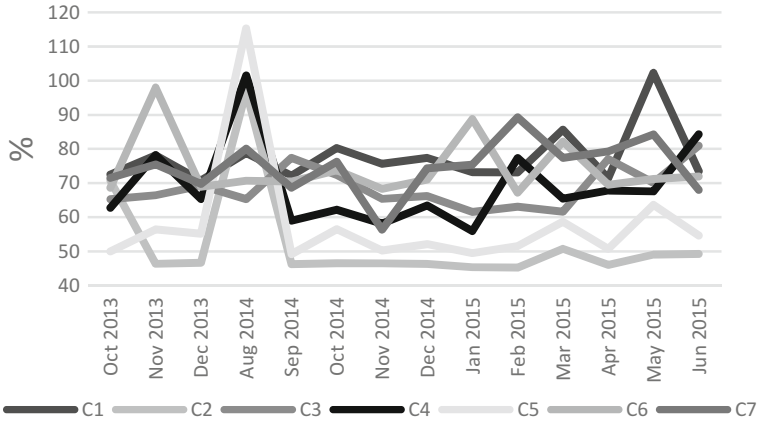


Fig. 6 OEE metric results

5 Summary

The articles Amadi-Echendu [6] and Amadi-Echendu et al. [5] make that point that performance of an engineered asset should be approached from a value doctrine. This article has briefly described a case study on the challenges encountered in measuring the parameters for measuring asset performance and calculating OEE in

practice. The challenges surround the proper definition of the parameters and collection of pertinent data. Muchiri and Pintelon [15], Ljungberg [14], Dal et al. [11], Kullh and Josefine [13], and Jonsson and Lesshammar [12] all reiterate that data should be collected electronically and analysed. Our case study reveals that, in practice, the cost doctrine still remains the primary driver for the application of the OEE concept. Hence, the measurement of asset utilization is not a priority. Since an asset is a thing of value, the effectiveness of an asset must depend on how it is used, therefore, we argue that the measurement and assessment of the performance of an engineered asset like a platinum mining concentrator should be approached from a value doctrine so that the utilization is appropriately defined and measured.

References

1. Wilburn DR, Bleiwas DI (2004) Platinum-group metals—world supply and demand. US geological survey open-file report 2004-1224
2. Survey G (2015) Mineral commodity summaries, 2015. U.S. Government Printing Office
3. Mudd GM (2012) Key trends in the resource sustainability of platinum group elements. Ore Geol Rev 46:106–117
4. Too EGS (2009) Capabilities for strategic infrastructure asset management. PhD Thesis
5. Amadi-Echendu JE, Willett R, Brown K, Hope T, Lee J, Mathew J, Vyas N, Yang B-S (2010) What is engineering asset management? Definitions, concepts and scope of engineering asset management. Springer, London, pp 3–16
6. Amadi-Echendu J (2004) Managing physical assets is a paradigm shift from maintenance. In: Proceedings IEEE international engineering management conference 2004, pp 1156–1160
7. Henderson K, Pahlenkemper G, Kraska O (2014) Integrated asset management—an investment in sustainability. Proc Eng 83:448–454
8. Ahuja IPS, Khamba JS (2008) Assessment of contributions of successful TPM initiatives towards competitive manufacturing. J Qual Maint Eng 14(4):356–374
9. Madu CN (1994) On the total productivity management of a maintenance float system through AHP applications. Int J Prod Econ 34(2):201–207
10. Kumar SV, Mani VGS, Devraj N (2014) Production planning and process improvement in an impeller manufacturing using scheduling and OEE techniques. Proc Mater Sci 5:1710–1715
11. Dal B, Tugwell P, Greatbanks R (2000) Overall equipment effectiveness as a measure of operational improvement—A practical analysis. Int J Oper Prod Manag 20(12):1488–1502
12. Jonsson P, Lesshammar M (1999) Evaluation and improvement of manufacturing performance measurement systems-the role of OEE. Int J Oper Prod Manag 19(1):55–78
13. Kullh A, Josefine A (2013) Efficiency and productivity improvements at a platinum concentrator. Masters of Science in Mechanical Engineering, Chalmers University of Technology
14. Ljungberg Õ (1998) Measurement of overall equipment effectiveness as a basis for TPM activities. Int J Oper Prod Manag 18(5):495–507
15. Muchiri P, Pintelon L (2008) Performance measurement using overall equipment effectiveness (OEE): literature review and practical application discussion. Int J Prod Res 46(13):3517–3535
16. Andersson C, Bellgran M (2015) On the complexity of using performance measures: enhancing sustained production improvement capability by combining OEE and productivity. J Manuf Syst 35:144–154

Mobile Technologies in Asset Maintenance

Faisal Syafar, Andy Koronios, and Jing Gao

Abstract Assets are the lifeblood of most organizations. Maintenance is critical in asset management. Whenever a machine stops due to a breakdown, or for essential routine maintenance, it incurs a cost. Unlike consumer applications, in heavy industry and maintenance, the uses of mobile solutions have not yet become very popular. However, it is believed that mobile solutions can bring maintenance management closer to daily practice in the field, and lead to more efficient maintenance operations. This research has adopted a multi-case studies in order to identify the role of mobile technologies in as-set maintenance activities. The findings will contribute to the development of mobile technologies in facilitating effective and efficient maintenance in engineering asset management organisations.

Keywords Mobile technology • Collaborative maintenance

1 Introduction

Assets are the lifeblood of most organizations. They may include digital assets, human assets, and financial assets. Most companies also have physical assets. These physical engineering assets (e.g. machinery, plant and equipment, etc.) can be used to turn raw material into finished goods, supply electricity and energy, provide transportation services, or control huge utility operations. Many organizations rely heavily on these engineering assets to maintain and monitor daily operations. During the lifecycle of these engineering assets, an enormous amount of data is produced. The data is captured, processed and used in many computer information systems such as Supervisory Control and Data Acquisition (SCADA) systems, Facility Maintenance and Management Systems (FMMS), and Geographic Information Systems (GIS).

F. Syafar (✉)
Universitas Negeri Makassar, Makassar, Indonesia
e-mail: faisal@mymail.unisa.edu.au

A. Koronios (✉) • J. Gao
University of South Australia, Adelaide, Australia
e-mail: Andy.Koronios@unisa.edu.au; jing.gao@unisa.edu.au

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_22

2 Importance of Asset Maintenance

Maintenance is critical in asset management. Whenever a machine stops due to a breakdown, or for essential routine maintenance, it incurs a cost. The cost may simply be the cost of labour and any materials, or it may be much higher if the stoppage disrupts production [1]. In order to define how far such interruptions (due to wear, tear, fatigue and sometimes corrosion) has impacted plant and/or machinery of engineering assets, systematic inspection is required. Routine or systematic maintenance plays an important role as a requirement to achieve certain production targets.

As explained by Dekker [2] the maintenance role can be defined by the four objectives it seeks to accomplish. They are:

- Ensuring system function (availability, efficiency and product quality). For production equipment this is the main objective of the maintenance function. Here, maintenance has to provide the right reliability, availability, efficiency and capability to produce at the right quality for the production system, in accordance with the need for these characteristics.
- Ensuring the system's or plant's life refers to keeping systems in proper working condition, reducing the chance of condition deterioration, and thereby increasing the system's life.
- Ensuring human wellbeing or equipment shine. This objective has no direct economic or technical necessity, but is primarily a psychological one of ensuring the equipment or asset looks good.
- Ensuring safety refers to the safety of production equipment and all engineering assets in general.

Gouws and Trevelyan [3], and Soderholm et al. [4] state that maintenance stakeholders are the individuals in the organisational structure involved directly or indirectly with maintenance. Some people are very visible in the maintenance workflow process (such as managers, maintenance engineers, maintenance supervisors, and maintenance technicians) while others are less visible, but not less important (e.g. reliability engineer inspectors, accountants, purchase buyers, and computerised maintenance management systems [CMMS] administrators).

Maintenance is a combination of actions intended to retain an item in, or restore it to, a state in which it can perform the function that is required for the item to provide a given service. This concept leads to a first classification of the maintenance actions in two main types: actions oriented towards retaining certain operating conditions of an item, and actions dedicated to restoring the item to supposed conditions. Retention and restoration are action types that are then converted into preventive and corrective maintenance types in the maintenance terminology by the European Committee for Standardization (CEN [5]).

The following sections present the European Committee for Standardization [5] explanations of corrective, preventive and condition-based maintenance.

Corrective maintenance (CM), also called breakdown maintenance or run-to-failure [6], is maintenance carried out after fault recognition, and is intended to put the equipment into a state in which it can perform a required function.

Preventive maintenance (PM), also called planned maintenance or time-based maintenance [6], is defined as maintenance carried out at predetermined intervals or according to prescribed criteria and intended to reduce the probability of failure or the degradation of the functioning of the equipment. It involves preventive actions such as inspection, repair or replacement of the equipment. It is performed in fixed schedules and regardless of the status of a physical asset.

Condition based maintenance (CBM), From Jardine, Lin and Banjevic's point of view 'Condition based maintenance (CBM) is a maintenance program that recommends maintenance actions based on the information collected through condition monitoring techniques' ([7], p. 77). CBM is PM based on performance and/or parameter monitoring and subsequent actions. Performance and parameter monitoring may be scheduled, on-request or continuous. CBM includes predictive maintenance that can be defined as CBM carried out following a forecast derived from the analysis and evaluation of the significant parameters of the degradation of the equipment.

As mentioned above, there are three types of assets maintenance including CM, PM and CBM. CM is a kind of maintenance method based on a failure shutdown, and its basic idea is not to repair until breakdown. PM is a proactive maintenance method, including predetermined PM. CBM is an effective PM that carries out equipment maintenance work based on the real-time status of and use plan of the assets.

3 Mobile Technologies in Asset Maintenance

Sokianos et al. [8] emphasise that in order to manage the sophisticated AM process and to provide its data requirements, particular technology and systems are required. The system that captures, maintains, and manages all the needed asset information throughout the entire asset lifecycle is critical in providing effective AM.

In contrast, mobile technologies and solutions are very popular in consumer applications, and the exploitation of mobile technologies will keep on expanding. In heavy industry and maintenance, the uses of mobile solutions have not yet become very popular. One reason is the lack of competence and knowledge in adopting mobile solutions successfully for professional use. Many companies have poor experience of adopting mobile solutions in maintenance due to previously inoperative telecommunication connections, lack of suitable devices, or just because the organisation had insufficient preparation for the adoption and implementation process. Another reason is that the benefits of mobile solutions are unseen or unknown, for example, in the maintenance domain [9]. Mobile technologies

nowadays are mature enough to face the challenge and requirements of professional use in the engineering industry.

The use and implementation of mobile services has been studied globally and extensively from a context-driven organisational problem-solving view [10]. When considering the use of mobile solutions in industry, and especially in maintenance, the available studies and research focuses mainly on e-maintenance [11–13]. The term e-maintenance still refers to quite a large concept where mobile solutions can be just one part. Some e-maintenance specific case studies focus on mobile device architectures where the mobile device can, for example, help the maintenance engineer perform maintenance tasks [11]. Mobile solutions can bring maintenance management closer to daily practice in the field, and lead to more efficient maintenance operations.

Some research has been conducted on the role of mobile technology in the workplace, but only few applied to asset maintenance works. Moreover, several mobile maintenance systems have been invested in by EAM organisations to enhancing their AM and maintenance systems. But these technologies/systems do not adequately support the maintenance collaboration requirements associated with different maintenance stakeholders.

4 Research Question and Design

A multiple case-study approach was adopted for the case-study methodology in this research. It is aim to identify the role of mobile technologies in asset maintenance activities with specifics focuses on

a. The current use of mobile technologies in asset maintenance
b. Collaborative asset maintenance requirements
c. Issues and problems associated with the current mobile technologies

The reasons for choosing a multiple case study approach over a single case approach was its capacity to handle the complexity of the phenomenon under study [14], and the fact that it augmented external validity, helping guard against observer bias [15]. It is recommended to be of assistance in capturing the complexity of social settings, and facilitating the comparison of activities across a variety of settings and situations [16]. The multiple case-study approach uses replication logic to achieve methodological rigour [14, 17], and triangulate evidence, data sources and research methods [18].

Eight Australian and Indonesian engineering asset organisations were selected for the case study in this research. All were chosen from large sized organisations taking into consideration the complexity of maintenance process, such as having more functions, covering more operation and maintenance perspectives, involving more and variety of maintenance stakeholders, and more importantly, having strong motivation to improve their maintenance productivity. They truly reflect the engineering asset organisations that need, or have been implementing, collaborative

Table 1 Overview of case organisations

Case	Description	Organisation size	Business nature	Interview period
A	Government organisation	Large	Telecommunications	July 2013
B	Private enterprise	Large	General trades (multi areas)	July 2013
C	Government organisation	Large	Petroleum	August 2013
D	Private enterprise	Large	Telecommunications	August 2013
E	Government organisation	Large	Electricity	October 2013
F	Government organisation	Large	Airline services	October 2013
G	Private organisation	Large	Electricity	June 2013
H	Private organisation	Large	Oil and gas	September 2013

maintenance systems in supporting their routine maintenance activities. The eight case-study organisations also represent the typical engineering industries of telecommunications, electricity, airline services, and oil and gas, in both the public and private sectors. In order to respect the privacy of the participating organisations and individual interviewees, they are not identified by their real names or actual position titles. The cases are referred to as Case A through to Case H. Table 1 provides an overview of the eight organisations. It includes a description of each organisation, the organisation's size, the main business, and the period when interviews were conducted. The cases include four public (government) organisations, and four private organisations. The case-study interviews were carried out between July 2013 and September 2013.

This study employed a pragmatism stance in the eight case studies in order to determine and identify the collaborative maintenance requirements for successful implementation. Therefore, the qualitative data were collected and organised using two different methods. First, interview responses were transcribed and tested for accuracy through a couple of run-throughs by comparing the recording with the transcriptions. Second, thematic analysis was performed to identify patterns and themes within the data. Thematic analysis is a method that allows the researcher to report the experiences of the study's respondents captured during the interview process. The interpretation identifies new information and findings based on the interview questions that become progressively focused towards the research questions.

All case study interviews were transcribed. A very intensive content analysis of those documents and interview transcripts was conducted. All transcript material was coded [19] according to the research developed framework and the refined interview protocol questions. Coding of the data made it easier for the researcher to detect trends and commonalities among the interviewees.

Table 2 Collaborative maintenance requirements

No.	Requirements	Frequency
1.	Mobile technology competence	14
4.	Clear maintenance vision (maintenance strategy-business objective)	11
5.	Data and information accessibility	10
6.	Cross-organisational management communication	10
7.	Common understanding of maintenance processes	10
8.	Specific mode for each specific maintenance roles	9
9.	Mobility of the users, devices and services	9
10.	Trust and commitment the other crews will do their part	9

Table 3 Current mobile technologies being used to support asset maintenance

No.	Statements	Frequency
1.	Preventive maintenance expert availability	15
2.	Job information library	13
3.	Copy and printing facilities	12
4.	Display data/information in the form of text, audio, picture, visual and video format	11
5.	Hyperlinks	11
6.	Work list	11
7.	Expandable	11
8.	Document in the form of word, spreadsheet and pdf file	10
9.	Wireless (3G or LTE)	10
10.	Wi-Fi	10

5 Research Findings

Current maintenance circumstances are more complex because engineering assets having an increasing number of functions, requiring maintenance processes to be managed by multiple and interlinked activities. Hence, an integrated high-level maintenance system, which contains multiple sub-systems, requires inter-departmental collaboration of multiple stakeholders. Operation and maintenance is the longest and most complex lifecycle stage, thus needing additional attention. Due to complexity, long process, and multiple stakeholders and departments involved, coordinating and sharing AM data from all disparate sources into operational business intelligence requires many skills in intra-organisation and inter-partner collaboration [20]. Through the interview, it can be clearly identified that mobile technologies play an important role in facilitating collaboration activities in maintenance [21, 22]. Some findings are summarised below (Tables 2, 3, 4, and 5).

Table 4 Current problems with mobile enabled collaborative maintenance systems

No.	Problems	Frequency
1.	H/W and S/W limitations or lack of functions	10
2.	Lack of responsiveness of skilled maintenance people	9
3.	Unavailability of skilled maintenance people	9
4.	Establishing common ground is a crucial activity for collaboration	9
5.	Difficult to access the history of previous maintenance work	8
6.	System security become even more important and complex	8
7.	Lack of support from corporate offices	8
8.	Lack of commitment on maintenance resources	8
9.	Technology does not operate as expected in real world, energy is still an open problem for many contexts, e.g.: bridge maintenance shifts have to be adapted to battery availability/charge.	7
10.	Limited use in large industry in developing countries only	7

Table 5 Perceived mobile technology roles supporting collaboration technology systems

No.	Statements	Frequency
1.	Mobile technology allows at the right place to access directly to a set of information coming from all the potential actors involved in the decision (CMMS, ERP, sensors, etc.)	11
2.	Visualising of collected data, parameter history and trending	11
3.	Contextualising access over remote data and services: task-related services and data entry ubitously available to authorised users	11
4.	Critical for response time for data or information that can lead to early correction and or identification of failures	11
5.	Allowing to take the right maintenance decision, at the right time, at the right place, from the right information	10
6.	Comprehensive failure report	9
7.	Reports actual working hours and availability	8
8.	Enhancing accuracy of critical data entry for maintenance history	8
9.	Detecting the location of skilled maintenance personel nearby an asset that has experienced a failure through GPS	8
10.	Resources management (material, maintenance people) facilitator for continous task monitoring/assignment/reporting	8

6 Conclusion

Engineering asset organizations will be better able to identify problems associated with the current mobile technologies as well as critical requirements including the relationships among these key requirements for effective and efficient mobile maintenance operations. The research findings have suggested that by utilising mobility solutions, maintenance crews (as the users) can access vital information as and when they need to. The mobility of devices enables faster access to critical

information for informed decision-making on the fly. On site they can monitor workload, fill in expenses and work done, and continuously report job progress so an engineering organisation's entire workforce can be optimised on the right job at the right time, and meet its service level agreement. However, in order to fully take advantages of mobile technologies, it is an ongoing journey for asset management organisations to undertake.

References

1. Pintelon L, Muchiri PN (2009) Safety and maintenance. In: Handbook of maintenance management and engineering, pp 613–648. isbn 1848824718
2. Dekker R (1996) Applications of maintenance optimization models: a review and analysis. Reliab Eng Syst Saf 51(3):229–240
3. Gouws L, Trevelyan J (2006) Research on influences on maintenance management effectiveness. In: Paper presented at the 2006 world conference on engineering asset management. Surfer's Paradise, Queensland, 11–14 July 2006
4. Soderholm P, Holmgren M, Klefsjo B (2007) A process view of maintenance and its stakeholders. J Qual Maint Eng 13(1):19–32
5. CEN, European Committee for Standardization (2001) Guide on preparation of maintenance contracts. European Standard, Brussels: CEN ENV-13269:2001
6. Koochaki J (2009) Collaborative learning in condition based maintenance. Lect Notes Eng Comput Sci 2176(1):738–742
7. Jardine AKS, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech Syst Signal Process 20(7):1483–1510
8. Sokianos N, Drke H, Toutatoui C (1998) Lexikon productions management. Landsberg, Germany
9. Backman J, Helaakoski H (2011) Mobile technology to support maintenance efficiency-mobile maintenance in heavy industry. In: Proceedings of the 9th IEEE international conference on industrial informatics, pp 328–333
10. Burley L, Scheepers H (2002) Emerging trends in mobile technology development: from healthcare professional to system developer. Int J Healthc Technol Manag 5(3):179–193
11. Campos J et al (2009) Development in the application of ICT in condition monitoring and maintenance. Comput Ind 60(1):1–20
12. Koc M, Ni J, Lee J, Bandyopadhyay P (2003) Introduction of E-manufacturing. In: Refereed proceedings of the 31st North American manufacturing research conference (NAMRC), Hamilton, Canada
13. Muller A, Marquez AC, Iung B (2008) On the concept of E-maintenance: review and current research. Reliab Eng Syst Saf 93(8):1165–1187
14. Yin RK (2003) Case study research: design and methods, 3rd edn. Sage, Thousand Oaks, CA
15. Leonard-Barton D (1988) A dual methodology for case studies: synergistic use of longitudinal single site with replicated multiple sites. Organ Sci 1(3):248–266
16. Adams ME, Day GS, Dougherty D (1998) Enhancing new product development performance: an organisational learning perspective. J Prod Innov Manag 15(5):403–422
17. Donnellan E (1995) Changing perspectives on research methodology in marketing. Ir Mark Rev 8:81–90
18. Eisenhardt KM (1989) Building theories from case studies. Acad Manag Rev 14(4):532–550
19. Neuman WL (2006) Social research methods: qualitative and quantitative approaches. Pearson Education, Boston

20. Snitkin S. (2003) Collaborative asset lifecycle management vision and strategies. Research report, ARC Advisory Group, Boston, USA
21. Syafar F, Gao J (2013) Mobile collaboration technology in engineering asset maintenance: a delphy study. In: Paper presented at the 2013 I.E. 17th international conference on CSCWD, Whistler, BC, Canada, 27–29 June 2013
22. Syafar F, Gao J, Du T (2013) Applying the international delphi technique in a study of mobile collaborative maintenance requirements. In: Refereed proceedings of the 17th Pacific Asia conference on information systems (PACIS), Jeju Island, Korea

Essential Elements in Providing Engineering Asset Management Related Training and Education Courses

Peter W. Tse

Abstract Engineering Asset Management (EAM) is a prime field in managerial and technical importance to well-developed countries and global industry. Proper asset management can make the difference in competitiveness on a global scale. Asset management involves systematic and coordinated activities and practices through which an organization optimally and sustainably manages its assets and asset systems, their associated performance, risks and expenditures over their life cycles for the purpose of achieving its organizational strategic plan. The recent release of the ISO standards on Asset Management, the ISO 55000-2, does attract a number of organizations to offer certificate training programs in physical asset management and a handful of universities to establish EAM based curriculum. As of today, there is no formal regulation to monitor and evaluate the quality of these training programs. The purpose of this forum is to gather expert and general opinions on the development of EAM-based training program and the surveillance of quality in offering related training programs to public. The goal is to establish suitable guidelines and code-of-practice so that any authorized EAM organizations or institutes can use them to grant recognition to these programs for training and education purposes.

Keywords Asset management • Education and training • ISO standards • Certification • Engineering management

1 Introduction

Nowadays, the practice and certification of Engineering Asset Management (EAM) is popular in managerial and technical importance to Hong Kong and global industry. Proper asset management can make the difference in competitiveness on a global scale. Nearly all kinds of industrial sectors require engineering asset management. Examples are public utilities, transportations, manufacturing,

P.W. Tse, Ph.D., C.Eng. P.Eng. (✉)
Department of Systems Engineering & Engineering Management (SEEM), City University of Hong Kong, Tat Chee Ave., Kowloon, Hong Kong, PR China
e-mail: Peter.W.Tse@cityu.edu.hk; http://www6.cityu.edu.hk/seam

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_23

property development, building services, chemical and power plants, finance, information and telecommunication etc. Asset management involves systematic and coordinated activities and practices through which an organization optimally and sustainably manages its assets and asset systems, their associated performance, risks and expenditures over their life cycles for the purpose of achieving its organizational strategic plan. From this course, the students will comprehend a managerial perspective to the maintenance and physical asset management, and introduce an effective strategy for routine maintenance and asset control so that the students are capable of selecting suitable engineering asset management systems for different industrial sectors. The content of this course is especially designed to partially comply with the British Standards on Asset Management (BSi—PAS 55) [1].

Based on PAS 55, asset management is a sophisticated discipline, which embraces management techniques, organizational strategic planning, policy making, finance and asset planning, long term optimized and sustainable asset management, maintenance corrective actions and life cycle management. The importance of engineering asset management has been uplifted significantly in global industry as well as in local (Hong Kong) industry. For instant, the CLP Power Hong Kong Ltd., the Hong Kong and China Gas Co. Ltd., the MTR Corporation Ltd., have changed the name 'Maintenance' to 'Asset Management' after being certified by the BSi standards. Based on the content of PAS 55, in 2014, the International Standards Organization (ISO) released the ISO standards on Asset Management, the ISO 55000-2 [2]. It does attract a number of organizations to offer certificate training programs in physical asset management and a handful of universities to establish EAM based curriculum. Some of the well-established global organizations in asset management include the International Society of Engineering Asset Management, the World Congress on Engineering Asset Management, the asset management council etc.

1.1 Existing Organizations that Provides Services in EAM

The International Society of Engineering Asset Management (ISEAM) is a new multidisciplinary professional learned society dedicated to the development and recognition of Asset Management as an integrated and important body of knowledge [3]. The Society provides thought-leadership and influence on a global basis to coordinate the discipline's advance with academics, practitioners, and policy makers in the emerging trans-discipline of EAM. The ISEAM will also act as a formal organization to recognize future EAM related curriculum and training programs and monitor their progress as well as quality [3]. The World Congress on Engineering Asset Management has been held annually since 2006 commencing with the inaugural event on the Gold Coast in Queensland, Australia [4]. Subsequent congresses have been holding at Harrogate, England (2007); Beijing, China (2008); Athens, Greece (2009) and Brisbane, Queensland (2010), Cincinnati USA (2011),

Korea (2012), Hong Kong (2013), South Africa (2014), Finland (2015) etc. The objective of WCEAM is to bring together academics, practitioners and scientists from all around the world to promote research, development and application in the field of EAM and to strengthen the link between academic researchers and industrial practitioners in the EAM field [4].

1.2 Current Universities that Provide Similar EAM Related Courses

Maintenance Engineering and Asset Management is a critical field of Managerial and technical importance to United Kingdom and international industry [5]. It is estimated that 10% of typical plant cost is spent every year maintaining that plant. Maintenance can make the difference in competitiveness on a global scale. Maintenance managers can make major impacts on their companies' bottom line, and often report at board level. The University of Manchester has offered a program in such aspect [5]. It is a sophisticated discipline, which embraces management techniques, organization, planning and the application of substantial electronic, engineering and analytical know-how to manufacturing processes, transport, power generation and the efficient operation of industrial, commercial and civic buildings. The aim of the program is to give companies the technical and managerial expertise to thrive in the global marketplace. The student can earn the postgraduate certificate or the postgraduate diploma. The University of Western Australia offers a program of Master of Business and Engineering Asset Management [6]. It is an interdisciplinary field that combines the technical issues of asset reliability, safety and performance with financial and managerial skills. The emphasis is on achieving sustainable organizational outcomes and competitive advantage by applying holistic, systematic and risk-based processes to decisions concerning the physical assets of an organization. Physical assets include buildings and fixed plant, mobile equipment and civil, electrical and mechanical infrastructure. The postgraduate degree program is targeted at experienced engineers and managers in operational, maintenance, engineering and asset management roles [6]. Students engage in a balanced interdisciplinary program of asset management, reliability and business units with a focus on practical applications and the challenges faced by today's organizations. The program put particular emphasis on the development of technical skills, decision-making, program implementation, change-management and communication.

2 Essential Subjects Required by Proper Training in EAM

As of today, there is no formal regulation to monitor and evaluate the quality of these training programs and curriculum. The purpose of this investigation is to recommend the important subjects to be developed and taught in EAM-based curriculum and the surveillance of quality in offering related training programs to public. Hence, any institution or university has an intention to offer such EAM-based curriculum can benchmarking her program with the recommended subjects listed here.

From the official document named 'The Asset Management Landscape', issued by the Global Forum on Maintenance and Asset Management, it has stated the scope and definition for asset management training courses. It said that the asset management training courses could be described within an asset management framework by linking the training to the 39 subjects to demonstrate coverage of scope of each course. The 39 subjects are [7]:

- In the group of strategy and planning, it includes AM policy, AM strategy and objectives, demand analysis, strategic planning and AM planning.
- In the group of AM decision-making, it includes capital investment decision-making, operations and maintenance decision-making, lifecycle value realization, resourcing strategy, shutdowns and outage strategy.
- In the group of lifecycle delivery, it includes technical standards and legislation, asset creation and acquisition, systems engineering, configuration management, maintenance delivery, reliability engineering, asset operations, resource management, shutdowns and outage management, fault and incident response, asset decommissioning and disposal.
- In the group of asset information, it includes asset information strategy, asset information standards, asset information systems, data and information management.
- In the group of organization and people, it includes procurement and supply chain management, AM leadership, organizational structure, organizational culture, competence management.
- In the group of risk and review, it includes risk assessment and management, contingency planning and resilience analysis, sustainable development, management of change, asset performance and health monitoring, AM system monitoring, management review, audit and assurance, asset costing and valuation, stakeholder engagement [7].

3 A Tentative Curriculum for EAM to Satisfy the Required Standards and Subjects

A typical post-graduate curriculum that offers EAM-based courses should consider to include the following subjects in the program. The overall Student Intended Learning Outcomes (SILOs) for this EAM curriculum is after the completion of this curriculum, the student should possess the required fundamental knowledge and skills (as listed below) to become a qualified practitioner in EAM. The content and the SILOs are drafted according to of the PAS 55 or ISO 55000-2 and can fulfill the required 39 subjects as follows:

SILO-1: recognize the importance of engineering asset management in manufacturing, public utilities, transportations and building services,
SILO-2: understand the philosophies and international compliance on engineering asset management,
SILO-3: apply all essences of engineering asset management, including the needs, the design, the investment appraisal, purchase, installation, commissioning, operation, quality inspection, maintenance and replacement, disposal of asset etc., into routine practices of asset and maintenance management, and
SILO-4: formulate reliable and cost-effective engineering asset management in selected equipment/process operating in a particular kind of public utility and industry.

To be a fully and qualified graduate studying under this EAM's curriculum, a student must take a total of 30 credit units (CUs). The 30 CUs include 18 CUs taken from the core courses and 12 CUs from a pool of various electives. The 18 CUs of core courses or a total of six courses (3 CU per course) should be selected from the following courses. The Asset and Maintenance Management, the Quality and Reliability Engineering, the Risk and Failure Analysis, the Managerial Decision-making Systems, the Policy Making and Management, the Financial Engineering for Engineering Managers, the Life Cycle Management, the Condition Monitoring, Diagnosis and Prognosis, Statistical & Information Analysis and the Project Management. The other 12 CUs taken as elective courses can be selected from the list core courses that have not been taken as the core courses and the pool of electives courses. Some of suggested electives are the International Business and the Global Business Enterprise, Global Human Resource Management, Emerging Issues in Multinational Strategic Management, Management Science, Quality Improvement: Systems and Methodologies, Marketing Strategy for Engineers, Engineering Enterprise Management, Engineering Economic Analysis, Business Process Improvement and Innovation, Operations Management, Supply Chain Management, Management of Technological Innovation, Dissertation (9 CU per dissertation). After completing the EAM program, a student should possess the required fundamental knowledge and skills to become a qualified practitioner in engineering asset management.

4 Matching of the Courses' Content to the Required 39 Subjects

For the core courses, as aforementioned, a student must select 18 CUs or 6 courses from the following courses that can match with most of the required 39 subjects. The italic words are the course names. The rest of the sentence contains the matched subjects.

1. *Asset and Maintenance Management*: shutdowns and outage strategy and management, asset creation and acquisition, maintenance delivery
2. *Quality and Reliability Engineering*: contingency planning and resilience analysis, audit and assurance
3. *Risk and Failure Analysis*: fault and incident response, reliability engineering, risk assessment and management
4. *Managerial Decision-making Systems*: capital investment decision-making, operations and maintenance decision-making, management review
5. *Policy Making and Management*: AM policy, AM strategy and objectives, demand analysis, strategic planning and AM planning, management of change, management review
6. *Financial Engineering for Engineering Managers*: asset creation and acquisition, resourcing strategy, resource management, asset costing and valuation, stakeholder engagement
7. *Life Cycle Management*: lifecycle value realization, asset operations, asset decommissioning and disposal, sustainable development, management review
8. *Condition Monitoring, Diagnosis and Prognosis*: asset operations, asset performance and health monitoring, AM system monitoring
9. *Statistical and Information Analysis*: asset information strategy, asset information standards, asset information systems, data and information management
10. *Project Management*: resourcing strategy, configuration management, organizational structure, organizational culture, competence management

Besides the requirement of taking six core courses, the student must also select 12 CUs or four courses from the following elective courses listed in the EAM program.

1. *Global Business Enterprise and Planning*: organizational structure, organizational culture, competence management
2. *Global Human Resource Management*: resourcing strategy, resource management
3. *Emerging Issues in Multinational Strategic Management*: organizational structure, organizational culture, competence management
4. *Management Science*: AM leadership
5. *Quality Improvement, Systems and Methodologies*: systems engineering
6. *Marketing Strategy for Engineers*,
7. *Supply Chain Management*: procurement and supply chain management,

8. *Business Law and Contract for Engineers*: technical standards and legislation
9. *Management of Technological Innovation*,

After taking the core courses and the electives listed above, the students should possesses sufficient knowledge and skills to handle and practice the required 39 subjects. The graduated student can classify himself as a qualified specialist in EAM.

5 Recognition Rules and Process

ISEAM is one of the professional organizations that may consider to be an official body for providing audit and certification of AM according to the Asset Management Systems Scheme [8]. Hence, in 2015, ISEAM has setup a task force to draft the recognition process of any curriculum or training program related to EAM that are offered by universities and professional organizations (UPOs) [9]. The recognition process may be initiated by ISEAM, especially to recognize the programs offered by UPOs and represented by ISEAM fellows and members or by a UPO requesting recognition for academic programs that embrace and cover EAM body of knowledge. In either case, a UPO so recognized may be identified as an ISEAM partner institution. Based on the above rules and process, a university that offers a special curriculum in EAM or a UPO provides EAM related training program to industry may seek recognition from official body like ISEAM. Hence, the offered curriculum or program can be recognized by professional body, like ISEAM, and its quality can be continuously monitored and guaranteed.

6 Conclusion

The recent release of the ISO standards, the ISO 55000-2 and the BSi PAS 55 on Asset Management do attract a number of organizations to offer certificate training programs in physical asset management and a handful of universities to establish EAM based curriculum. However, as of today, there is no formal regulation to monitor and evaluate the quality of these training programs and curriculum. Based on the standards' requirements, this paper presents a list of essential subjects and courses in the development of EAM-based curriculum and the surveillance of quality in offering related training programs to public. The ultimate goal is to establish suitable proper guidelines and code-of-practice so that any institution or university can use the recommended curriculum to benchmarking her intended curriculum to ensure it is comply with the requirements of the standards. Hence, the graduated students can be recognized by world-wide organizations and accredited by well-known professional organizations that serve EAM.

Acknowledgments The work described in this paper was fully supported by two grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CityU 11201315).

References

1. The British Standards: BSi PAS 55
2. The International Standards Organization (ISO) released the ISO standards on AM, the ISO 55000-2
3. The website of the International Society of Engineering Asset Management (ISEAM). http://www.iseam.org
4. The website of the World Congress on Engineering Asset Management (WCEAM). http://wceam.com
5. The website of the University of Manchester. http://www.manchester.ac.uk/
6. The website of the University of Western Australia. http://www.uwa.edu.au/
7. 'The asset management landscape', the global forum on maintenance and asset management, 2nd Edn, March 2014. isbn 978-0-9871799-2-0
8. Asset management systems scheme – requirements for bodies providing audit and certification of Asset Management Systems, issue 1, 13 Apr 2015, Asset Management Council
9. 'Brief for recognition of EAM program at higher education institutions – policy and process', a draft proposal for recognition of EAM body of knowledge in academic programs, ISEAM, 2015

Optimizing the Unrestricted Wind Turbine Placements with Different Turbine Hub Heights

Longyan Wang, Andy C.C. Tan, Michael Cholette, and Yuantong Gu

Abstract Wind farm layout optimization is an effective means to mitigate the wind power losses caused by the wake interventions between wind turbines. Most of the research on this field is conducted on the basis of fixed wind turbine hub height, while it has been proved that different hub height turbines may contribute to the reduction of wake power losses and increase the wind farm energy production. To demonstrate this effect, the results of simple two-wind-turbine model are reported by fixing the first wind turbine hub height while varying the second one. Then the optimization results for a wind farm are reported under different wind conditions. The optimization with differing hub heights is carried out using the unrestricted coordinate method in this paper. Different optimization methods are applied for the wind farm optimization study to investigate their effectiveness by comparison. It shows that the selection of the identical wind turbine hub height yields the least power production with the most intensive wake effect. The value of optimum wind turbine hub height is dependent on several factors including the surface roughness length, spacing between the two wind turbines and the blowing wind direction. The simultaneous optimization method is more effective for the complex wind conditions than for the simple constant wind condition.

Keywords Wind farm • Layout optimization • Different hub heights

L. Wang • M. Cholette • Y. Gu (✉)
School of Chemistry, Physics and Mechanical Engineering, Queensland University of Technology, Brisbane 4001, Australia
e-mail: yuantong.gu@qut.edu.au

A.C.C. Tan
School of Chemistry, Physics and Mechanical Engineering, Queensland University of Technology, Brisbane 4001, Australia

LKC Faculty of Engineering & Science, Universiti Tungku Abdul Rahman, Bandar Sungai Long, Cheras, 43000 Kajang, Selangor, Malaysia

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_24

1 Introduction

Wind energy plays a very important role as an alternative energy supply nowadays, due to its properties of renewability and generality on earth. In the year of 2011, the worldwide installed wind power capacity was reported to reach 237 Gigawatt (GW). It is expected that the total installed wind power capacity will be 1000 GW by 2030, as reported by the World Energy Association. For most of them, the utilization of wind energy is achieved in the form of transformation to the electricity using wind turbines. In order to make full use of local wind energy resources, multiple wind turbines are placed in cluster which is called the wind farm or wind park. However, the non-isolated wind turbines bring about the wake interactions, namely, wake interference or wake effect, which greatly reduce the total power production of a wind farm. By optimizing the wind farm layout, the power losses can be regained to a large extent.

Among all the previous wind farm layout optimization studies reported, most of them have considered the scenario with constant wind turbine hub height [1, 2]. However, for a wind farm mostly using the identical type of wind turbine, the hub height of turbines can be selectable. Meanwhile, it is reported that the use of different wind turbine hub heights has the potential to reduce the power losses caused by the wake interaction and hence contribute to the wind farm layout optimization. For the existing literature that report the optimization with different hub heights [3], none of them have claimed to apply the unrestricted coordinate method (use wind turbine coordinate to determine the position in wind farm) to study the layout optimization, which is believed to be superior than the counterpart method. Therefore, this paper attempts to discuss the wind farm layout optimization using the unrestricted coordinated while considering different wind turbine hub heights. The effect of applying different wind turbine hub heights on the power production for both single wind turbine and wind farm is discussed in detail through simple two turbine model and wind farm model. In the meantime, different optimization methods including the single layout optimization, hub height optimization and simultaneous optimization are applied to evaluate their effectiveness under different wind conditions.

The rest of the paper is organized as follows. Section 2 discusses the methods applied for the wind farm layout optimization studies. It includes the optimization algorithm applied for the studies, the calculation of the objective function, the representation of the solution for the objective function using different optimization methods and finally process of the wind farm layout optimization studies using different methods. Section 3 discusses the results and Sect. 4 draws the conclusion.

2 Methods

The methods for the three different types of optimizations are introduced in this section. They are presented in the aspects of optimization solution for the methods, optimization algorithm, objective function calculation and finally the optimization process for the methods

2.1 Optimization Algorithm

For all the three different optimization methods as described above, one feature that they have in common for the solution X is that they all applied the simple real coding method with different number of variables. Therefore, simple Single Objective Genetic Algorithm (SOGA) is employed in this paper to study the optimization of wind farm with different hub heights. GA is a search heuristic that mimics the process of natural selection. It begins with encoded solutions to the optimization problem. The main principle of GA is the maintenance of these encoded solutions which are evolved with the generations to be guided towards the optimum solutions step by step. A simple GA works as follows [4]: a random initial population (a set of encoded solutions) is created. The fitness of individual (the single encoded solution) is evaluated based on the optimization objective function.

1. The raw fitness values are transformed into the range of values that are suitable for the selection process through the fitness scaling procedure.
2. The individuals with the best fitness values are guaranteed to survive to the next generation, while other individuals are used to select parents to produce new population individuals for the next generation.
3. Other new population individuals are generated through the crossover and mutation operators.
4. The current population is replaced with the new generation.
5. Steps 2–6 are repeated until the stopping criteria is met.

2.2 Calculation of Objective Function

Before calculating the wind farm power, the single wind turbine power P_{WT} and the wind speed approaching the rotor of every single wind turbine should be determined first. In this paper, the wind turbine model applied in the reference [5] is employed, and means of determining the wind speed for the turbines affected by the multiple wakes can be referred in the reference [1, 2].

For the discrete wind condition applied in this paper, based on the individual wind turbine power model and the approaching wind velocity for each wind turbine, i-th wind turbine power P_i can be obtained. The total power output P_{tot} with N_{WT} number of wind turbines and finite number of wind directions N_{d} can be calculated as:

$$P_{\mathrm{tot}} = \sum_{j=1}^{N_{\mathrm{d}}} p_j \left[\sum_{i=1}^{N_{\mathrm{WT}}} P_i \right]_{\text{j-th direction}} \tag{1}$$

where p_{j} is the probability of occurrence of j-th wind direction. The wind farm power output is calculated as the accumulation of all wind turbine power output:

$$P_{\text{tot}} = \sum_{i=1}^{N_{\text{WT}}} P_i \tag{2}$$

Based on the calculation of wind farm power output and the wind farm cost models, the Cost of Energy (CoE), which is the objective function for the wind farm optimization study in this paper, can be represented by:

$$\text{CoE} = cost/P_{\text{tot}} \tag{3}$$

where *cost* is the wind farm cost given in the reference [6] and P_{tot} is the wind farm power output calculated above.

2.3 Optimization Solution

The optimization solution X indicates the individual of encoded solution, and each encoded element among them indicates the variable that needs to be optimized through the algorithm.

a. Layout Optimization

For the simple wind farm layout optimization with identical wind turbine hub height, the optimization solution X can be represented as follows:

$$X = \underbrace{\boxed{x_1}\cdots\boxed{x_{\text{i}}}\cdots\boxed{x_{\text{N}}}\boxed{y_1}\cdots\boxed{y_{\text{i}}}\cdots\boxed{y_{\text{N}}}}_{2\ \times\ \text{N}\ \text{ decimal digits}} \tag{4}$$

Where x and y store the X and Y coordinates for different wind turbines. N is the number of wind turbines to be optimized which is predetermined by the user. So X applies the real variable value codification with altogether 2*N number of variables.

b. Hub Height Optimization

For the simple wind turbine hub height optimization, it is conducted based on the result of the optimized wind farm layout through the simple wind farm layout optimization method. The optimization solution X for the optimization method can be represented as follows:

$$X = \underbrace{\boxed{H_1}\cdots\boxed{H_{\text{i}}}\cdots\boxed{H_{\text{N}}}}_{\text{N decimal digits}} \tag{5}$$

where H stores the wind turbine hub height value for different wind turbines and X has N number of variables in total.

c. Simultaneous Optimization

For the simultaneous wind farm layout and wind turbine hub height optimization, the optimization solution X can be represented as follows:

$$X = \underbrace{\boxed{x_1}\cdots\boxed{x_i}\cdots\boxed{x_N}\,\boxed{y_1}\cdots\boxed{y_i}\cdots\boxed{y_N}}_{3 \times N \text{ decimal digits}}\boxed{H_1}\cdots\boxed{H_i}\cdots\boxed{H_N} \tag{6}$$

where it is obvious that the solution X for the method can be regarded as the integration of the two solutions for the above two optimization methods.

2.4 Optimization Process

Figure 1 depicts the general process of the three methods applied in this paper for the wind farm layout optimization studies. Initially, the wind farm design parameters are randomly generated for the different optimization approaches. For the simple wind farm layout optimization method, the random layout parameter (wind turbine position) of the initial population is generated with fixed wind turbine hub height. For the simple hub height optimization method, the random wind turbine hub height value of the initial population is generated while the wind turbine positions are fixed which are imported from the layout optimization results. Unlike these two optimization methods, both wind turbine positions and wind turbine hub height are variable and optimized during the process for the simultaneous optimization. Hence both parameters are initialized randomly as the initial population for this method. After the initialization, all the related wind turbine parameters used for the calculation of wind farm power output are ready using different optimization approaches. Then individuals of the initial population are evaluated according to

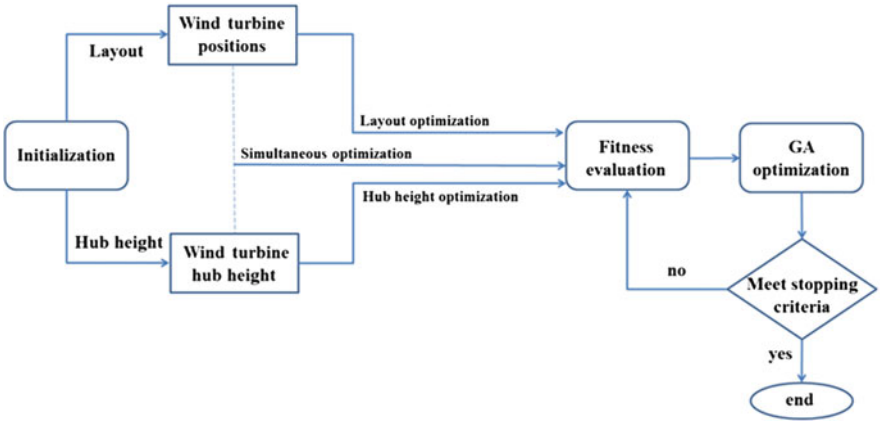


Fig. 1 Process of the wind farm layout optimization study using different optimization methods

the objective function followed by GA optimization procedures. The process excluding the initialization part is repeated until the GA stopping criteria is met.

3 Results and Discussion

Before applying different wind turbine hub heights while designing the wind farm, the simple two wind turbine model is employed to study the influence of the factors including surface roughness, wind turbine spacing and wind direction on the wake-affected wind turbine power production with different hub heights. After this, the wind farm layout optimization study using different optimization methods are carried out and the effectiveness of the methods are compared.

3.1 Two Wind Turbine Model Results

For the two-turbine model, we fix the first wind turbine hub height and study the variation of the second wind turbine power affected by the wake effect (normalized to the free stream wind turbine power) with different hub heights. First, the influence of the surface roughness length (Z_0) on the results is studied with wind direction aligned with the wind turbines and fixed spacing of 15D (D is rotor diameter). Figure 2a depicts the study model while four different surface roughness values are employed from the smooth surface of open sea ($Z_0 = 0.0002$) to the tough surface of high crop obstacle ground ($Z_0 = 0.3$). As can be seen from Fig. 2b, the tougher surface is, the more power wake-affected wind turbine produces in general. When the surface roughness length value is big (0.3), there is no need to

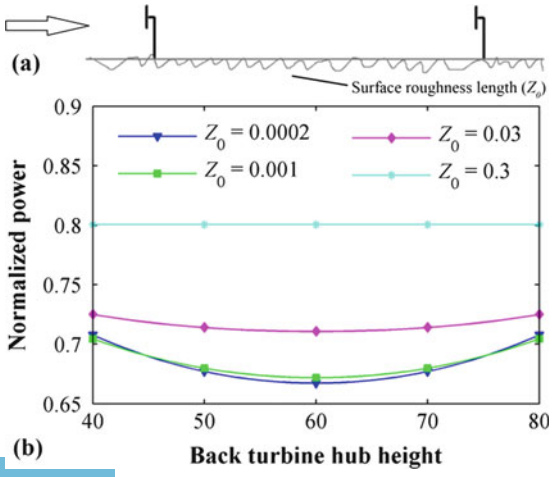


Fig. 2 Variation of normalized wake-affected wind turbine power production with different wind turbine hub height for four characteristic surface roughness lengths

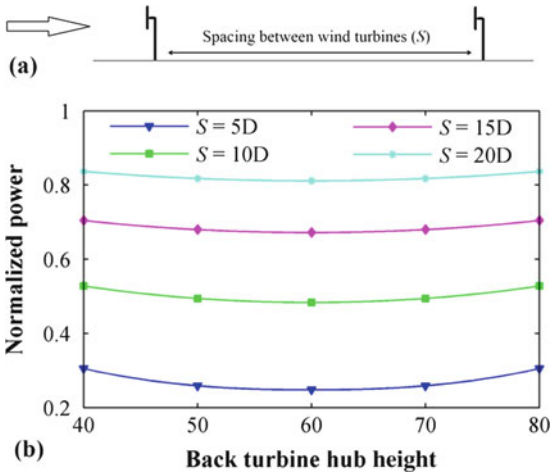


Fig. 3 Variation of normalized wake-affected wind turbine power production with different wind turbine hub height for different spacing between wind turbines

apply different wind turbine hub heights for the back turbine in this case since it is in full wake of the front turbine anyway. When the surface roughness length value is small (Z_0 is less than 0.001), the variation of Z_0 won't have much impact on the power production, and increase of power production by applying different hub height is more prominent. The bigger difference back turbine hub height has from front turbine hub height, the larger power production increase it yields compared with constant hub height.

Figure 3 indicates the results of back turbine power production with different hub heights for different spacing (S) between the two turbines from 5D to 20D (D is the rotor diameter). Obviously, the more distance between turbines, the less wake power losses it has. Also it is found that the percentage of power increase due to the enlarged distance becomes less as it continues. Still when applying different wind turbine hub heights, there is basically the same power increase regardless of the spacing it applied. For the wake-affected back turbine, its power productions with different hub heights are symmetrical with that of the front turbine hub height value (60 m), at which point the back turbine has the most wake power losses.

Next we study the effect of wind scenarios on the power production of wake-affected wind turbine by applying different hub heights. When varying the wind speed with fixed wind direction still aligned with the two wind turbines, it is found that the wake effects are the same for the back turbine with constant normalized power production, and the results are not quantitatively shown here. Figure 4 reports the variation of normalized power production of wind turbine with different hub heights when the wind blowing angle varies from 0° to 10°. The model is illustrated in Fig. 4a and the results are shown in Fig. 4b. As can be seen when the wind direction is very much aligned with the two wind turbines (up to 4°), the power production of the back turbine is the same regardless of the hub heights applied. The power production increases as the wind direction continues to increase, and the power increase due to the use of different wind turbine hub

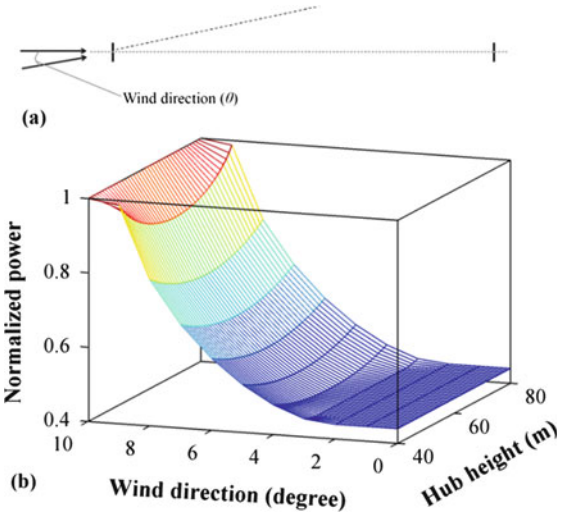


Fig. 4 Variation of normalized wake-affected wind turbine power production with different wind turbine hub height under different wind blowing angles

heights is also increasingly pronounced along with the wind direction (up to 9°). When the wind direction exceeds 10°, the back turbine is not located in the wake of the front turbine anymore, which reaches its maximum power production with free stream wind.

3.2 Wind Farm Results

Three characteristic wind farm conditions, which have been popularly employed for the wind farm layout optimization study, are applied in this paper to compare the effectiveness of different optimization methods for the wind farm with different hub heights.

a. Constant Wind Speed and Wind Direction

Figure 5 reports the wind farm layout optimization results under constant wind speed of 12 m/s and constant wind direction that is aligned with the wind farm length direction (defined as 0°). Figure 5a shows the optimized fitness of the objective function, which is the normalized value for cost of energy production (CoE) using the three methods. The standard deviation of the fitness results with repeated calculation are incorporated in the figure as well. It can be seen that the simultaneous optimization method achieves the best optimization results (the smallest fitness value) through repeated calculation, while it has large deviation and hence relies on the repetition to obtain the best optimization result. When using simple layout optimization method, it also has large deviation. Based on its best optimization result through repeated calculation, the hub height optimization attempts to further optimize the wind turbine hub height but has little improvement.

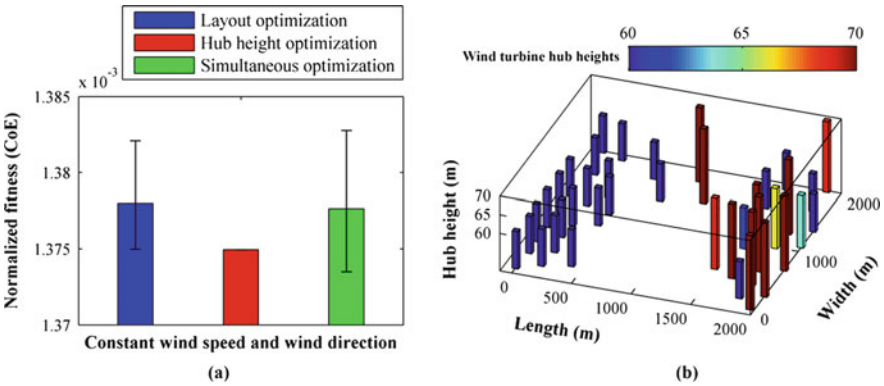


Fig. 5 Results of constant wind speed and constant wind direction aligned with wind farm length: (**a**) normalized fitness of the objective function using different optimization methods (standard deviation depicted) and (**b**) optimal wind farm layout with different turbine hub heights (turbines are represented by the cuboid with heights indicated)

It is also found that the hub height optimization method has no deviation through repeated calculation which means that the results are unchanged and the performance is stable. Figure 5b shows the optimal wind farm layout using different wind turbine hub height after the optimization comparison. It should be noted that the selectable range of wind turbine hub height is from 60 to 70 m in this paper for the wind farm optimization study with all three different wind conditions. The distribution of the wind turbine is obvious for the simple constant wind condition. Nearly half number of turbines is located near to the one side of the wind farm boundary perpendicular to the wind direction in windward direction, and they all have the lower bound of the hub height value which is 60 m. Most of the rest of wind turbines are distributed near to the other normal side of the wind farm boundary in leeward direction. As can be seen, most of them have higher hub height than that of the wind ward turbines to escape from the wake of upstream wind turbines.

b. Constant Wind Speed and Variable Wind Directions

Figure 6 reports the optimization results for the constant wind speed of 12 m/s and variable wind directions. The wind is coming from 36 wind directions evenly distributed with 10° interval, that is 0°, 10°, . . ., and 350°. It can be from Fig. 6a that for this wind condition the simultaneous optimization achieves way much better results than the counterpart method. Meanwhile, the deviation of repeated calculation for the method is relatively small compared with the results of above constant wind condition, which implies the performance of the method is more stable. In comparison, the simple layout optimization method has large deviation and obtains worst optimization results, while the results of hub height optimization are also inferior to the simultaneous optimization method and the deviation of the hub height optimization with repeated calculation is still zero. By observing the optimal wind farm layout shown in Fig. 6b, it is obvious that most of wind turbines are distributed along the four sides of wind farm boundaries. Seven of them have

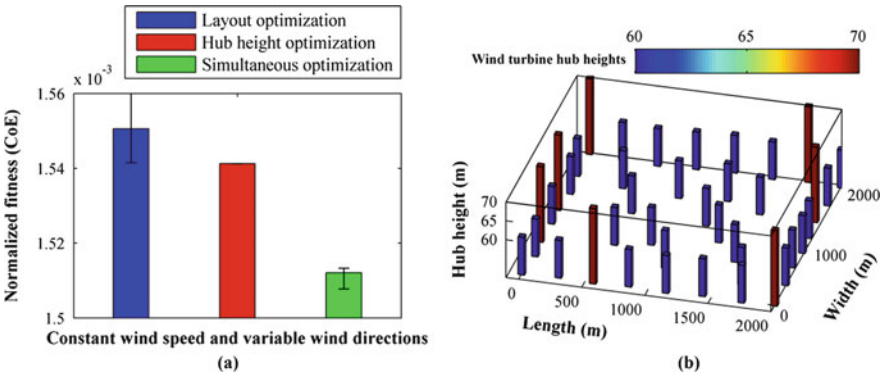


Fig. 6 Results of constant wind speed and variable wind directions: (**a**) normalized fitness of the objective function using different optimization methods and (**b**) optimal wind farm layout with different wind turbine hub heights

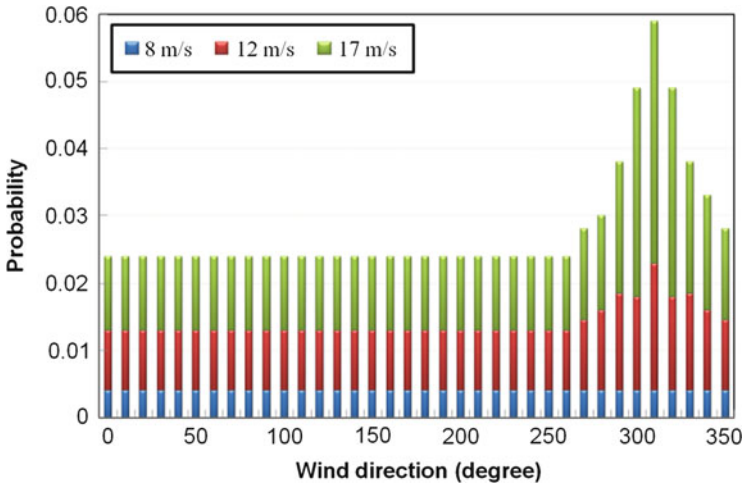


Fig. 7 Histogram of wind condition for variable wind speeds and wind directions

the upper bound of hub height value (70 m), and the rest of wind turbine has hub height of 60 m.

c. Variable Wind Speeds and Variable Wind Directions

Before discussing the results under the variable wind speeds and variable wind directions, the wind condition is introduced here as shown in Fig. 7. The wind condition also consists of 36 wind directions with 10° interval. Every component of the wind directions comprises three different wind speeds and each wind speed in each wind direction is assigned with a value to represent the probability of occurrence. The same wind condition is also applied in the reference [7].

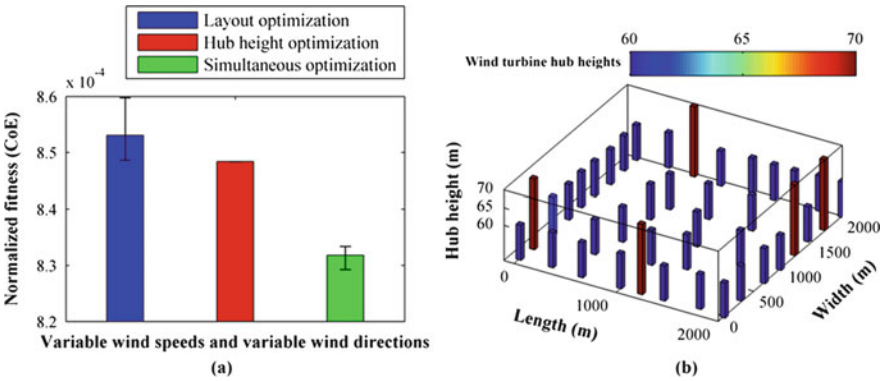


Fig. 8 Results of variable wind speeds and variable wind directions: (**a**) normalized fitness of the objective function using different optimization methods and (**b**) optimal wind farm layout with different wind turbine hub heights

Results of the wind farm layout optimization study with different hub heights for this wind condition are shown in Fig. 8. Like the above condition of constant wind speed and variable wind directions, the simultaneous optimization achieves the best fitness results with the small deviation of repeated calculations. The results using the layout optimization method are the worst with large deviation, based on which there is a small improvement for the results of hub height optimization without deviation of repeated calculation for it. According to the optimal wind farm layout for this wind condition shown in Fig. 8b, most of the wind turbines are spread along the four sides of wind farm boundaries to enlarge the distance between wind turbines with blowing wind from all 360°. Five of them employ the largest wind turbine hub height with 70 m, and for the rest of wind turbines the hub heights of them are near to 60 m.

4 Conclusions

The wind farm layout optimization, which applies the unrestricted coordinate method to determine the wind turbine position, is carried out in this paper for the first time by considering different wind turbine hub heights with same rotor diameter (same wind turbine type). Before introducing the wind farm design results, the simple two wind turbine model is studied in order to discuss the influence of related factors on the wake-affected turbine power output when applying different hub heights. It is found out by altering the back turbine hub height (either increasing or decreasing based on the front turbine hub height value), the power output of the back turbine can be regained. The bigger difference of hub height between the two turbines, the more power output increase it has. However, depending on the distance between the two turbines, the variation of the hub heights

becomes ineffective when the surface roughness length exceeds certain value as the back turbine is always located in the full wake of front turbine. Based on the wake model applied in this paper, with all other fixed factors except for the blowing wind speed the normalized power output for the back turbine is unchanged regardless of the hub height employed. Nonetheless, it is closely related to the blowing wind direction. Under certain value of wind angle, the wake-affected turbine power is fixed as it is always in the full wake of upstream turbine regardless of the hub height employed. As wind direction increases, the effect of altering the turbine hub height on the increase of power output becomes pronounced since it enables the back turbine rotor to jump out of the wake zone or at least decrease the wake-affected area. When the wind direction exceeds certain value, the back turbine is not affected by the wake of front turbine even with constant hub height, and hence the use of different hub heights becomes needless.

Three characteristic wind conditions, which have been widely employed in the wind farm layout optimization studies, are applied in this paper to test the effectiveness of different methods for the different hub height optimization studies of wind farm. For the simplest wind condition of constant wind speed and constant wind direction, the layout optimization is effective enough to obtain the good results by repeated calculation. Based on the results, the improvement of applying hub height optimization is not evident. In comparison, the simultaneous optimization has large deviation with repeated calculation and the improvement of the best optimization results obtained by the simultaneous method is relatively small. For the other two wind conditions, the effectiveness of applying simultaneous optimization method is obvious compared with the counterpart method and the performance is more stable with relatively small deviation of repeated calculation. This is because for these two wind conditions incorporating all wind directions of 360° the single layout optimization is inefficient with constant wind turbine hub height. Based on the inferior fixed wind turbine positions, the hub height optimization has little contribution to the increase of the final wind farm results. In contrast, the simultaneous optimization method facilitates the design to both optimize the wind turbine position and hub height at a time and hence much better results can be achieved.

Acknowledgments The High-Performance Computer resources provided by Queensland University of Technology (QUT) are gratefully acknowledged. The study is financially supported by China Scholarship Council (CSC) from the Chinese government as well as the Top-up scholarship from QUT.

References

1. Wang L, Tan AC, Gu Y, Yuan J (2015) A new constraint handling method for wind farm layout optimization with lands owned by different owners. Renew Energy 83:151–161
2. Wang L, Tan AC, Gu Y (2015) Comparative study on optimizing the wind farm layout using different design methods and cost models. J Wind Eng Ind Aerodyn 146:1–10

3. Chen Y, Li H, Jin K, Song Q (2013) Wind farm layout optimization using genetic algorithm with different hub height wind turbines. Energy Convers Manag 70(0):56–65
4. Mitchell M (1998) An introduction to genetic algorithms. MIT Press, Cambridge, MA
5. Grady S, Hussaini M, Abdullah MM (2005) Placement of wind turbines using genetic algorithms. Renew Energy 30(2):259–270
6. Mosetti G, Poloni C, Diviacco B (1994) Optimization of wind turbine positioning in large windfarms by means of a genetic algorithm. J Wind Eng Ind Aerodyn 51(1):105–116
7. Chen L, MacDonald E (2014) A system-level cost-of-energy wind farm layout optimization with landowner modeling. Energy Convers Manag 77:484–494

Predicting Maintenance Requirements for School Assets in Queensland

Ruizi Wang, Michael E. Cholette, and Lin Ma

Abstract In this paper, a maintenance prediction model is developed for school building assets using a large data set provided by the Queensland Department of Education and Training (DET). DET data on the asset condition, historical maintenance expenditure, and asset characteristics, was analyzed to evaluate which characteristics affect the maintenance needs of the school assets. The condition of the assets was quantified using data on the estimated maintenance backlog. Using statistical methods, models for key building element groups were constructed and the statistical significance of each factor was evaluated. It was found that the school region, the gross floor area, and the maintenance expenditure significantly affected the degradation of key building element groups.

Keywords Maintenance prediction • Funding allocation • Analysis of variance (ANOVA) • Multivariate regression • Influential factors

1 Introduction

School facilities are crucial assets that are designed to provide safe and comfortable places for learning. It is generally acknowledged in the relevant literature that the physical condition can significantly affect teaching, learning, and the health and safety of students and teachers [1–4], making maintenance a critical part of the mission of schools. Yet, setting the appropriate level and allocation of maintenance funds remains a challenge. In Queensland, this is evident from the significant and persistent maintenance backlog, which was estimated at $232 million since the efforts to clear the 2011–2012 school maintenance backlog of $298 million [5].

Funding allocation models for buildings are typically based on the value of the building asset. For instance, in Queensland, the Department of Housing and Public Works [6] recommends a minimum allocation of 1% of the building Asset

R. Wang (✉) • M.E. Cholette • L. Ma
School of Chemistry, Physics and Mechanical Engineering, Faculty of Science and Engineering, Queensland University of Technology, 2 George Street, Brisbane, QLD 4000, Australia
e-mail: r9.wang@qut.edu.au

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_25

Replacement Value (ARV) for annual maintenance budgeting. Such valuation-based funding allocation models are prevalent, however, they fail to account for key attributes that affect facility maintenance needs, e.g. current condition and age [7]. In response to the deficiencies of the valuation-based methods, much of the existing government documents and literature recommend moving toward a predictive approach. However, in order to support this practice, a prediction of future degradation is needed.

This paper details research which aims to lay the foundation for a sophisticated predictive maintenance methodology by constructing a statistical model for forecasting future maintenance needs. To this end, historical maintenance and condition data of Queensland schools have been utilized to identify statistically significant asset characteristics (e.g. geographical location, physical size, enrolment) affecting maintenance needs.

The paper is structured as follows. Section 2 details the data that was available for statistical analysis. Section 3 presents some preliminary analysis targeted at uncovering patterns in the maintenance expenditures. In Sect. 4, a statistical approach to analyze influential factors is described. The resulting significant factors driving the non-uniformity of school maintenance costs are given with discussions in the following Sect. 5. Finally, the conclusions of the research are presented in Sect. 6.

2 Data Description

A considerable amount of data to support this study was collected in conjunction with the Queensland Department of Education and Training (DET). Queensland schools are organized into regions (Fig. 1) which have a number of constituent schools. Each school has a unique identifier (*CIS*) and is comprised of a number of buildings, each of which also has unique identifiers (*BID*). The majority of *BID* s have a construction date which can be used to determine the building's age. Most buildings have a recorded gross floor area (*GFA*) as well. Buildings have a number of element groups (e.g. external finishes and internal finishes) whose condition will be the subject of analysis.

Among the numerous element groups present in the buildings, three critical element groups (*EGC*s) were selected per the advice of at DET: external finishes (EFIN), internal finishes (IFIN), and building structure (BLDG), which are significant maintenance cost drivers for DET. Motivated by literature, expert opinion at DET, and the preliminary analysis (detailed in Sect. 3), the following attributes were considered to potentially affect the degradation of the considered *EGCs*:

1. School gross floor area, $GFA_{CIS}(t)$
2. Building gross floor area, $GFA_{BID}(t)$
3. Enrolment, E_{CIS}
4. Region, r_{CIS}
5. Utilisation, $U_{CIS}(t)$
6. Distance from the Coastline CL_{CIS}

Fig. 1 Regional map of Queensland from [8]

7. Capital expenditures + Planned/unplanned maintenance expenditures, $EXP_{CIS}(t)$[1]
8. Age of the buildings, $AGE_{BID}(t)$
9. Heritage (historical) listing, H_{BID}

The subscript indicates if the attribute is associated with the school or the building.

[1]We actually use the moving average of the past 4 years of maintenance expenditures due to the infrequent inspection intervals and the challenges with aligning the precise dates of maintenance with the inspection times.

3 Preliminary Analyses of Influential Factors to School Maintenance Cost

The condition of a particular element group was quantified by the *indicative cost*, or the estimated cost to repair, and a degradation rate is thus a change indicative cost per unit time. The indicative costs for a selected *EGC* for a particular building (*BID*) can be computed by summing up the indicative costs for all elements (E) that belong to the *EGC*

$$I_{BID,EGC}(t) = \sum_{E \in EGC} I_{BID,E}(t) \tag{1}$$

where $I_{BID,E}(t)$ is the indicative cost of element E at time t and the expression $E \in EGC$ (in an abuse of notation) indicates that an element belongs to a particular *EGC*. It is implicitly assumed that the indicative cost on un-inspected elements is zero. That is, a report only appears when the element is in need of repair.

The *BID* level indicative costs may be used to assess the impact of the building-level candidate factors: building age, building *GFA*, and heritage listing. However, most of the analysis of the indicative cost patters was conducted at the school level (*CIS*) by summing the indicative costs for each *BID* for the *EGC* of interest

$$I_{CIS,EGC}(t) = \sum_{BID \in CIS} I_{BID,EGC}(t) \tag{2}$$

The reason for conducting the analysis at the school level is that the majority of the candidate influential factors are specified at the *CIS* level. By examining the indicative cost of the school, the effect of sparse of building-level condition assessments is less pronounced, provided that the element group is inspected on *some* buildings for a school in a particular year.

Figure 2 shows the time history of EFIN indicative cost per m^2 (i.e. *GFA*), which we use as an indication of school size.[2] The results clearly indicate that there are significant regional differences in the various regions of Queensland: Darling Downs South West, North Queensland, and Central Queensland have significantly higher EFIN costs per m^2 than the remaining regions. This is important to note for any prediction, since increasing the GFA in one of these regions would have a larger effect on total EFIN maintenance costs than elsewhere.

Figure 3 shows per m^2 indicative costs for IFIN. From these plots it is observed that the total maintenance work to be done for IFIN has stabilized and is now (slightly) decreasing for most regions. It is also noted that Darling Downs South West and Central Queensland are the regions that have the worst IFIN indicative costs per m^2.

[2]One could also use the number of students to normalize the costs. However, motivated by our later statistical analysis, we only show *GFA*.

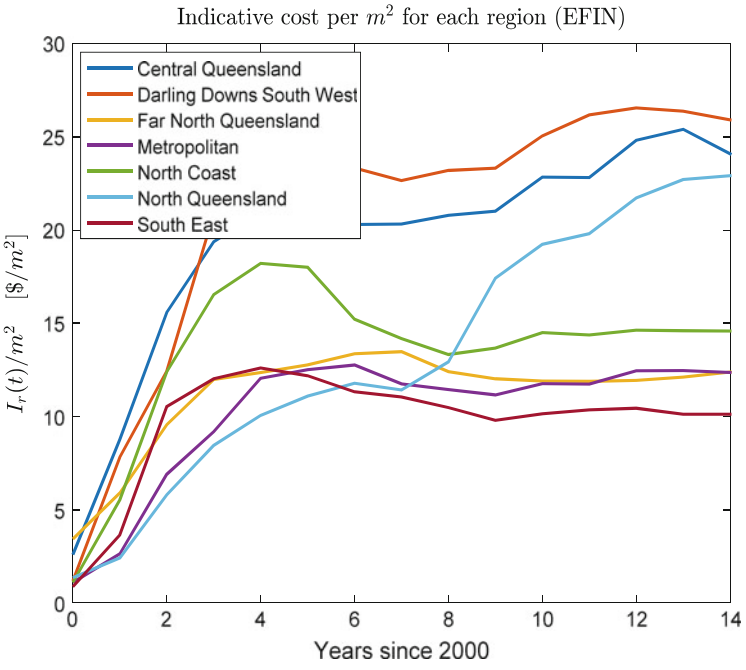


Fig. 2 Indicative cost per m^2 of EFIN for each region

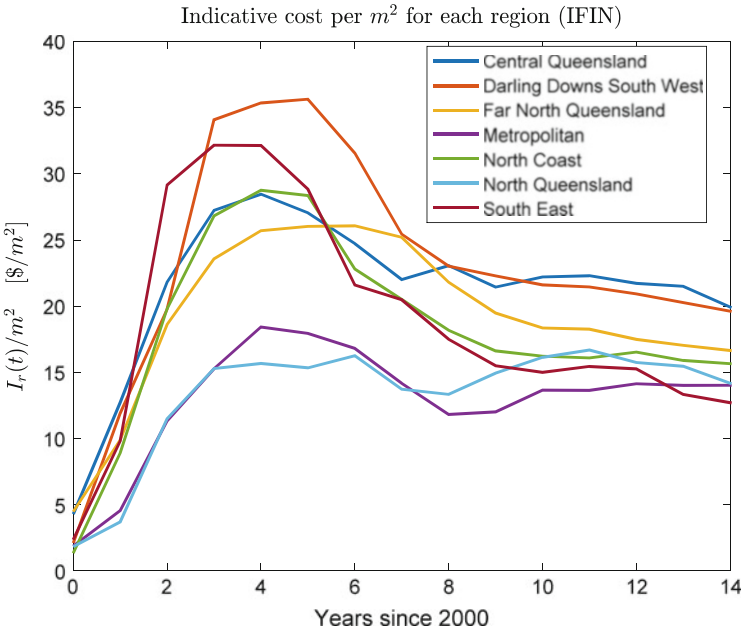


Fig. 3 Indicative cost per m^2 of IFIN for each region

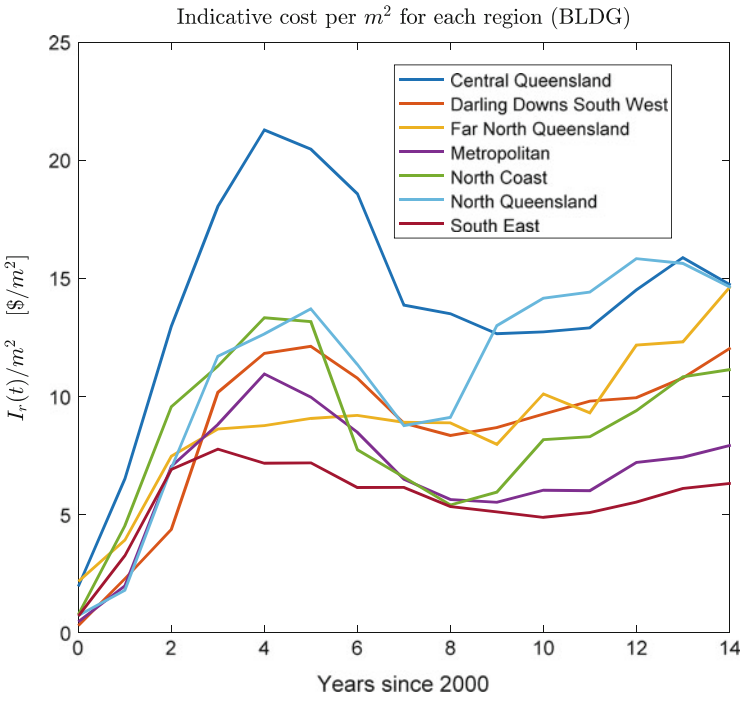


Fig. 4 Indicative cost per m^2 of BLDG for each region

Finally, the regional trends in BLDG condition were examined. Figure 4 shows per m^2 indicative costs that Central Queensland, and North Queensland, and Far North as have the highest indicative costs per m^2. It is also noted that most of the regional indicative costs have decreased significantly, but have started to increase again.

Clearly, there are regional differences even after normalisation by *GFA*. This variation motivates the inclusion of the region as a potential influential factor. Examining the map of the regions in Queensland in Fig. 1, we can see a common factor of the high-cost regions: they all have significant inland regions, motivating the inclusion of the distance from the cost as a potential influential factor as well.

4 Model Development

In this section a model for the degradation of the *EGCs* is developed. The condition of a particular *EGC* was quantified by the indicative cost, so degradation is simply the change in total indicative cost of an *EGC* per unit time.

$$\Delta I_{CIS,EGC}\ (t) = \frac{I_{CIS,EGC}(t) - I_{CIS,EGC}(t_\ell)}{t - t_\ell} \tag{3}$$

where t_ℓ is the time of the last inspection. We will hereafter refer to $\Delta I_{CIS,EGC}\ (t)$ as the *school indicative cost rate* and analogously define $\Delta I_{CIS,BID}$, the building indicative cost rate as an obvious modification to Eq. (1).

Two statistical tools were employed to assess the significance of the potential factors: multivariate regression will be used to assess the strength of the effect of each factor while and analysis of variance (ANOVA) will be used to assess the strength of the evidence that the factor is important. The difference between these "strengths" is subtle but important: a high strength of effect indicates that the factor has a large influence on the indicative cost rate while a high strength of evidence means that this effect is real and not just a statistical fluke. Accordingly, our working model for examining the significance of each factor at the school level was formed as follows:

$$\begin{aligned}\Delta I_{CIS,EGC}(t) = {} & \alpha_1 GFA_{CIS} + \alpha_2 U_{CIS}(t) + \alpha_3 EXP_{CIS}(t) + \alpha_4 E_{CIS}(t) + \tau_r \\ & + \alpha_5 CL_{CIS} + b + \varepsilon(t)\end{aligned} \tag{4}$$

and for the building level as

$$\Delta I_{BID,EGC}(t) = \alpha_1 GFA_{BID} + \tau_r + \alpha_2 CL_{BID} + \alpha_3 Age_{BID}(t) + \tau_{HBID} + b + \varepsilon(t) \tag{5}$$

where $\varepsilon(t)$ is the prediction error and the remaining variables are defined in Sect. 2. To assess the strength of evidence, we will examine the p-values resulting from the ANOVA. We will adopt the conventional (albeit somewhat arbitrary) $p < 0.05$ as statistically significant factors, i.e. we wish to have less than a 5% chance that the factor is falsely stated as significant.

The school level model of Eq. (4) will be our primary tool for exploring the strength of the candidate factors. However, the age of the school isn't meaningful since each building can be of significantly different age. Heritage listing is also a building level property. Thus, we also employ a building level model to assess the influence of heritage listing and age.

The procedure for assessing a factor's significance was as follows:

1. Begin by considering one factor at a time. Eliminate all factors that are not statistically significant on their own.
2. For all the statistically significant factors from step 1, fit a model with the first level factors (i.e. "main effects" in ANOVA terminology).
3. If all $p < 0.05$ stop. Report factors. If not, exclude the factor with the highest p-value and repeat steps 2–3.

In the following sections, the results of statistical analysis for each *EGC* at the school and building level will be detailed. For the following results, the regions are encoded as in Table 1.

Table 1 Region numbers

Region name	Number
Central Queensland	1
Darling downs south west	2
Far North Queensland	3
Metropolitan	4
North coast	5
North Queensland	6
South east	7

5 Results and Discussions

5.1 School-Level Results

For EFIN, the identified influential variables were displayed in Table 2. The results demonstrate GFA_{CIS} positively influences the changes in the indicative cost while $EXP_{CIS}(t)$ negatively influences it. This is of course intuitive: a larger floor area is more difficult to maintain and maintenance spending should improve the condition, decreasing the indicative cost.

The region, r, was also identified as significant influential factor with approximately zero p-value. Regional coefficients τ_r in red of Darling Downs South West, North Queensland and Central Queensland are significantly higher than other regions, confirming our *ad hoc* analysis in Sect. 3. Figure 5 shows a 1D projection of the prediction along the GFA factor to give a sense for the quality of the fit.

Table 3 shows the results for IFIN. We once again find that GFA_{CIS}, $EXP_{CIS}(t)$, are significant, however r is just outside our significance cut-off, and we briefly depart from convention to show its coefficients. We see that the high-cost regions are once again represented (1, 2, and 6), but we also note that the regional coefficient of the South East (7) region is high. The South East region has the third highest GFA, which is directly in line its third-highest indicative cost (Fig. 3). However, the fast rise at the beginning of the series leads to the large estimated τ_7 (Table 3). This may be due in part to the "start-up" effect noted in Sect. 3, but it was decided that the data series was too short to leave out these early points for this exploratory analysis.

The results for BLDG can be seen in Table 4. Following our procedure, the influential variables would have been GFA_{CIS} and, to a lesser extent, E_{CIS} (p-value $= 0.015$). However, E_{CIS} was excluded from the influential variable list for two reasons:

1. E_{CIS} is highly correlated with GFA_{CIS} (correlation coefficient of 0.88), meaning that they contain very similar information.
2. The coefficient of E_{CIS} was negative, which makes no physical sense since more students would *decrease* the indicative cost.

Table 2 Influential variables for EFIN on school level

Variable	Coefficient(s)	p-Value
GFA_{CIS}	$\alpha_1 = 1.54$	$p \approx 0$
U_{CIS}	–	Insignificant
EXP_{CIS}	$\alpha_3 = -0.73$	$p \approx 0$
E_{CIS}	–	Insignificant
r	$\tau_1 = 4722$	$p \approx 0$
	$\tau_2 = 3919$	
	$\tau_3 = -6113$	
	$\tau_4 = -5441$	
	$\tau_5 = 1064$	
	$\tau_6 = 7110$	
	$\tau_7 = -5260$	
CL_{CIS}	–	Insignificant
b	$b = 8022$	

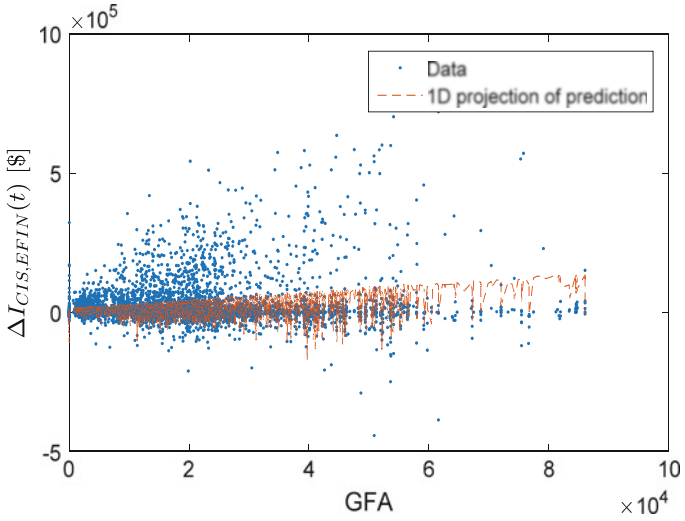


Fig. 5 Prediction of change of indicative cost from the fitted model

Point 1 is supported when the second level factors of ANOVA are considered (i.e. "interaction" terms). We find that the interaction term is significant, but the main effect for E_{CIS} is not. This means that E_{CIS} influences the result only through GFA. We thus neglect it as a significant factor on its own, but a more sophisticated prediction model may need to consider this nonlinear effect for optimal accuracy.

We also note that maintenance expenditure was not significant for BLDG. This is likely due to the fact that in the data available, there is no straightforward way to extract only maintenance expenditures that are related to BLDG alone. Thus, the

Table 3 Influential variables for IFIN on school level

Variable	Coefficient(s)	p-Value
GFA_{CIS}	$\alpha_1 = 3.05$	$p \approx 0$
U_{CIS}	–	Insignificant
EXP_{CIS}	$\alpha_3 = -1.90$	$p \approx 0$
E_{CIS}	–	Insignificant
r	$\tau_1 = 2828$	$p \approx 0.0758$(insignificant)
	$\tau_2 = 1940$	
	$\tau_3 = 327$	
	$\tau_4 = -9874$	
	$\tau_5 = -3475$	
	$\tau_6 = 4021$	
	$\tau_7 = 4231$	
CL_{CIS}	–	Insignificant
b	$b = 7135$	

Table 4 Influential variables for BLDG on school level

Variable	Coefficient(s)	p-Value
GFA_{CIS}	$\alpha_1 = 0.59$	$p \approx 0$
U_{CIS}	–	Insignificant
EXP_{CIS}	–	Insignificant
E_{CIS}	–	Neglected (See text)
r	–	Insignificant
CL_{CIS}	–	Insignificant
b	$b = 4072$	

entire maintenance expenditures must be used, much of which did not pertain to the external finishes.

Based on the analysis for the three element groups EFIN, IFIN and BLDG on school level, *GFA* and region (r) have been identified as strongest influencing the degradation on school level. Maintenance expenditure does have an effect, but only when expenditure data related directly to the EGC can be extracted, as was done for EFIN and IFIN.

5.2 *Building-Level Results*

A building-level model was proposed in Eq. (5) and the identified influential variables for EFIN and IFIN are displayed in Tables 5 and 6, respectively. For EFIN, the identified influential variables at the building-level are AGE_{BID} and region, while for IFIN the influential variable is region alone. Interestingly, GFA_{BID} is not a significant factor. This may be due to the lack of maintenance information, which adds uncertainty to the analysis and makes it more difficult to find smaller correlations. It might also be that at the school level, *GFA* is a proxy for the number

Table 5 Influential variables for element group EFIN on building level

Variable	Coefficient(s)	p-Value
GFA_{BID}	–	Insignificant
AGE_{BID}	$\alpha_2 = 11.5$	$p = 0.013$
r	$\tau_1 = 91.5$	$p \approx 0$
	$\tau_2 = 217$	
	$\tau_3 = -273$	
	$\tau_4 = -751$	
	$\tau_5 = -236$	
	$\tau_6 = 1414$	
	$\tau_7 = -461$	
H_{BID}	–	Insignificant
b	$b = 325$	

Table 6 Influential variables for element group IFIN on building level

Variable	Coefficient(s)	p-Value
GFA_{BID}	–	Insignificant
AGE_{BID}	–	Insignificant
r	$\tau_1 = -4.0$	$p \approx 0$
	$\tau_2 = 111$	
	$\tau_3 = -344$	
	$\tau_4 = 598$	
	$\tau_5 = 712$	
	$\tau_6 = 1140$	
	$\tau_7 = -1837$	
H_{BID}	–	Insignificant
b	$b = 536$	

of buildings to be maintained, while this is not so at the building level (there's only one).

Examining the EFIN regional coefficients, we see that the three largest regional coefficients are indeed the same as at the school level, further supporting the idea that these regions are indeed significantly different in terms of EFIN degradation. On the other hand, for IFIN we see that the regional effects with the high τ_r values correspond to the high un-normalized indicative cost regions. This is likely due to the fact that the lack of maintenance expenditure information has left *region* as the only significant factor at the building level, forcing it to explain all of the variation in the indicative cost rates alone.

The results for BLDG at the building level can be seen in 7. We clearly see that all considered variables are significant, with a small effect from the heritage listing. Interpreting the regional effects with regard to Fig. 5 also suggests that Central and Far North Queensland are confirmed as being more severe regions for BLDG degradation. However, the high degradation rates in North Queensland appear to be explained in part by other factors included in the model.

Table 7 Influential variables for element group BLDG on building level

Variable	Coefficient(s)	p-Value
GFA_{BID}	$\alpha_1 = 0.57$	$p \approx 0$
AGE_{BID}	$\alpha_2 = 15.05$	$p \approx 0$
r	$\tau_1 = 866$	$p \approx 0$
	$\tau_2 = -254$	
	$\tau_3 = 720$	
	$\tau_4 = -1525$	
	$\tau_5 = 172$	
	$\tau_6 = 111$	
	$\tau_7 = -90$	
H_{BID}	$\tau_{H_{WIC}} = \begin{cases} -304 & H_{WIC} = 0 \\ 304 & H_{WIC} = 1 \end{cases}$	$p = 0.003$
b	$b = 438$	

6 Discussion and Conclusions

Based on the statistical analysis on both school and building levels, influential factors can be concluded as follows:

- Region, *GFA*, *Age* are confirmed to be influential factors correlated to indicative cost rates
- Maintenance expenditure effect is a significant factor when we can extract the maintenance that pertains to a particular *EGC*.
- Heritage listing has a statistically significant effect on BLDG indicative cost, but the effect is small relative to *GFA* and *AGE*.

The models established here confirm the feasibility of constructing a model to predict the indicative costs over time. Furthermore, for the models that include expenditure, we can have a coarse estimate of the expenditure required to stabilize the indicative cost by setting $\Delta I_{CIS,EGC}(t) = 0$ and solving for $EXP_{CIS}(t)$ in Eq. (4).

The significant influential factors on school maintenance cost have been identified through statistical analysis of Multivariate Regression and ANOVA. From the available data region, *GFA*, and *Age* are verified as influential factors in the degradation school facilities. Maintenance expenditure is clearly significant as well, but only when more targeted maintenance expenditure information can be obtained using the descriptions or maintenance programs.

Based on the model presented in this paper, the effects of future school assets can be predicted. For instance, it can be seen that expanding school assets in the inland regions (e.g. Darling Downs South West) will have a larger impact on external finish maintenance costs than expansion elsewhere. Additionally, ageing schools are likely to lead to larger maintenance costs; a factor which will play an important role in budget planning and asset renewal decisions.

Clearly, the proposed (linear) models can provide a prediction of the maintenance cost for these new assets. However these simple statistical models here were

indented as tools for exploring the statistical significance of various factors and not focused on the accuracy of the prediction. We have also ignored the interaction effects of various factors, which a more sophisticated regression could use to enhance the prediction accuracy. For a future funding allocation model, it is recommended for inclusion in a funding allocation model.

Acknowledgments This project was funded by the Queensland Department of Education and Training (DET). The authors wish to thank Ariane Panochini, Greg Duck, Malvin White, and Nadeia Romanowski for the in-depth discussions and insight which greatly aided the research presented in this paper.

References

1. Lawrence BK (2003) Save a penny, lose a school: the real cost of deferred maintenance. Policy Brief Series on Rural Education
2. Lyons JB (2001) Do school facilities really impact a child's education. Council of Educational Facility Planners International, Scottsdale, Arizona (ERIC reproduction service no. ED 458791)
3. Mahli M, Che-Ani A, Tawil MA-RN, Yahaya H (2012) School age and building defects: analysis using condition survey protocol (CSP) 1 matrix. World Acad Sci Eng Technol 6 (7):1830–1832
4. Schneider M (2002) Do school facilities affect academic outcomes? Academic Achievement 25
5. Queensland Audit Office (2015) Maintenance of public schools (Report 11: 2014–15), edited by Queensland Audit Office
6. Queensland Department of Housing and Public Works (2012) Policy for the maintenance of Queensland government buildings, maintenance management framework
7. Bello MA, Loftness V (2010) Addressing inadequate investment in school facility maintenance. School of Architecture, Paper 50
8. Department of Education and Trainin, Queensland Government (2016) Regional map of Queensland. http://education.qld.gov.au/hr/recruitment/teaching/locations.html. Accessed 8 Feb 2016

Research on Armored Equipment RMS Indexes Optimization Method Based on System Effectiveness and Life Cycle Cost

Zheng Wang, Lin Wang, Bing Du, and Xinyuan Guo

Abstract Requirements of equipment performance improve continually as the combat power generation mode keeps transmitting. It becomes more and more significant to improve reliability, maintainability and supportability (RMS) level of newly-developed equipment. Aiming at the optimization and design of newly-developed armored equipment RMS indexes, this paper analyses the relationship between RMS indexes and system effectiveness, life cycle cost, respectively. On condition of the certain life cycle cost, it sets up an RMS indexes optimization model of armored equipment to maximize system effectiveness. By using parameter time-varying particle swarm optimization (PSO), this paper attains a series of RMS indexes value, which provides methods and gist for armored equipment RMS argumentation and design in future.

Keywords RMS indexes • System effectiveness • Life cycle cost • Optimization

1 Introduction

As the composite development of our army armored equipment mechanization and informatization in new century, the combat power generation mode transmits faster, especially, requirements of equipment performance are advanced by military conflict improve continually. The level of RMS would directly determine the exertion of battle effectiveness; it has great significance to upgrade the ability of winning the battle. Although the newly-developed equipment of our army armored equipment continually increase in recent years, it becomes a huge problem to design equipment RMS indexes so that we could maximize the system effectiveness on condition of a certain life cycle cost, which is also a key research content.

General models of system effectiveness contain WSEIAC model (ADC model), ARINC model and navy system effectiveness model [1]. Lots of researchers in domestic and overseas have already carried out related research on equipment RMS

Z. Wang (✉) • L. Wang • B. Du • X. Guo
Department of Technology Support Engineering, Academy of Armored Forces Engineering, Beijing, PR China
e-mail: pro.wangzheng@hotmail.com

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_26

indexes tradeoff analysis based on system effectiveness. Huixin Xiao etc. provided a gist to integrate tradeoff and optimization of unrepairable equipment by putting forward a method of RMS synthetical effectiveness model and indexes design analysis [2]. Directing at unrepairable and repairable faults of equipment during executing missions, Yong Yu etc. put forward a system effectiveness evaluation model based on RMS characters and got an expression of system effectiveness [3]. Jinzhuo Li etc. studied RMS indexes design problem of military aircraft with simulation. They analyzed an instance of the military aircraft RMS indexes design with Monte Carlo simulation method and got a RMS indexes solution set which could afford the support effectiveness requirements [4]. Dianfa Ping etc. built an index evaluation system of airborne EW system with combination of ADC model. They chose appropriate partition granularity system based on the objective structure of airborne EW system, to analyze the system reliability, availability, capabilities and building an evaluation model [5]. Wenhua Peng etc. established a mapping relationship between RMS and system effectiveness with WSEIAC model, and got an intuitional curve of RMS-effectiveness by computer simulation, which provided a gist for RMS indexes tradeoff in naval gun weapon system [6]. Kun Han etc. established a system effectiveness model of armored vehicles, which consists of operational readiness, battle dependability and capability. Aiming at system effectiveness, they proposed two RMST tradeoff analysis method, which provided an efficient means to reasonably determine the armored vehicles RMST quantitative requirements and improve system effectiveness [7]. Haidong Du etc. proposed synthetic expression of system effectiveness and related constraints based on equipment RMS indexes, and they also built a tradeoff dynamic simulation model of system effectiveness. Via simulation, they screened and optimized index schemes so that they could provide a reference for RMS design and verification of armored equipment [8]. Considering the relationship between equipment RMS indexes and life cycle cost, Geng etc. presented the tradeoff for RMS indexes based on the availability calculation models and LCC models, on the condition that the availability value is not below the given value, the decision-making is optimal if the life cycle cost is the minimum value [9]. American expert Charles E. Ebeling proposed a life cycle cost general model [10] and modified model etc. [11], which provided gist for the calculation of life cycle cost.

This paper establishes a system effectiveness model and life cycle cost model based on RMS indexes, and proposed a RMS indexes optimization model which aims to maximize system effectiveness of armored equipment, on condition of the certain life cycle cost. The optimal RMS indexes are gotten by using parame-ter time-varying PSO algorithm, which provides reference for practical equipment development.

2 Optimization Model For RMS Indexes Design

2.1 System Effectiveness Model Containing RMS Indexes

System effectiveness refers to the ability of fulfilling given quantitative traits and service requirements in given condition, which compositely reacts system availability, mission dependability and inherent capability. Relationship between system effectiveness and the three elements could be reflected from model, which is the most widely applied model.

2.1.1 Availability Function

As a measure parameter of operational readiness, operational availability A_0 is used to describe availability A. Taking no account of alert time, non-operation time, reaction time and administrative delay time, A could be described as formula (1):

$$A = A_O = \frac{T_{MTBF}}{T_{MTBF} + T_{MTTR} + T_{MPT} + T_{MLDT}} \tag{1}$$

where T_{MTBF} is mean time between failures, belonging to reliability indexes; T_{MTTR} is mean time to repair, belonging to maintainability indexes, which reflects the complexity of corrective maintenance; T_{MPT} is mean time to preventive maintenance, belonging to maintainability indexes, which reflects the complexity of preventative maintenance; T_{MLDT} is mean logistics delay time, belonging to supportability indexes, which reflects the complexity of planned support resource adequacy and applicability, as well as support system effectiveness. Apparently, availability A is totally decided by RMS indexes, preventative maintenance regime, administration, operation, support and various kinds of elements.

2.1.2 Dependability Function

Mission availability refers to the ability of usage and fulfilling given function at any random moment in specified mission section. Mission availability describes whether equipment could continuously keep working or not, which measures the level of required function that panzer could reach during the mission. It is affected by mission reliability, mission maintainability, security and survivability etc. In the mission interval, equipment come into failures at a certain probability, and the faulted equipment go back to the mission at a certain renovating probability. The condition conversion of equipment between good condition and failure obeys the Markov process. Mission dependability in moment t could be described as formula (2):

$$\begin{aligned} D(t) &= e^{-\lambda t}(1-\lambda) + \left(1-e^{-\lambda t}\right)K\mu \\ &= K\mu + (1-\lambda-K\mu)e^{-\lambda t} \end{aligned} \tag{2}$$

Equipment have to be prepared to next mission after mission at present, it is both concerned that the ratio of required function the equipment possess during and after mission. Mean mission dependability $\overline{D_b}$ has a relationship with the whole mission process, it could reflect reliability in specific mission and the ability of maintenance support. Mission dependability D could be described by Mean mission dependability $\overline{D_b}$ as formula (3).

$$\begin{aligned} D = \overline{D_b} &= \frac{1}{T}\int_0^T D(t)dt = K\mu + \frac{1-\lambda-K\mu}{\lambda T}\left(1-e^{-\lambda T}\right) \\ &= \frac{K}{T_{MRF}} + \frac{T_{MRF}T_{MTBCF} - T_{MRF} - KT_{MTBCF}}{T_{MRF}T}\left(1-e^{\frac{-T}{T_{MTBCF}}}\right) \end{aligned} \tag{3}$$

where λ is failure rate, belonging to supportability indexes, which is the probability equipment convert from good to failure; μ is repair rate, belonging to maintainability indexes, which is the probability equipment convert from failure to good; K is spare sufficiency, belonging to supportability indexes; T_{MRF} is mean time to restore function, which is related to mission function restoration; T_{MTBCF} is mean time between critical failures, belonging to reliability indexes, which is related to mission function maintenance; T is mission execution time.

2.1.3 Capability Function

Capability is a measurement of mission execution, which describes whether equipment could finish given mission under normal condition during the whole mission and is affected by elements such as operating distance, accuracy, power, lethality and etc. Conceptually, RMS is not concluded in capability, but it determines how the capability could be exerted. In this model, capability C is supposed to be a constant.

2.1.4 System Effectiveness Model

According to availability function A, dependability function D and capability C, we could get system effectiveness by formula (4) with model $E = ADC$:

$$\begin{aligned} E = ADC = &\frac{T_{MTBF}}{T_{MTBF} + T_{MTTR} + T_{MPT} + T_{MLDT}} * \\ &\left(\frac{K}{T_{MRF}} + \frac{T_{MRF}T_{MTBCF} - T_{MRF} - KT_{MTBCF}}{T_{MRF}T}\left(1-e^{\frac{-T}{T_{MTBCF}}}\right)\right) * C \end{aligned} \tag{4}$$

2.2 Life Cycle Cost Model Containing RMS Indexes

RMS is important influencing factors for equipment, as well as factors that affect equipment life cycle cost. According to the influence that RMS act on life cycle cost, we modify the general model set up by American expert Charles E. Ebeling and build a life cycle cost mathematical model based on RMS indexes [10] which is described as formula (5).

$$LCC = F_z\left(\frac{T_{MTBF_0}}{T_{MTBF}}\right)^{\beta} + F + A_0C_0 + A_0\frac{t_0}{T_{MTBF}}C_f - S_a \tag{5}$$

where LCC is life cycle cost; F_Z is the cost of equipment and labor to produce a reliable system; β is a constant; T_{MTBF_0} is the statistic of mean time between failures when equipment is used in practical mission; F is the fixed use cost; C_0 is annual cost of equipment; t_0 is annual use time of equipment; C_f is the fixed cost of every failure; S_a is net residual value of equipment.

With those formulas above, RMS indexes and other parameters, we could predict the life cycle cost of equipment.

2.3 Optimization Model for RMS Indexes Design

Aiming at maximizing system effectiveness, we build a RMS indexes optimization model under a certain life cycle cost as follows.

$$\text{Max } E = E(x_1, x_2, \cdots, x_n) \tag{6}$$

$$\text{S.t.} \quad LCC < C_{LCC} \tag{7}$$

$$a_i \le x_i \le b_i \quad i = 1, 2, \cdots, n \tag{8}$$

$$a_i, b_i, C_{LCC} \ge 0 \tag{9}$$

where x_i is design variable, indicates basic performance indexes of RMS, such as T_{MTBF}, T_{MTTR} and so on; n is the amount of basic performance indexes; a_i, b_i indicate domain of definition of x_i. To fulfill military requirement, indexes need to set threshold value. To fulfill plan feasibility, indexes need to set target value. So, in condition that when indexes are smaller, system effectiveness is larger, a_i indicates threshold value, indicates target value; to the opposite, a_i indicates target value, indicates threshold value; $E(x_1, x_2, \cdots, x_n)$ indicates system effectiveness as formula (4) shows; LCC is life cycle cost as formula (5) shows; C_{LCC} indicates the constrained value of life cycle cost.

3 Parameter Time-Varying Particle Swarm Optimization

Particle swarm optimization, an evolutionary computation algorithm proposed by Kennedy, is, in fact, an optimizer, based on the food-searching behavior of birds flocking. In standard PSO algorithm, a swarm of particles, on behalf of candidate solutions, are initialized and move through the search space towards the compromise between the best positions historically and the best among all particles until converging to the optimum [12].

3.1 Structure of Parameter Time-Varying PSO

3.1.1 Particles Coding Method

Particle $X_j = \{x_{j1}, x_{j2}, \cdots, x_{jn}\}$ indicates a design scheme of RMS parameters, where x_{ji} is the value of the ith parameter in the design scheme j.

3.1.2 Evaluation Function

Evaluation function is calculated by formula (4), which is the effectiveness function.

3.1.3 Initial Position and Velocity

Initial position of particle X_j is x_{ji}, random value in domain of definition $[a_i, b_i]$. Initial velocity of particle V_j equals half of the length of interval, $V_j = (a_i + b_i)/2$.

3.1.4 Updating Velocity

Each particle updated the velocity based on the last velocity and the best position of history and the whole particle swarm:

$$V_j^{k+1} = \omega V_j^k + c_1 r_1 \left(P_j^k - X_j^k\right) + c_2 r_2 \left(P_g^k - X_j^k\right) \tag{10}$$

where r_1, r_2 are random number between (0, 1).

3.1.5 Updating Particles

The position of particles is updated by adding a change velocity:

$$X_j^{k+1} = X_j^k + V_j^{k+1} \tag{11}$$

3.1.6 Inertia Weight

ω is time-varying inertia weight, as formula (12):

$$\omega_k = \omega_{\max} - \frac{\omega_{\max} - \omega_{\min}}{kmax} \times k \tag{12}$$

where k is iterative times, $k = 1, \cdots, kmax$. Time-varying inertia weight makes particles to have the good global search ability in initial searching period, and a good local search ability in inertia later period.

3.1.7 Learning Factor

c_1 and c_2 are time-varying learning factors, shown in formulas (13) and (14)

$$c_1 = (c_{1f} - c_{1g})\frac{k}{kmax} + c_{1g} \tag{13}$$

$$c_2 = (c_{2f} - c_{2g})\frac{k}{kmax} + c_{2g} \tag{14}$$

Asynchronous time-varying searching factor makes particles possess more self-learning ability and less social-learning ability, so particles prefer moving in whole searching space rather than moving fast to optimal solution. In late optimization, particles possess more social-learning ability and less self-learning ability, so particle tend to move to optimal solution quickly.

3.1.8 Termination Criterion

The procedure would save ever best solution and stop when iteration frequency reaches *kmax*.

3.2 Algorithm Process

Step1: set parameters in algorithm, including $\omega_{\max}$, $\omega_{\min}$, c_{1f}, c_{1g}, c_{2f}, c_{2g} and $k$max.

Step2: set $k=1$ and generate initial particle swarm $S=\{X_1, X_2, \cdots, X_m\}$ randomly, set E_i^*, X^* and E^* as particle best solution, global best solution and optimal target value respectively.

Step3: update particle velocity and location with formulas (10) and (11).

Step4: calculate evolution function value and renew best solution E_i^* of every particle. If $E_i > E^*$ and the constraints are satisfied, update global best solution X^* and optimal target value E^*.

Step5: if $k < k$max, $k = k+1$, otherwise, stop calculation and output global best solution X^* and optimal target value E^*.

4 Experimental Results and Analysis

4.1 Decision Variable

To simplify calculation, we consider four RMS indexes for decision variable x_i as follows. Other indexes are considered as constants.

T_{MTBF}: mean time between failures, $4 \le T_{MTBF} \le 200$.

T_{MTTR}: mean time to repair, $0.5 \le T_{MTTR} \le 5$.

K: spare sufficiency, $0.6 \le K \le 0.9$

T_{MTBCF}: mean time between critical failures, $30 \le T_{MTBCF} \le 250$.

4.2 Parameter Setting

4.2.1 Model Parameters Setting

Values of parameters in model are showed in Table 1.

4.2.2 Algorithm Parameters Setting

Inertia weight: $\omega_{\max}=0.9$, $\omega_{\min}=0.1$, learning factor: $c_{1f}=0.5$, $c_{1g}=2.5$, $c_{2f}=2.5$, $c_{2g}=0.5$, particles size: $m=300$, iterations: $kmax=200$.

Table 1 Model parameter setting

Parameter	Value	Parameter	Value
T_{MPT}	3	t_0	135
T_{MLDT}	9	C_f	20
T_{MRF}	1	S_a	50
F_z	800	T	35
F	200	β	0.5
C_0	290	C	0.9
LCC	1230		

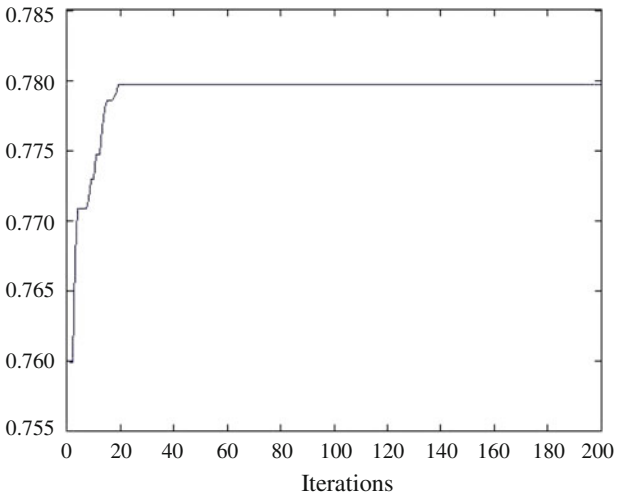


Fig. 1 Experimental result

4.3 Result and Analysis

Programming with Parameter Time-Varying PSO for 100 times in Matlab, the possibility to get best solution is 95%. Averagely, we get the best solution in 145th iteration and the result is quite accurate. Figure 1 shows a calculation procedure of getting best solution.

As shown in Fig. 1, algorithm converges in 146th iteration, RMS indexes is $X^* = (113.3, 5.0, 0.9, 250.0)$, optimal effectiveness is $E^* = 0.78$. Therefore, time-varying PSO is an effective method in solving RMS indexes optimization problem.

5 Conclusions

This paper establishes a system effectiveness model and life cycle cost model based on RMS indexes, and proposes a RMS index optimization model which is aimed at maximize armored equipment system effectiveness under a certain life cycle cost. Particle swarm optimization (PSO) with parameter time-varying is used to get the optimal RMS indexes. View from the model and algorithm structure, PSO with parameter time-varying is same efficient as other models such as system effectiveness optimization model based on RMS index description, which could also provide reference for armored equipment RMS demonstration and design.

References

1. Gan MZ, Kang JS (2005) Military equipment maintenance engineering. National Defence Industry Press, Beijing
2. Xiao HX, Wang JB (2008) Study on system effectiveness evaluation model of weapons and equipment based on RMS characteristics. Fire Control Command Control 33(5):130–132
3. Yu Y et al (2007) Efficiency evaluation model of weapons and equipments in use phase based on R&M&S. Armament Autom 26(3):6–7
4. Li JZ, Wang NC (2010) RMS research based on three dmiension model of support system with simulation method for military aircraft. J Beijing Univ Aeronaut Astronaut 36(12):1485–1489
5. Ping DF, Liu ZY (2013) Effectiveness evaluation of airborne ECM system based on the ADC model. Aerosp Electron Warf 1:34–37
6. Peng WH, Wang MH (2007) Study of influence on naval gun system efficiency by RMS. Ship Electron Eng 27(4):192–194
7. Han K, He CM (2014) Trade-off analysis of reliability/maintainability/supportability/testability of armored vehicle based on system effectiveness. Acta Armamentarii 35(2):268–272
8. Du HD, Wu W (2013) Simulation research for RMS indexes of armored equipment based on system effectiveness analysis. J Syst Simul 25(8):1947–1950
9. Geng JB, Jin JS, Zhang J (2010) Tradeoff analysis for RMS of equipment based operation availability and life cycle cost. Int J Plant Eng Manag 15(1):18–21
10. Ebeling CE (2010) Reliability and maintainability engineering. Tsinghua University Press, Beijing
11. Qiao SB et al (2012) Warship's RMS based on LCC of balance technology analysis. Equip Manuf Technol 1:90–93
12. EberHart RC, Kennedy J (1995) A new optimizer using particle swarm theory. In: Proceedings on 6th international symposium on micromachine and human science, pp 39–43

Dealing with Uncertainty in Risk Based Optimization

Ype Wijnia

Abstract Asset management is the coordinated activity to realize value from assets. This in most cases concerns optimization of costs, benefits, risks and opportunities. Theoretically, the optimum is the input that gives the best outcome, which should be precisely one value. In practice, however, optimizations encounter many uncertainties. Depending on the assumptions about these uncertainties the optimum may change. This may give the impression that optimization is not a very useful concept in case of significant uncertainty. However, this is not necessarily true. In this paper this is demonstrated for a typical decision problem in electricity distribution: the timing of a circuit upgrade in relation to the overload risk. Under the assumption that all relevant aspects of the risk can be captured into a single equivalent value, optimizations could reach more than 90% accuracy in total cost of ownership despite some 50% uncertainty in the underlying assumptions. This regards both the failure probability as the failure consequences. Apparently, in some cases optimization produces very good outcomes over a wide range of assumption, that is, the optimization is robust. Given that all optima are balance points, this may suggests that the robustness is a fundamental characteristic of optimization. Further research has to be conducted to test under which conditions that hypothesis holds.

Keywords Asset management • Risk management • Risk based optimization

1 Introduction

Asset management is, according to the international standard, the coordinated activity to realize value from assets. As mentioned in the standard, this in general involves "balancing of costs, benefits, risks and opportunities". Technically, this is not entirely correct, as in optimization it is about balancing the derivatives of those aspects and not the aspects itself (as pointed out by Woodhouse [1]). In practice though most asset managers will understand balancing as optimizing though.

Y. Wijnia (✉)
Asset Resolutions B.V., P.O. Box 30113, 8003CC Zwolle, The Netherlands
e-mail: Ype.wijnia@assetresolutions.nl

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_27

In theory, optimization is a fairly straightforward concept. One input is varied to see what value gives the best output. A typical example is the optimization of the maintenance interval for a specific equipment. Increasing the interval reduces the planned costs (both the expense for maintenance and the lost production) for preventive maintenance. However, the failure probability tends to increase over time due to wear and tear. This means that the likelihood of unplanned costs for the equipment and thus the expected value increases with increasing maintenance intervals. The optimal interval then is the value for which the sum of planned and unplanned costs is minimal [2].

Unfortunately, in practice optimization is not that simple. First of all, the cost functions may not be smooth over the range. If the production chain is periodically down for other reasons (like maintenance of other equipment or changes in the setup) the cost of maintenance precisely at that moment is significantly lower (only the expense, not the production loss) than when the maintenance is conducted at other intervals. But variations in the circumstances may influence the optimum as well. If the equipment is used under high load conditions, wear and tear will increase and thus the optimal interval will decrease. Furthermore, the optimization would be correct for continuous operation, but what if start/stop situations because of market demand and/or other equipment failures were included? Whereas the internal deviations can be included reasonably well in optimization, the uncertain environment is much harder to include. Optimization as a concept really is context dependent.

Does this mean that optimization as a concept does not hold if significant uncertainties are encountered? Fortunately, this is not necessarily the case. Some optimizations retain their value very well under a wide range of circumstances. They are so called robust optima. In this paper we will explore the robustness of optimization under a range of circumstances. We will start with the concept of risk based optimization. The next step is the analysis of risk based optimization with an uncertain failure rate. This will be followed by an analysis of risk based optimization if the impact is uncertain. The paper ends with a reflection on the robustness of optimization and recommendations for further research.

2 The Concept of Risk Based Optimization

2.1 The Concept of Risk

According to Aven [3], there is no generally agreed upon definition of the concept of risk. In everyday language risk may be used to indicate threat, probability, consequence or even the object at risk [4], though in literature many more definitions can be found [3]. If the many definitions of the concept have to be structured, the most differentiating aspect is whether risk is used to indicate the thing that could result in unexpected outcomes (risk as an entity) or that risk is the measure for the

amount of misery to be expected (risk = probability*effect). More modern definitions of risk tend to include the concept of uncertainty, as probability (in the meaning of relative frequency) is not very appropriate for discussing future unique events. The ISO standard on risk management [5] (which is used in the ISO55k serie [6–8]) defines risk for example as the effect of uncertainty on objectives. However, the ISO definition also attracted criticisms like not being precise [9, 10], illustration again of the lack of agreement. Uncertainty can result both in better than expected as in worse than expected. Whereas in some professions risk is used both for positive as for negative (especially finance) [11], in general risk is understood as the potential for a negative outcome. Risk is this paper will be used in this negative annotation.

Negative, however, is a value judgment, which makes the concept of risk inherently subjective. But even so called objective risk assessments can have hidden subjectivities. In modeling risk, experts make judgments that are by default value laden [4]. Even the choice of the metric can influence the outcome. Besides, people do not always judge risk in line with that objective assessment.

2.2 Risk Decision Making

To align risk management approaches with those irregularities in the concept, Klinke and Renn [11] developed a risk escalator: the more complex, uncertain and ambiguous a risk is, the more deliberation is needed in dealing with it. Risks that do not score on any of those criteria can be regarded as normal risks, to be dealt with in the form of a cost benefit analysis by experts. Interestingly, the role of deliberation for the non-normal risks is to bring the risks back to the normal area so that Cost Benefit Analysis (CBA) can be applied. Fortunately, most technical risks can be regarded as normal risks. They are well understood, often have a reasonable record of failures (making probability a meaningful concept) and they do not score high on ambiguity (technical failures are regarded as a bad thing). Furthermore, most people trust the experts in dealing with technical failures.

The basis method of CBA for risks is the concept of net present value. Measures should only be taken if the cost (in present value) of the intervention is less than benefit (also in present value) of the mitigated risk. If a part of the risk impact is measured in non-financial terms like safety and reliability, these aspects have to be translated into equivalent monetary impacts. A practical method for doing so is the risk matrix, which allows impacts on different values to be aligned on their severity [12], though it has to be recognized that risk matrices are also criticized [13]. Alignment in this case means that the decision maker is indifferent between impacts in the same severity class but on different values. Replacing the non-financial impact with the equally severe financial impact then gives the monetary equivalent. Figure 1 holds such an aligned risk matrix, with the three basis values of asset management. In this example matrix, 1 Customer Minute Lost (CML, a minute outage for one client) equals 0.50 €, whereas the equivalent value of a human live is

Potential / Severity class	Finance	Safety	Reliability	Likelihood: Unlikely <0,003	Remote <0,03	Probable <0,3	Annually <3	Monthly <30	Weekly >=30
Extreme	> 10 M€	Several fatalities	> 20 M cml	M	H	VH	U	U	U
Serious	1-10 M€	Single fatality or disability	2-20 M cml	L	M	H	VH	U	U
Considerable	100k-1M €	Serious injuries and significant	200k-2M cml	N	L	M	H	VH	U
Moderate	10k-100k €	Lost time incidents	20-200k cml	N	N	L	M	H	VH
Small	1k-10k€	Near misses, first aid	2-20k€ cml	N	N	N	L	M	H
Negligible	<1k€	Unsafe situations	<2k cml	N	N	N	N	L	M

Fig. 1 Risk matrix for energy distribution, derived from NTA8120 [14]. Risk levels are indicated by Negligible, Low, Medium, High, Very High, Unacceptable

between 1 and 10 million (if geometric mean is used about 3M). This allows virtually any asset management decision on risk to be monetized.

2.3 Risk Based Optimization

Even though the cost benefit analysis of a single intervention can be regarded as an optimization (it is still about the best option), normally optimization requires multiple alternatives to be evaluated. This can be a choice between technologies (like several types of switchgear), the optimal size within a single technology (like the cable diameter for expansions) or just a timing option for a single decision. Most decisions on existing assets fall into the timing category. The typical example (as referred to in the introduction) is the optimization of the maintenance interval, but upgrades, expansions and replacements also have an element of timing.

3 Case Introduction

In the electricity distribution grid cables make up for more than 50% of the total asset value. Decisions on the replacement and upgrades of those cables thus are of vital importance. Figure 2 gives a typical situation for decision making. A medium voltage substation is connected to a high voltage substation by means of two cables.

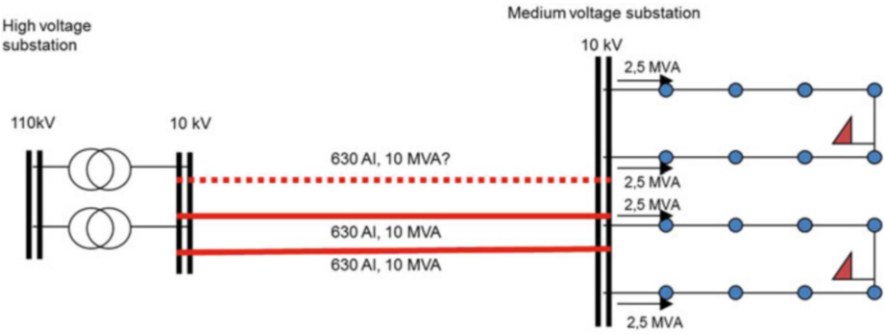


Fig. 2 Typical decision problem for circuit upgrade after [15]

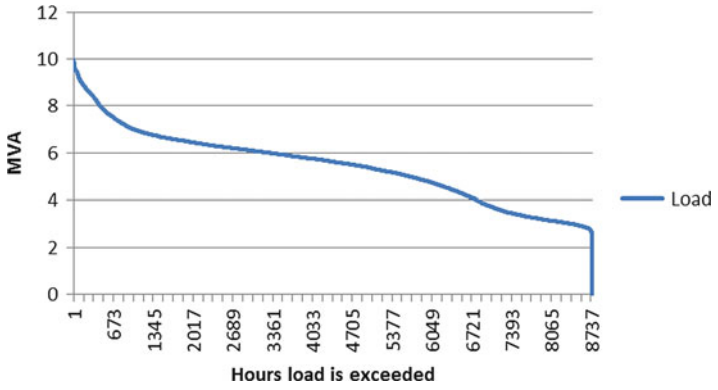


Fig. 3 Typical load duration curve of electricity distribution

Peak loads in the distribution grid tend to grow continuously, with some 3% per year. The peak load of the MV substation above has reached the redundant power (i.e. the power that can be provided with one of the cables out of operation). This means that in case of a single cable failure the other cable could be in an overload situation, potentially resulting in an overload induced failure. What is the right moment to upgrade the circuit with a third cable?

Classically, this decision would be resolved in by comparing the potential peak load with the nominal rating of the cable. This would result in upgrading right now. Two considerations result in postponement of the decision. First of all, the peak load occurs only 1 h/year. Figure 3 shows a typical load duration curve.

Only if the first cable would fail precisely during this peak load, an overload would result. But that probability is currently 1 in 8760, and will only slowly increase over time. The use of peak demand is therefore risk averse. Unfortunately, the load pattern of cables is normally not measured, only the peak load.

The second consideration is that peak load is compared with the nominal rating, which is the load the cable can endure for 30 years continuously. At least part of the

degrading mechanism is related to cable temperature. As cables have a large thermal time constant (i.e. they require time to heat up), it is clear that 1 h of overload does not result in a temperature induced failure, especially given that the cable is normally operated at less than half the nominal power. The use of the nominal rating thus also is risk averse. Changing to dynamic rating (which accounts for cooling of the cables during the nightly low load situation) could increase the safe capacity of the cable by as much as 20% [16].

However, there is also a factor that could limit the postponement of the investment. The energy loss in cables is quite significant, in net present value of the same order of magnitude as the total construction costs of a cable [17]. Upgrading the circuit with a third cable would reduce the energy losses by about one third, paying at least for part of the investment. This means the optimal timing of the investment depends on the load pattern, load growth, actual capacity of the cables and the future value of the losses, which all are uncertain. Furthermore, part of the investment has to be paid by the increased reliability, which is inherently subjective and uncertain. The decision thus has to be made under uncertainty. In this paper this will be decomposed in two fundamental uncertainties: one about the probability of failure and one about the precise value of the consequences. For each of those dimensions an analysis will be made whether a robust optimization is possible.

4 Dealing with Uncertainty in Failure Probability

The classical comparison of peak load with the nominal rating of the cable can be regarded as a risk free approach. No matter at what moment the failure occurs, the remaining cable will be able to supply the full load. However, this is not the maximum load level to be risk free. Even at the dynamic power rating, there will be no risk of thermally induced failures, simply because the cable temperature will not rise above that of the nominal rating due to the load cycle. But beyond this dynamic rating there is a risk of thermal overload.

Unfortunately, the mechanism of thermally induced overload is not well understood. On the one side, there is the understanding that chemical degradation is strongly related to temperature, but that a few hours of overload will not eat away too much lifetime. Cables may run at a double load (twice the nominal rating) for several hours without considerable life expectancy effects. If this is the case (the best case approach, with no overload induced risk), the decision should only be based on the energy losses. The left part of Fig. 4 compares the cost of the energy losses with the annual equivalent investment costs. According to this diagram, the investment should be postponed until (roughly) the full capacity was used in an everyday situation, but beyond that the losses would pay for a new cable.

On the other hand, it is also known that every cable has hotspots (points which heat up more than the normal cable), like the joints. Furthermore, the resistance of metals is generally positively correlated with the temperature, meaning that hot cables heat up faster. Therefore the overload risk may possess runaway

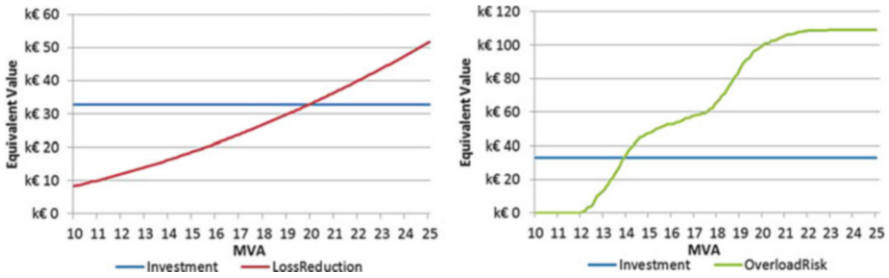


Fig. 4 Optimization of upgrade based on upgrade costs and energy loss (*left*) or upgrade costs and overload risk (*right*)

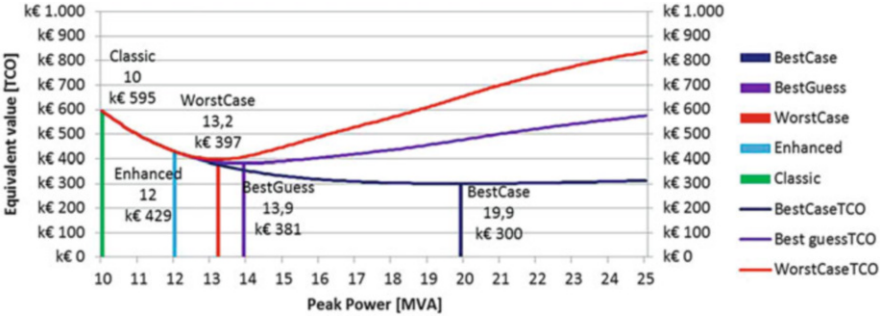


Fig. 5 Comparing optimizations based on different assumptions

characteristics, with a trouble free operation and then sudden failures. The worst case approach is that these failures occur at even 1 h above the safe limit. The right part of Fig. 4 shows the overload risk compared to the investment costs. According to this diagram (comparing risk and annual equivalent costs of the investment) some risk can be taken, but not much. At about 10% above the safe dynamic limit of 12 MVA the new cable pays out.

The difference between these extreme approaches is significant, about 6 MVA in load level. Given the growth rate of 3% per year, this is almost 20 years difference. Figure 5 holds all optimizations discussed so far. Additionally a best guess (average between worst and best case) has been added.

Especially given the large differences in Total Cost of Ownership (TCO), it seems very difficult to make the right decision. However, this changes if the optimizations are reviewed in terms of relative performance (i.e. the cost relative to the optimal value in the scenario). This is shown in Fig. 6. In the enclosed optimum (where the relative curves of best and worst case intersect) the performance is equally bad with regard to both worst and best case, thought this choice is

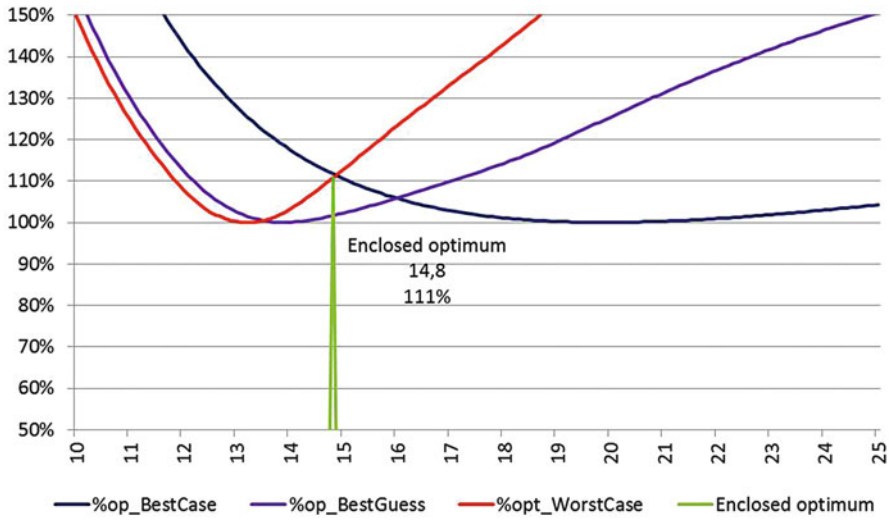


Fig. 6 Relative comparison of the various optimizations

in either case only 10% more expensive. As reality is somewhere in between the extremes, the premium towards reality is less. For example, the premium (or regret) towards the best guess is only 2%. This means that even though the precise failure mechanism is not fully understood and large uncertainties in failure probabilities exist, it is possible to make a very good and robust decision.

5 Dealing with Uncertainty in Failure Consequences

The other major uncertainty in the decision is the actual value of the failure. Most obvious is the value of the interruption to the customers. The direct value of an interruption is estimated at some 0.50 € per CML, though that value can rise significantly if the customers regard the network operator to blame. Another uncertain factor is the precise duration of the repair. The first failed cable can be expected to be repaired in 8 h, but if the second cable then fails due to overload, the first cable cannot be used to restore power completely. Additional measures like emergency power supplies are needed. It is probably more like 24 h before every customer is back on. The third uncertain factor is the damage to the overloaded cable. Thermally induced failures cannot be repaired, as the insulation will be degraded too much. Therefore, at least parts of the cables will have to be replaced, with a significant probability that the whole cable needs replacement. Total expected cost estimate would be in the neighborhood of the cost of replacing the cable, but with significant uncertainty.

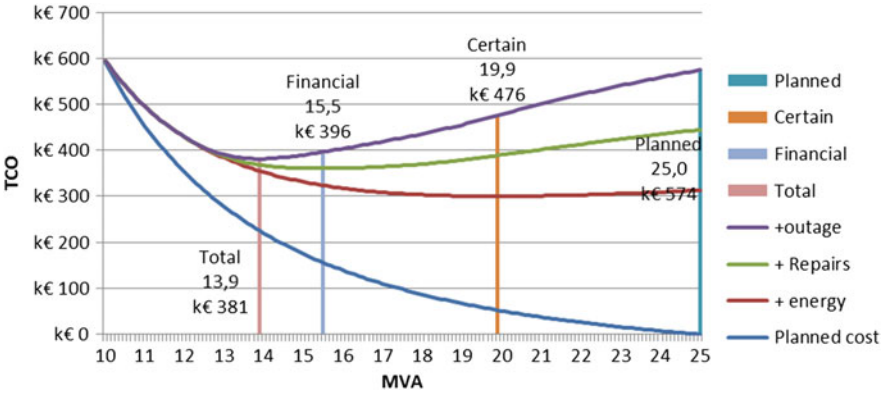


Fig. 7 Impact of partially capturing the value at risk on the optimum

Interestingly, optimizations do not depend that much on the precise value of the consequences. This is demonstrated in Fig. 7. Even if only the relatively certain costs (upgrade costs plus energy loss) are regarded in the best guess assumptions, it results in an outcome some 25% more expensive that the true optimum. With half of the risk captured (either repair costs or outage, which are about equal in this example) the difference is less than 5%. Optimization based on 50% of the risk then is more than 95% accurate. For normal practice that would be good enough.

6 Conclusions and Discussion

Even though optimizations in practice may encounter many uncertainties, it can be possible to find optima which are robust under extremes in the assumptions on the uncertain variables. The case used gave about 90% accuracy for both extreme scenarios with regard to failure probability as with regard to the value of failure. This is no coincidence. Optima are by nature balance points, in which the increase of benefits is exactly equal to the increase of costs. Both factors develop at a different rate, but around the optimum their values are very comparable over a significant range. In case of the electricity grid with its extremely long lived assets this may account for 10s of years. The case demonstrated a robust outcome of optimization despite large uncertainties. However, it is just one case. For an asset manager it would be very valuable to understand how accurate a decision would need to be to produce good outcomes for a variety of cases. This could help define in advance whether it would pay out to search for better data or that the decision given uncertainties is that good that this effort simply cannot pay out.

References

1. Woodhouse J (2014) Asset management is growing up. Tutorial at the 9th WCEAM, Pretoria
2. Wijnia YC (2015) Towards quantification of optimality in asset management (to be publsihed). WCEAM2015. Tampere, Finland
3. Aven T (2012) The risk concept—historical and recent development trends. Reliab Eng Syst Saf 99:33–44
4. Slovic P, Weber EU (2002) Perceptions of risk posed by extreme events. In: Risk management strategies in an uncertain world, New York
5. ISO (2009) ISO 31000: risk management- principles and guidelines
6. ISO (2014a) ISO 55000 asset management-overview, principles and terminology, Geneva
7. ISO (2014b) ISO 55001 asset management-management systems-requirements, Geneva
8. ISO (2014c) ISO 55002. Asset management-management systems-guidelines for the application of ISO 55001, Geneva
9. Aven T (2011) On the new ISO guide on risk management terminology. Reliab Eng Syst Saf 96:719–726
10. Leitch M (2010) ISO 31000:2009—the new international standard on risk management. Risk Anal 30:887–892
11. Klinke A, Renn O (2002) A new approach to risk evaluation and management: risk based, precaution based and discourse based strategies. Risk Anal 22:1071–1094
12. Wijnia Y (2012) Asset risk management: issues in the design and use of the risk matrix. In: Mathew J, Ma L, Tan A, Weijnen M, Lee J (eds) Engineering asset management and infrastructure sustainability. Springer, London
13. Cox LA Jr (2008) What's wrong with risk matrices? Risk Anal 28:16
14. NEN (2009) NTA 8120:2009 asset management—Eisen aan een veiligheids-, kwaliteits- en capaciteitsmanagementsysteem voor het elektriciteits- en gasnetbeheer
15. Wijnia YC, Herder PM (2005) Options for real options: dealing with uncertainty in investment decisions for electricity networks. International conference on systems, man and cybernetics, Hawaii
16. IEC (1985) IEC 60853: calculation of the cyclic and emergency current rating of cables. Part 1: cyclic rating factors for cables up to and including 18/30 (36) kV
17. Wijnia YC, Peters JCFM (2008) Integrating sustainability into risk based asset management. International conference on infrastructure systems: building networks for a brighter future, Rotterdam

The Design and Realization of Virtual Maintenance and Training System of Certain Type of Tank

Longyang Xu, Shaohua Wang, Yong Li, and Lijun Ma

Abstract Starting from the practical equipment maintenance and training, the designing thought, scheme and key technology of the virtual maintenance and training system of certain type of tank are elaborated in this paper. This article also introduced the method to realize the virtual maintenance and training system. Furthermore, an expectation on the developmental direction of virtual maintenance and training is also given in the paper.

Keywords Virtual Maintenance • Training • Tank

1 Introduction

Along with the constant deepening of the preparation of military struggle, new types of armament have been equipped to troops in succession. The new equipment has the prominent features of small number, complex structure, high scientific and technological content, and expensive materials of new types of equipment. The traditional real equipment maintenance and training are highly priced and low efficiency. In addition, tank components and parts are often damaged because of man-made mistakes and the high cost-benefit ratio. All these problems are particularly standing out; therefore, it is difficult to meet the requirements for new equipment maintenance and training. Fortunately, with the development of computer science and technology, it is mature to adopt the thought of virtualization and the technology of simulation to construct a training system and platform as an important auxiliary means for new equipment maintenance, support and training [1, 2].

L. Xu (✉)
Department of Technical Support Engineering of Academy of Armored Force Engineering, Beijing 100072, P. R. China
e-mail: aafe77330@163.com

S. Wang • Y. Li • L. Ma
Department of Basic Courses of Armored Force Engineering, Beijing 100072, P. R. China
e-mail: 13401060975@139.com

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_28

2 System Design

2.1 Design Principles

1. Visualized forms of representation. The virtual training system should be able to fully reflect the structural features of certain type of tank, enable the train-ees to understand the structure and assemble relationship of each component and part explicitly.
2. Real maintenance operation. The maintenance training procedure, action and technical requirements in the virtual environment should be completed in consistence with those on real equipment, and the virtual training should be able to replace training on real equipment basically.
3. Complete system functions. The system can not only complete the technical training for maintenance, but also has the functions for testing and grade evaluation, compensating for the shortcomings of being unable to conduct accurate testing (examination) step by step in real equipment training.
4. Convenient operation. The system installation and operation should be visualized, simple and convenient; furthermore the system should have good compatibility so as to be generalized and popularized in use.

2.2 Overall Design Scheme

Due to the particularity of maintaining activities, it makes great difference to conduct virtualized simulation for the maintenance and to carry out general scene virtualization and environmental simulation. The existing constraint conditions for supportive resources and technical procedures make the steps and procedures of the dynamic simulation very complex. Therefore, the main aims of this paper are to deposit the modules of tank components and tools in the form of module bank, to introduce virtual interactive engine for maintaining actions, to call the necessary parts, tools and materials according to the equipment maintaining technology; and to construct the virtual maintenance and training system for certain type of tank with the help of the desktop display. After all models are constructed, the system has three modes: demonstration, training and testing which can not only describe the equipment maintaining process visually but also provide man-machine interactive maintaining (such as tools, material selections, and selections of recovering repairing methods) etc. End user can freely select or switch over one of the three modes and control the training process, enabling relevant personnel in service to conduct training under different training stages. System managerial personnel can control the users' logging-in, conduct inquiry, statistics and analysis over the process and effect of the training. In addition, the system has reserved interface for secondary development and upgrading (Fig. 1).

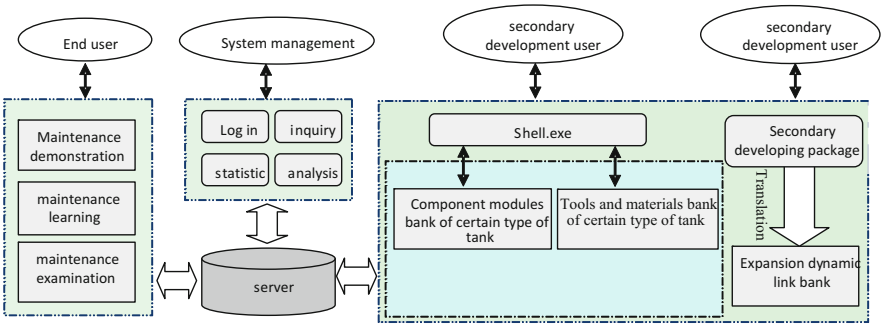


Fig. 1 Virtual maintenance and training system for certain type of tank

2.3 Development and Running Environment

The virtual maintenance and training system for certain type of tank can be run on single computer and on the LAN, using SolidWorks as cartographic software to establish 3-D models for tank parts and components, maintaining tools and materials. Using Cortona3D Rapid Learning platform as interactive engine for maintenance actions, and using HTML files for independent release.

3 Key Technology

3.1 Equations Three Dimensional Interactive Modelling for Equipment Maintenance

Looking at the past modelling experience and process, we find that the most virtual technologies adopted the software like 3D MAX, MAYA. The type of modelling system is based on polygonal modelling system, for example, adopting Loft, Edit, Poly, Mesh, and HSDS. Comparatively, it is appropriate to consider this type of modelling method under static art modelling is presented, or the required accuracy is not high.

However, due to the high requirements of the arm or equipment maintenance and training system for the size, accuracy, complexity, and consistency with real object structure, it is very hard to achieve the desired effect if the polygon-based sys-tem is adopted to construct primitive 3-D modelling data. Therefore, the comprehensively selected 3-D digital modelling methods are considered which is described as follows:

1. Adopting entity modelling software such as SolidWorks, ProE, etc, to design the primitive digital models of the equipment assemblies, parts and components.

2. Lightening the weight of the primitive digital models and save them as poly-gon-based 3-D digital models (*.wrl) that can be directly used by the system.
3. Taking the files of *.wrl format as standard 3-D digital model guiding format for the virtual maintenance and training system.

3.2 Transformation of Visual Scene Coordinates

The 3-D visual scene control of the virtual maintenance and training system involves hidden-surface elimination, illumination model and surface algorithm, cut-ting algorithm, topographic data display and topography matching optimization. This article highlights the exploration of the coordinate transformation within three- dimensional space. Coordinate transformation can be divided into subject position transformation, and coordinate transformation from one system to another system.

As for the subject position transformation within a coordinate, it involves three kinds of transformations such as translation, zoom and rotation. Sometimes it also includes reflection and transvection transformation. Taking a translation as ex-ample, in the 3-D homogeneous coordinates, the arbitrary point P (x, y, z) can be transformed into point P(x′, y′, z′) using the transformation formula Eqs. (1) and (2):

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \bullet \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{1}$$

$$P' = T \bullet P \tag{2}$$

Or, in the 3-D space the translation of an object can be realized through the object to be translated. As for the object represented by a group of polygons, the vertexes of each surface can be translated and then the transformed position can be expressed as.

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} S_x & 0 & 0 & 0 \\ 0 & S_y & 0 & 0 \\ 0 & 0 & S_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \bullet \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{3}$$

$$P' = S \bullet P \tag{4}$$

Limited by the length of this paper, Zoom, Rotation, Reflection and Transvection are not discussed in the paper.

3.3 Maintenance Procedures and Constraint Model

The premise and basis for improving the training quality with virtual maintenance and training system is to establish a set of scientific and reasonable maintenance procedures. In the course of virtual maintenance training, the subjects (equipment parts and components, tools and materials) in the training scene are constantly changed, so it is needed to establish a special model to describe the maintenance procedures and their constraints. This article adopts Finite State Ma-chine (FSM) to describe the maintenance procedures and uses starting condition controlling functions to realize the procedure constraints.

$$S_i = \begin{cases} 1 & \textit{condition that meets the starting of procedure } i \\ 0 & \textit{condition that doesn't meets the starting of procedure } i \end{cases} \tag{5}$$

$$F_i = \begin{cases} 1 & \textit{condition that procedure } i \textit{ is done} \\ 0 & \textit{condition that procedure } i \textit{ isn't done yet} \end{cases} \tag{6}$$

For every working procedure, a corresponding starting condition controlling function is established:

$$F_i = \phi(\mathrm{S}_1, S_2, \ldots, \mathrm{S}_\mathrm{n}) \tag{7}$$

When a user's operation comes to procedure i, the controlling function set is scanned and a corresponding controlling function is obtained to determine instantly whether its starting condition is met.

4 System Realizations

According to the functions required by the virtual maintenance and training system for certain type of tank, the virtual maintenance and training system framework is composed of visual scene controlling platform, data bank system, virtual maintenance inactive engine and system management platform. The visual scene controlling platform mainly realizes setting simulation; the data bank system is used to store equipment parts and components, tools and materials; the virtual maintenance inactive engine is used to realize man-machine interaction for the maintenance actions; the system management platform is used to control the system running for demonstration, learning and examination.

The realization process of system involves SolidWorks 3-d modelling; using Cortona3D interactive engine to disintegrate, set up and edit the maintenance procedures and interactions; using Rapid Learning platform to edit the functions like demonstration, learning, and examination; using Virtual Training Viewer to make independent release with HTML files (see Figs. 2, 3, and 4).

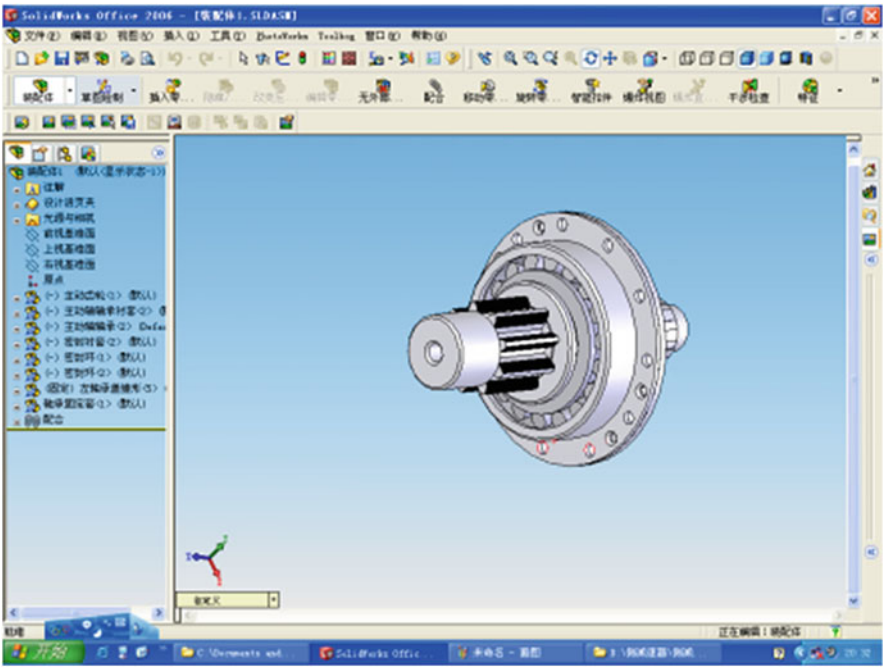


Fig. 2 Equipment parts 3-D modelling

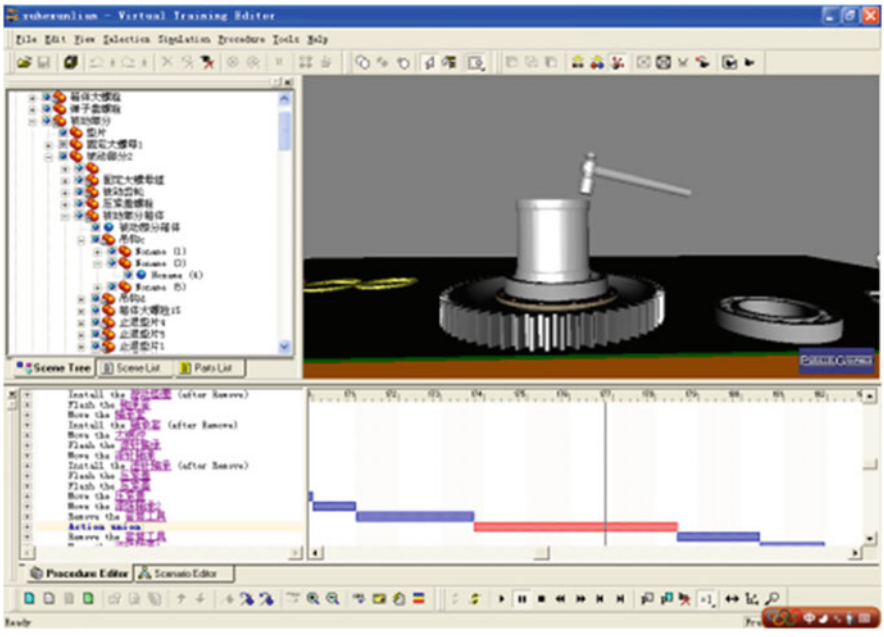


Fig. 3 Maintenance procedure setting up

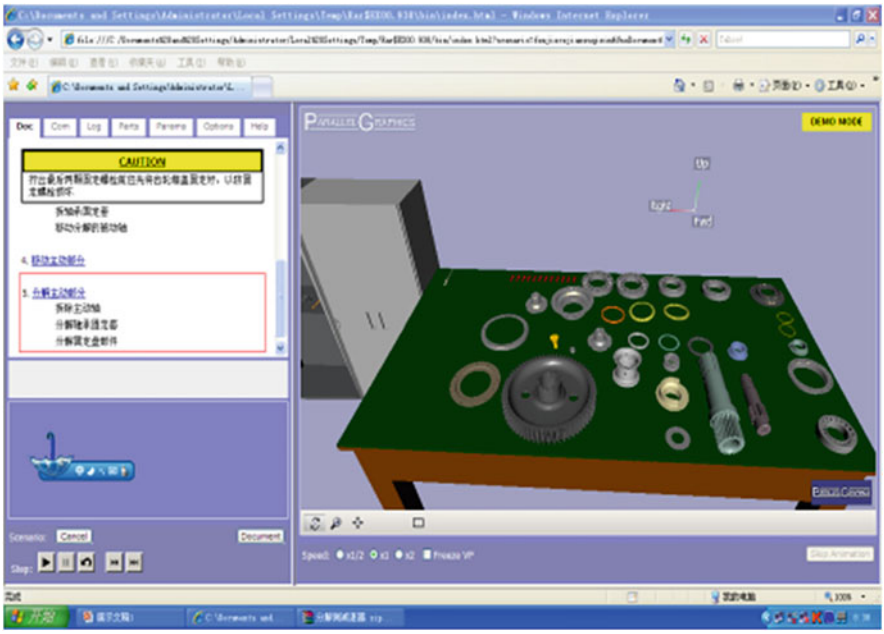


Fig. 4 HTML file release

5 Conclusions

As a brand new training mode, the virtual maintenance and training system have aroused the attention of many experts and scholars in the field of equipment support. The virtual maintenance and training system can provide advanced experimental environment and simulating means for the arm or equipment maintenance and training. It can also greatly facilitate the training effect, promote talent cultivation quality and improve the level of scientific re-search. The virtual maintenance and training system has become a main developmental trend of the maintenance and training technologies, hence it should be given extensive attention on it.

References

1. Xing HG, Wang GH, Zhou SH (2014) Virtual maintenance system for certain type of tank gun. J Sichuan Ordnance 35(7):1–4
2. Zhao C, Li XX, Xia K et al (2014) Control method of concurrent operation for collaborative virtual maintenance. Comput Syst Appl 23(10):198–201

Measures to Ensure the Quality of Space Mechanisms

Jian-Zhong Yang, Jian-Feng Man, Qiong Wu, and Wang Zhu

Abstract Space mechanisms (SMs) refer to the devices attached to the spacecraft to perform defined functions by mechanical movements. It is very important to assure the quality of SM to meet the requirements of spacecraft, because most quality defects of SMs can result in severe performance degradation or failure of the spacecraft. Based on 20 years' experience of SMs research and development, a set of critical measures that have been found to be extremely important in quality assurance in the most demanding environments are identified. The measures can be adopted in different phases of the development process, such as design phase, manufacture phase, ground validation phase, storage and transportation phase, final assembly and integration phase, etc. It has been proved that these measures can ensure the quality of SM effectively, and can also contribute to the research and development project for new SM.

Keywords Reliability • Performance • Spacecraft mechanism • Quality control • Product assurance

J.-Z. Yang (✉)
Beijing Institute of Spacecraft System Engineering, Beijing, PR China

College of Aerospace Engineering, Nanjing University of Aeronautics and Astronautics, Jiangsu, PR China
e-mail: jzyang1234@sina.com

J.-F. Man
Beijing Institute of Spacecraft System Engineering, Beijing, PR China

School of Mechanical Engineering and Automation, Beijing University of Aeronautics and Astronautics, Beijing, PR China
e-mail: manjf@163.com

Q. Wu • W. Zhu
Beijing Institute of Spacecraft System Engineering, Beijing, PR China
e-mail: wing21@126.com; juwong@china.com.cn

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_29

1 Introduction

Space mechanisms (SMs) refer to the devices which are attached to the spacecraft to perform defined function by mechanical movements [1, 2]. Their main functions contain the hold-down and release of two components of a spacecraft or a component and a spacecraft, the docking and separation of two spacecrafts, the deployment and locking or shape preserving of component, landing shock alleviation of spacecraft or astronaut, on-orbit vibration attenuation of high precision payload, tracking and pointing of the payload, orientation adjusting of whole spacecraft, etc. The spacecraft is usually equipped with many kinds of SMs with different functions. It's very important to assure the quality of SMs, because most quality defects of SMs can result in a fatal performance degradation or failure of the spacecraft [3, 4]. SMs will generally suffer severe environment conditions of launch and on-orbit, which can result in unignorable faults. Besides, it's extremely difficult, risky and costly to repair or maintain SMs on-orbit. The high reliability of SM is therefore quite important. For example, in order to provide steadily energy for the communication satellite during the whole life cycle, the driving mechanism which drives the solar wing pointing to the sun is ordinarily required to work normally at least 15 years in the space [5–7].

To insure SMs work reliably on-orbit, some researchers have discussed the measures of design [8–10]. However, measures of only design stage cannot assure the quality of SM. To insure SMs' reliability, measures of the whole development process should be considered. It has been proved in engineering that appropriate control measures in all of SM development phases of design, manufacture, ground validation, storage and transportation, final assembly and integration should be taken into consideration to assure the SMs' reliable work on-obit.

2 Design Measures

2.1 Design Process Management

2.1.1 Multi-phase Design

Similar to the development of the spacecraft, the development process of SM is also composed of several phases, such as concept phase, prototype phase and flight-type phase [1]. With the upgrade of the development phases, the control measures of the design modification will be stricter. When the design scheme of the product is finalized, none of the modification will be allowed. However, if the modification is highly necessary, the changed product is no longer regarded as the same product as the original one. Dividing the design process into several phases can meet the demand of frequent modification at the earlier phase and can also ensure the stability and credibility of the product at the later phase. The multi-phase design strategy is quite essential for the rapid development process as well as the strict quality control requirement for SMs.

2.1.2 Multi-level Checks and Signatures

For important documents such as the precept design report need to be multi-level checked and signed before they become effective. Ordinarily, the multi-level checks contain the signatures of designer, corrector, reviewer and approver. Sometimes it needs the countersignature of the technicians of other departments. A different signer has different and definite responsibility. With the development of the signature level, the signer will often ask some specialists to work together to dealing with the problems and to developing ideal solutions. Only when the related solutions are fully discussed, the signer will put his signature. The adoption of multi-level check and signature integrates the collective wisdom into the scheme, which can insure the scheme's feasibility, scientificity, comprehensiveness and pertinence. With the strategy of multi-level checks and signatures, the limitation and randomness of the scheme can also be avoided effectively. The quality of SMs can be fundamentally assured by this measure.

2.1.3 Multi-level Reviews

The important precept designs can only take effect when they have passed multi-level reviews. In general, it contains the reviews of engineering group, local department, user, and so on. Different levels of reviews will invite different specialists to discuss problems and evaluate the solutions. Only when the lower-level review has been gotten through can the higher-level review be implemented. In each review, the replies to all of the experts' questions should be summarized. If the advice is accepted, the implementation measures should be pinpointed. However, if the advice is denied, the reasons for rejection should be clarified in detail. All the replies will be fed back to the relevant experts. Only when all the experts agree with the answers and give their signatures, can the review be passed. The implementation of this regulation involves the knowledge and skills of many experts into the scheme, so the feasibility and scientificity of the scheme will be further ensured. The quality of SMs is further assured too.

2.1.4 Manufacturability Countersign

Drawings of any SM should be countersigned by process engineers before SM is manufactured. Process engineers will judge the manufacturability of the scheme by evaluating the performance characteristics of the product, the supply of the materials, the processing equipment, the cost, the special process and its stability, etc. Therefore, the manufacturability and reliability are both ensured. The quality and performance of SM will also be further assured.

2.2 Common Design Measures

2.2.1 Ensure the Toughness of Strong Parts

In order to decrease the mass or the volume of the SM, the high strength materials are usually adopted, or the special heat treatments are usually taken to improve the load bearing capacity of the part. To deal with heavy loads, especially the tension load, the strength and the toughness need to be both considered at the same time. If the strength is over emphasized and the toughness is neglected, the brittle fracture can be induced. It's important that both the strength and the toughness of the high load bearing part are required explicitly. As a result, the load bearing capacity can be ensured and the brittle fracture can be avoided. In addition, stress concentration should be also eliminated on high load conditions.

2.2.2 Selection of Mating Part Material

For some SMs, the surrounding environment temperature on-orbit may change greatly. Deferent parts for one joint (e.g., friction pair) should be manufactured with the same material, in order to avoid the failure of joint movement caused by the different coefficients of expansion. However, once the temperature changes too fast or too much, the uneven expansion caused by the temperature gradient cannot be avoided by the above mean. In this case, additional thermal control measures should be taken. So the temperature gradient can be limited within a safe range, and the movement fault of the joint can be avoided.

2.2.3 Selection of Lubricating Methods

All the joints of SM should be lubricated. The appropriate lubricating methods can be selected by the joint's running speed, on-orbit life, load conditions, etc. When solid lubrication film is selected, the effect of the film thickness on the properties of joint movement should be considered thoroughly. The lubrication film can be applied to the surface of one part of the two mating parts, if the other part is difficult to lubricate. For example, when the slender shaft-hole fitting is coped with, lubrication film can be applied only to the surface of the slender shaft.

2.2.4 Selection of Driving Methods

The spring or pyrotechnic device is usually used instead of electric motor to drive the SM which only works once on-orbit. Because the segments subject to potential failure in spring and pyrotechnic devices are much fewer than electric motor, so they are more reliable than electric motor. And the quality and performance of the

SM with spring or pyrotechnic device can be easily ensured [11–13]. Compression spring is often adopted instead of tension spring or scroll spring, because the compression spring can provide elastic driving force even the fracture happens.

3 Manufacture Measures

3.1 Technology Management

The management of the technology scheme is the same as the control of design scheme. In the implement process of the technology scheme the three-level checks are always emphasized, which is check by oneself, check through each other and check by professional. Detailed records for every check and for the handover of different procedures are all made and preserved for a long time. These records will provide convenience for the traceability of the probable problems that appear in the implement process. If the out-of-tolerance condition happens, the quality control department will have a meeting with relevant people, such as designer, technologist and processing worker to discuss the reason and the solution of the problem. Two goals will be achieved through the technology management. The first one is to ensure that the performance of each part meets its requirements. And then the quality of the whole mechanism can be assured. The other one is that the quality of the whole mechanism can be ensured by taking some necessary measures, even if the performance of some parts doesn't meet the requirements. For example, if the diameter of a shaft is smaller than required, the diameter of the relevant mating hole can be smaller and then the mating performance can be kept the same.

3.2 Common Technology Measures

3.2.1 Avoiding the Hydrogen-Induced Brittle

There is always no sign before the hydrogen embrittlement happens. Thus its influence on SM is often fatal. For the materials with strength of more than 1400 MPa, the electroplating corrosion protection is forbidden in order to avoid the hydrogen-induced brittle of the parts. And anticorrosion coating of physical or electroless nickel deposits will be adopted instead. In addition, the acid cleaning for the part should be also avoided in the whole process to eliminate the probability of the part to contact with hydrogen.

The heat treatment of the titanium alloy should be implemented in a vacuum tank to avoid the alloy contact with hydrogen in the air at the high temperature. Hydrogen-induced brittle of the part which is made of titanium alloy can be avoided effectively by this method [2].

3.2.2 Cleaning and Check Before Assembly

A list of all parts and materials is always made before assembly, which contains standard parts, purchased parts, manufactured parts, auxiliary materials, equipment and tools, etc. And then they will be cleaned and checked one by one according to the request. The unqualified samples should be picked out and a record will be made in detail. The sealing rings and sealing surface should be checked more carefully with magnifier to ensure that there is no scratch reducing the seal performance.

3.2.3 Usage Control of Adhesive

When adhesive is used, it should be selected correctively according to the environment conditions. The location and amount of adhesive should be controlled strictly to avoid the contamination to other parts. For example, if the adhesive flows into the joint, the movement performance of SM will be degraded or fail to work completely. By ensuring both the curing temperature and time meet the demands, the bonding quality of SM can be assured.

3.2.4 Anti-loose and Preload of the Thread Connection

Thread connection is one of the most common connection patterns for SM. Choosing the correct measures for anti-loose and controlling the preload of tightening is key point to ensure the reliability of the connection. To coat a little adhesive at the end of the male thread is often adopted to achieve the anti-loose. This method is both simple and reliable. For the connections which will be separated after connection, the adhesive may lead to damage of the thread during the disassembly process. Therefore, when the separated thread needs to be used again, it should be examined carefully.

The thread is usually tightened by the torque wrench according to the predefined torque. Tightening torque values can generally be found in related manuals.

3.2.5 Binding and Fixing of Cable

Electric power and control signal are necessary for the mechanism which will operate on-orbit for a long time. So the cable is usually an important component for SM. In order to avoid the damage of cable and plug in the launch period, it's necessary to bind and fix the cable properly. In order to eliminate the loosening of plug or the additional force acting on plug as well as the probable resulting resistance of the mechanism, the fixed point and diameter of bunched cable should be determined appropriately. In order to prevent the cable from getting hard due to

the low temperature environment, thermal control measures should also be taken to avoid addition resistance to the joint movement caused by the hard cable.

4 Ground Validation Measures

Comprehensiveness and safety of ground validation are both very important for quality assurance of SM. It contains the effectiveness and reasonability of validation scheme, the accuracy of measured parameters, pre-validation and the comprehensiveness of data analysis.

Effectiveness of validation scheme refers to effective simulation of on-orbit environmental conditions and validation of the mechanism performance under the conditions. The reasonability of validation scheme refers to low cost, mature equipment, low risk and easy implementation. Sometimes it's very difficult to simulate all the space conditions. In this situation, the influences of environmental condition parameters on the performance of the mechanism need to be fully analyzed, and only those with distinct effects will be simulated.

Accuracy of parameter measurement refers to that the method is reliability, the equipment is mature and all the critical measurement points or data have a backup. In order to measure the parameters accurately, the contractor will devise the implementation program in detail, and the program will be treated according to the design scheme. So the validation target will be achieved smoothly.

To ensure the safety of important validation, one or two pre-test will be conducted to valid the effectiveness of the program and the conditions of the equipment. Improvement measures will be taken according to the exposed problems. Thus the formal experiment can be conducted smoothly and effectively, and the safety of personnel and samples can also be assured.

The results of validation are often different from the expectation. The difference can always be explained by analyzing the measurement data. And the validity of the data and the success of the experiment can be judged correctly. In addition, some performances of SM can be determined only by the summary and analysis of the measurement data. For example, the reliability of SM can only be assessed based on the measurement data of the reliability test. The reliability of landing gear of Chang'E-3 Probe has been assessed by this way [11].

5 Control of Storage, Transportation and Final Assembly

Many parts of SM will be stored for a long time before the final assembly. In order to assure the stability of the quality, all the storage demands must be met including conditions of cleanliness, temperature and humidity, etc. For some special parts, such as slender rod and thin wall cylinder, need to be supported by special equipment, so that they will not be deformed by the action of the weight.

Transportation control contains the selection of package measures, transportation mode and the control of the overload in the transportation process. There are close relationships among them. The package measures should adapt to transportation mode, and the allowable overload and temperature should be controlled effectively. The measures of dustproof, damp proof and buffer should be adopted and the whole mechanism is supported and fixed reliably to prevent damage. For some special mechanisms, the variation of the air pressure in the transportation process should also be considered, because the variation can result in deformation of the mechanism.

When the mechanism is transported to the destination, it should be checked carefully. The check usually contains appearance inspections and basic performance tests. The key points of appearance inspections are to check whether the lubrication coating is peeling off or the surface is scratched, or the mechanism is deformed, etc. The key points of basic performance tests include the verifications of circuit conducting, resistance changes and switch on-off. For some special mechanisms, such as driving mechanism for solar wing deployment, the deployment property will be verified again after long-distance transportation.

Final assembly refers to attaching SM to the spacecraft, which is the last important step to ensure the quality. In this process, the video and photo will be both taken for the important operation. For large mechanisms, the effective measures should be taken to avoid the additional internal stress and deformation resulting from the weight.

6 Conclusions

It is important to ensure the on-orbit performance of SM by the quality control measures mentioned, which are summed up from long term practical experiences. These measures cover all of the steps of the research and development of mechanism from the scheme determination of design to technology control, ground validation and storage, transportation and final assembly, etc. The development of these measures is of great significance for the successful research and development of new SMs in the future.

References

1. Chen LM (2005) Spacecraft structures and mechanisms. China Science and Technology Press, Beijing
2. Yu DY, Yang JZ (2011) Spacecraft mechanism technology. China Science and Technology Press, Beijing
3. Cong Q (2012) The advance of integration technology for space mechanisms. Spacecr Environ Eng 4:384–387

4. Ma XR, Yu DY, Sun J (2006) The researching evolvement of spacecraft deployment and driving mechanism. J Astronaut 6:1123–1131
5. Chen J (2007) Development of foreign communication satellite technology. Aerosp China 2:38–46
6. Liu H (2012) Reliability analysis of foreign commercial communication satellite platform. Space Int 11:14–19
7. Shao RZ, Fan BY (1996) Reliability study of long-life communication satellites. Chin Space Sci Technol 4:24–33
8. Peng CR (2011) System design for spacecraft. China Science and Technology Press, Beijing
9. Yang JZ, Zeng FM, Man JF (2014) Design and verification of the landing impact attenuation system for chang'E-3 lander. Sci Sin Technol 5:440–449
10. Yuan JJ (2004) Design and analysis of satellite structure. Astronautic Publishing House, Beijing
11. Wu Q, Yang JZ, Fu HM (2014) Deployment reliability test and assessment for landing gear of Chang'E-3 probe. J Donghua Univ 6:782–784
12. Yang JZ (2015) Landing gear of spacecraft. Astronautic Publishing House, Beijing
13. Yang JZ, Wu Q, Man JF (2015) Reliability design and validation for the landing gear of Chang'E-3 lander. In: Proceedings of international conference on quality, reliability, risk, maintenance, and safety engineering (QR2MSE), Beijing, China

A Decision-Making Model of Condition-Based Maintenance About Functionally Significant Instrument

Xiang Zan, Shi-xin Zhang, Yang Zhang, Heng Gao, and Chao-shuai Han

Abstract Condition based maintenance (CBM) of functionally significant instruments (FSI) is critical for the reliable operation of the whole equipment. Based on the present moment and history monitoring information, a decision-making model of condition-based maintenance about FSI is established by WPHM (Weibull Proportional Hazards Model). Failure risk management is the decision-making target of the model. Through controlling the range of the failure risk, the result of the inspection interval of CBM can be acquired. Because of its favorable global search ability, genetic algorithm is used in parameter estimation to avoid the influence of the initial parameter value on the result of estimation. The decision-making result of engine shows that the model works well and effectively.

Keywords Condition based maintenance • Proportional hazard model • Inspection interval • Genetic algorithm

X. Zan (✉) • S.-x. Zhang
Department of Technical Support Engineering, Academy of Armored Force Engineering, Beijing 100072, China
e-mail: zan_xiang.student@sina.com; zsxzh@sohu.com

Y. Zhang
Military Deputy Office of PLA in 674 Factory, Harbin 150056, China
e-mail: 147267728@qq.com

H. Gao
Teach Room of Voluntary Artillery, Academy of Nanjing Artillery, Nanjing 102205, China
e-mail: 356165891@qq.com

C.-s. Han
Troop No. 63960 of PLA, Beijing 102205, China
e-mail: 929134680@qq.com

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_30

1 Introduction

With the development of high and new technology and fuzz of fault law and failure mode, limits for validity of time-based maintenance is reducing. CBM (Condition-based Maintenance) is valued gradually, which can insure reliability, improve operational availability, reduce cost of maintenance support.

Many decision-making models and methods of CBM are developed by scholar in home and abroad. Based on markov chain theory, markov decision-making model is established to describe law of condition changed. Levy process model can describe deterioration process on the condition of continuum damage. Delay-time model is used to divide change of condition into two stages. For example, Salvinder and cooperators established assessment model based markov chain to assess condition change of vehicle crank-link mechanism, and based the result of assessment the result decision-making can be acquired [1]. Sha and cooperators researched step-stress accelerated life testing through proportional hazards model, and combine proportional hazards model with Weibull model to establish decision-making model. At last, showing the model worked well and effectively through an example [2]. WANG Wen and cooperators analyzed the effect of drop height on the solder joints' life span through proportional hazards model [3].

2 Related Theory

2.1 Condition-Based Maintenance

Most faults occur in gradual process. The Process can show in Fig. 1. In Fig. 1, *O* shows fault-beginning point, which time fault begins to happen. *P* is Potential Failure point, which abnormal condition can be inspected. *F* is Functional Failure point, which fault finally happens. The time between *P* and *F* is *P-F* interval [4].

The fundamental principle of CBM is that if *P-F* interval is large enough, fault sign can be discovered to prevent fault from happening.

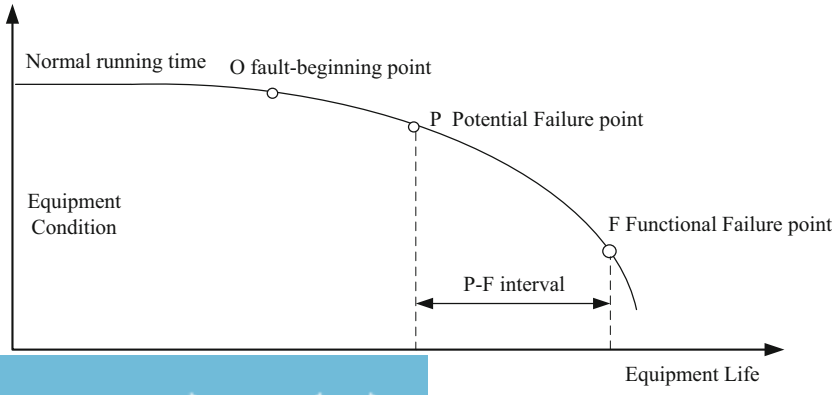


Fig. 1 P-F interval curvilinear

2.2 Concept About FSI

Failure consequence of FSI may be one of conditions below (GJB1378-92) [5].

1. May impact the service safety of equipment,
2. May impact the completion of tasks,
3. May lead to heavy economic losses,
4. Hidden function failure and other failure may together lead to one or some consequences above or quadratic effect may lead to the same consequences.

Based on related theory of CBM, FSI is the object of equipment CBM.

2.3 Weibull Proportional Hazards Model

PHM is present by Cox in 1972 [6]. Then Jandine and others put PHM in CBM, in which condition parameters, service load, fault and so on are looked as follow factors of equipment life. Product property theory can appear in real risk of equipment by PHM [7]. The property of PHM is the hazard of different individuals is proportion [8].

The form of PHM shows below.

$$\lambda(t, X) = \lambda_0(t)\exp(\beta X) \tag{1}$$

In the expression, $\lambda(t,X)$ is fault probability, $\lambda_0(t)$ is the only related to fault probability, X is condition number in t, β is regression variable efficiency, which is used for establishing the relation between condition number and fault probability.

Due to Weibull model can match most life distribution of mechanical products, WPHM (Weibull Proportional Hazards Model) is used for establishing reliability model of equipment.

Integration of PHM and two-parameter Weibull model, WPHM can be acquired, which can be used for describing the relation between equipment condition and fault probability. The expression can show below.

$$\lambda(t, X) = \frac{\delta}{\alpha}\left(\frac{t}{\alpha}\right)^{\delta-1}\exp(\beta X) \tag{2}$$

In the expression, α is shape factor, δ is locational parameter, X is dimensionality variable, which shows condition information, β is t dimensionality condition variable efficiency arrow. At the same time, some parameters can show below.

$$\begin{cases} \beta = (\beta_1, \beta_2, \ldots, \beta_t) \\ X = (X_1, X_2, \ldots, X_t)^{\mathrm{T}} \end{cases}$$

3 Decision-Making Model

3.1 Model Established

Newton-Raphson iterative algorithm was used for parameter estimation. However, disadvantage of the traditional method is the result to be inaccuracy, which is easy to get local optimized solution and influence by initial parameter. Genetic algorithm owns global convergence ability and can ignore the influence of initial parameter. So genetic algorithm is chosen.

The step of WPHM parameter estimation by genetic algorithm shows below [9].

1. Determination on encoding method

   Encoding method of chromosome is often decimal system and binary system. After determination encoding method, length and location of every parameter in chromosome also can be determinate.
2. Set on calculate parameter

   Calculate parameters are including population size, selection strategy, crossover strategy, mutation strategy, crossover probability, mutation probability and so on.
3. Determination on fitness function

   Fitness function is the only certainty indicator to estimate adaptability of chromosome [10]. The project of fitness function can show below.

   The distribution type of WPHM is known. Maximum likelihood function is used to structure fitness function of genetic algorithm. In assumption, there is n independent detected traffic data of equipment, joint probability density likelihood function can be acquired below.

$$L = \prod_{i=1}^{\mathrm{q}} \lambda(T_i) \prod_{i=1}^{n} R(T_i) \prod_{i=1}^{\mathrm{q}} \prod_{j=1}^{k} p\left(X_{j-1}^{i}, X_{j}^{i}\right) \tag{3}$$

In the expression, $R(T_i)$ shows reliability function of i, $\lambda(T_i)$ shows fault probability function of i. $p({X_{j-1}}^{i}, {X_j}^{i})$ shows probability from ${X_{j-1}}^{i}$ to ${X_j}^{i}$.

Based on Eq. (2), reliability function with condition parameters can be acquired.

$$R(\mathrm{t}|\mathrm{X}) = \exp\left(-\int_0^t \lambda(t, X)ds\right) = \exp\left(-\int_0^t \left(\frac{\delta}{\alpha}\left(\frac{\mathrm{s}}{\alpha}\right)^{\delta-1} \exp(\beta X)\right)ds\right) \tag{4}$$

Put Eqs. (2) and (4) into Eq. (3), the result shows below.

$$L(\delta,\alpha,\beta) = \prod_{i=1}^{\mathrm{q}}\left(\frac{\delta}{\alpha}\left(\frac{\mathrm{s}}{\alpha}\right)^{\delta-1}\exp(\beta X)\right)\prod_{i=1}^{n}\left(\exp\left(-\int_0^t\left(\frac{\delta}{\alpha}\left(\frac{\mathrm{s}}{\alpha}\right)^{\delta-1}\exp(\beta X)\right)ds\right)\right) \times\prod_{i=1}^{\mathrm{q}}\prod_{j=1}^{k}p\left(X_{j-1}^{i},X_{j}^{i}\right) \tag{5}$$

Logarithmic on both side of equation, the result shows below.

$$\begin{aligned}&\ln(L)\\&=\mathrm{q}\ln\left(\frac{\delta}{\alpha}\right)+\sum_{\mathrm{i}=1}^{q}\ln\left[\left(\frac{\mathrm{s}}{\alpha}\right)^{\delta-1}\right]+\sum_{\mathrm{i}=1}^{q}\beta X-\sum_{i=1}^{n}\int_0^t\left(\frac{\delta}{\alpha}\left(\frac{\mathrm{s}}{\alpha}\right)^{\delta-1}\exp(\beta X)\right)ds\\&+\sum_{i=1}^{n}\sum_{\mathrm{j}=1}^{m}\ln\left(p\left(X_{j-1}^{i},X_{j}^{i}\right)\right)\end{aligned} \tag{6}$$

In assumption

$$U=\int_0^{t_{\mathrm{j}}}\left(\frac{\delta}{\alpha}\left(\frac{\mathrm{s}}{\alpha}\right)^{\delta-1}\exp(\beta X)\right)ds \tag{7}$$

Because assuming the law of state transition is right indicial function, Eq. (7) can show below.

$$\begin{aligned}U&=\sum_{\mathrm{j}=1}^{m}\int_{t_{j-1}}^{t_{\mathrm{j}}}\left(\frac{\delta}{\alpha}\left(\frac{\mathrm{s}}{\alpha}\right)^{\delta-1}\exp(\beta X)\right)ds\\&=\sum_{\mathrm{j}=1}^{m}\exp\left[\beta X(\mathrm{t_j})\right]\left[\left(\frac{\mathrm{t_j}}{\alpha}\right)^{\delta}-\left(\frac{\mathrm{t_{j-1}}}{\alpha}\right)^{\delta}\right]\end{aligned} \tag{8}$$

Through logarithm log-likelihood function show below.

$$\begin{aligned}&\ln(L)\\&=\mathrm{q}\ln\left(\frac{\delta}{\alpha}\right)+\sum_{\mathrm{i}=1}^{q}\ln\left[\left(\frac{\mathrm{s}}{\alpha}\right)^{\delta-1}\right]+\sum_{\mathrm{i}=1}^{q}\beta_i X-\sum_{i=1}^{m}\sum_{\mathrm{j}=1}^{m}\exp\left[\beta_i X_{i,j}\right]\left[\left(\frac{\mathrm{t_j}}{\alpha}\right)^{\delta}-\left(\frac{\mathrm{t_{j-1}}}{\alpha}\right)^{\delta}\right]\\&+\sum_{i=1}^{n}\sum_{\mathrm{j}=1}^{m}\ln\left(p\left(X_{j-1}^{i},X_{j}^{i}\right)\right)\end{aligned} \tag{9}$$

The function is the fitness function of genetic algorithm structured.

4. Iteration

Based on meaning of Weibull model parameters, combined with experience of practice, every parameter can be determinate below.

$$\begin{cases} \delta \in [10^{-3}, 15] \\ \alpha \in [10^{-3}, 5000] \\ \beta_i \in [-10, 10] \end{cases} \tag{10}$$

Termination condition of genetic algorithm can be controlled by termination generation N. After N generation, genetic algorithm is terminating to output the best individual of population as the result.

Through many steps iteration to termination condition (N = 100), Matlab software is used to acquire estimation of α, δ, βi. The result is $\hat{\alpha}$, $\hat{\delta}$ and $\hat{\beta}_i$.

3.2 Decision-Making Target

Failure risk of equipment is probability of fault appearing in next inspection interval $\triangle T$, when affirm that the equipment is unfelt at t[13]. If life of equipment is T and failure risk of equipment is r, expression can show below.

$$\begin{aligned} r &= \mathrm{F}(t+\Delta T|t) = \mathrm{P}(\mathrm{T} < t+\Delta T|\mathrm{T} > t) \\ &= \frac{\mathrm{P}(\mathrm{T} > t)\text{-}\mathrm{P}(\mathrm{T} > t+\Delta T)}{\mathrm{P}(\mathrm{T} > t)} \\ &= 1 - \frac{R(t+\Delta T)}{R(t)} \end{aligned} \tag{11}$$

Based on failure risk to decision-making of CBM, the target is to acquire the best inspection interval through controlling failure risk. In general, through setting a threshold or acceptable range of failure risk, making dynamic decision about inspection interval of CBM.

Putting Eq. (2) into Eq. (11), expression can be acquired.

$$\mathrm{r} = 1 - \exp\left(- \int_{\mathrm{t}}^{t+\Delta T} \left(\frac{\delta}{\alpha}\left(\frac{\mathrm{s}}{\alpha}\right)^{\delta-1} \exp(\beta X)\right) ds \right) \tag{12}$$

Because assuming the law of state transition is right indicial function and putting $\hat{\alpha}$, $\hat{\delta}$ and $\hat{\beta}_i$, Eq. (12) can show below.

$$\mathrm{r} = 1 - \exp\{ - \frac{\exp(\hat{\beta} X)}{\hat{\alpha}^{\hat{\delta}}} \times \left[(t+\Delta T)^{\hat{\delta}} - t^{\hat{\delta}} \right] \tag{13}$$

From Eq. (13), the equation of inspection interval can show below.

$$\Delta T = \left[t^{\hat{\delta}} - \frac{\ln(1-r)}{\exp(\hat{\beta} X)} \times \hat{\alpha}^{\hat{\delta}} \right]^{\frac{1}{\hat{\delta}}} - t \tag{14}$$

4 Case Study

4.1 Case Background

In the same condition, engine oil data of two same type vehicles (called vehicle A and vehicle B) is detected at regular time. Main metallic wear data are chose as samples. The data of vehicle B is used for parameter estimation. The data of vehicle A is used for test data. After vehicle B data normalized, result shows in Table 1.

Through the method of principal component analysis in Document [11], the result shows the profit estimation of the first principal component is 92.43%. The first principal component almost includes all information of samples. At last, condition parameter of vehicle B can be acquired as in Table 2.

Table 1 Result of vehicle B data normalized

Density of element (ppm) motor-hour (h)	Fe	Cu	Pb	Cr	Mg
0	0.000	0.000	0.000	0.000	0.000
40	0.237	0.201	0.087	0.209	0.079
80	0.267	0.266	0.241	0.591	0.141
110	0.317	0.275	0.274	0.710	0.221
140	0.323	0.299	0.225	0.732	0.213
180	0.405	0.395	0.424	0.724	0.368
220	0.521	0.392	0.337	0.740	0.457
260	0.507	0.326	0.389	0.765	0.480
300	0.639	0.506	0.443	0.781	0.516
320	0.515	0.552	0.419	0.794	0.526
360	0.471	0.438	0.456	0.688	0.628
400	0.568	0.704	0.460	0.865	0.579
440	0.722	0.803	0.758	0.901	0.693
480	0.949	0.678	0.711	0.881	0.733
520	0.764	0.913	0.835	0.873	0.752
560	0.738	0.771	0.839	0.889	0.795
600	0.722	0.873	0.821	0.971	0.864
640	1.000	0.963	0.948	0.979	0.960
702	0.970	1.000	1.000	1.000	1.000

Table 2 Condition parameter of vehicle B

Motor-hour (h)	Density of element (ppm)	Motor-hour (h)	Density of element (ppm)
0	0	360	1.193
40	0.361	400	1.411
80	0.663	440	1.729
110	0.791	480	1.763
140	0.788	520	1.848
180	1.026	560	1.800
220	1.085	600	1.896
260	1.093	640	2.167
300	1.282	702	2.221
320	1.246		

4.2 Parameter Estimation

Trough data of Table 2, parameters in WPHM can be estimated by genetic algorithm, the result can show below.

$$\begin{cases} \hat{\alpha} = 628.75 \\ \hat{\delta} = 15 \\ \hat{\beta} = -0.63 \end{cases} \tag{15}$$

WPHM can show below.

$$\lambda(t, X) = \frac{15}{628.75}\left(\frac{t}{628.75}\right)^{14} \exp(-0.63X) \tag{16}$$

4.3 Decision-Making Result

1. Data processing
   Data of vehicle A is preconditioned, normalized, principal component analyzed; condition parameter of vehicle A can be acquired as in Table 3.
2. Determination on inspection interval
   Based on Eqs. (14) and (15), the expression of inspection interval can be acquired below.

$$\Delta T = \left[t^{15} - \frac{\ln(1-r)}{\exp(-0.63X)} \times 628.75^{15}\right]^{\frac{1}{15}} - t \tag{17}$$

Table 3 Condition parameter of vehicle A

Motor-hour (h)	Density of element (ppm)	Motor-hour (h)	Density of element (ppm)
0	0	350	1.593
40	0.573	394	1.578
82	0.738	430	1.844
110	1.004	470	2.026
150	1.037	510	2.080
192	1.207	550	2.164
230	1.304	590	2.204
272	1.377	610	2.236
311	1.457		

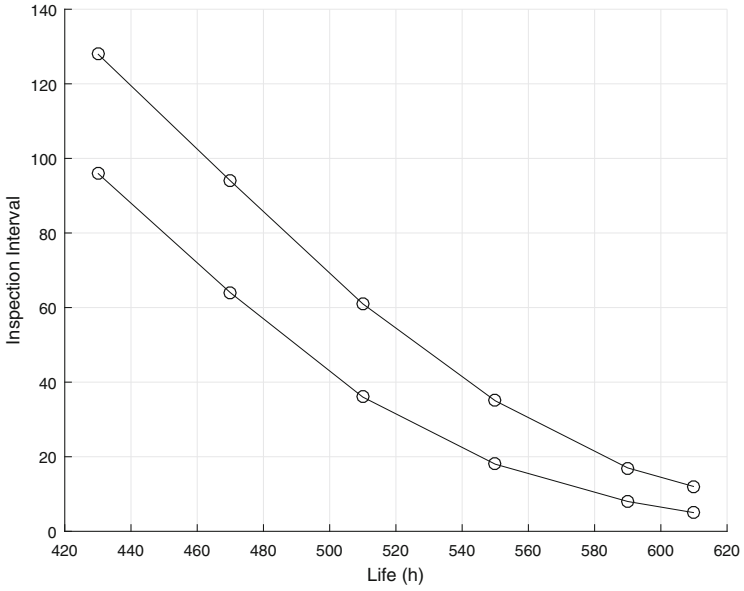


Fig. 2 Decision-making range of inspection intervals

The more serious failure risk, the more dangerously equipment runs. The lower serious failure risk, the more expenses equipment maintenance is. Based on experience, failure risk between 0.02 and 0.05 can be accepted.

After engine running a series time, condition descending and inspection interval need decrease to acquire condition of engine. The dynamic decision-making result of vehicle A engine after 430 motor-hours shows in Fig. 2.

Figure 2 shows decision-making range of inspection interval through controlling failure risk. In general, with life of equipment rising and degradation degree of condition rising, inspection interval need decrease to acquire change of equipment condition. It can confirm safety and task be controlled in acceptable range. Through

controlling failure risk, upper bound and lower bound of inspection interval at different stages can be acquired. Decision maker can make reasonable decision after considering different factors.

5 Conclusion

In the text, condition parameters be coverable is put into reliability model of equipment. Through Weibull proportional hazards model combine with both. The result of decision-making inspection interval can be acquired by controlling failure risk being target. In the process of decision-making, condition parameters is sufficiently considered. Through controlling, decision maker owns effective decision space.

References

1. Salvinder S, Shahrum A, Nik Abdullah NM et al (2015) Markov chain modelling of reliability analysis and prediction under mixed mode loading. Chin J Mech Eng 2(28):307–314
2. Sha N, Pan R (2014) Bayesian analysis for step-stress accelerated life testing using weibull proportional hazard model. Stat Pap 3(55):715–726
3. Wen W, Guang M, Fang L et al (2011) Lifetime analysis of lead-free solder joints under drop impact using proportional hazards model. J Vib Shock 30(3):124–128
4. Mehta P, Werner A, Mears L (2015) Condition based maintenance-systems integration and intelligence using Bayesian classification and sensor fusion. J Intell Manuf 36(2):331–346
5. GJB1378-92. Requirements and procedure of developing preventive maintenance program for material
6. Lawless JF (1998) Statistics model and method in life statistics. China Statistics Press, Beijing
7. Jiang ST, Landers TL, Rhoad TR (2006) Assessment of repairable-system reliability using proportional intensity model: a review. IEEE Trans Reliab 55(2):328–336
8. Zuo H, Cai J (2008) The theory and method of maintenance decision-making. Aeronautics Industry Press, Beijing
9. Bas E, Uslu VR, Yolcu U, Egrioglu E et al (2014) A modified genetic algorithm for forecasting fuzzy time series. Appl Intell 41(2):453–463
10. Ma M, Liu Y, Xu X et al (2014) Selection of shifting element design based on genetic algorithm. J Beijing Univ Aeronaut Astronaut 40(10):1372–1377
11. Grama SN, Subramanian SJ (2014) Computation of full-field strains using principal component analysis. Exp Mech 54(6):913–933

Research on the Fault Diagnosis Method of Equipment Functionally Significant Instrument Based on BP Neural Network

Xiang Zan, Shi-xin Zhang, Heng Gao, Yang Zhang, and Chao-shuai Han

Abstract Through BP neural network, a nonlinear mapping model between feature information and diagnosis result is proposed. The method is applied in diagnosing fault of equipment functionally significant instrument, which can fuse all kinds of information in the running process of the functionally significant instrument. For the supervisory architecture, first-sitting-weight and first-sitting-threshold for BP neural network are used. Genetic algorithm is applied in neural network optimization. At last, the method is applied in fault diagnosis of a transmission system. Through comparison between diagnosis results and real results, the method is validated.

Keywords Fault diagnosis • Neural network • Genetic algorithm • Functional significant instrument

1 Introduction

With the development of modern technology to push equipment, equipment fault mode changes from mechanical fault to mixed mode including mechanical, electricity, fluid, optical and software. The traditional maintenance method based on time is difficult to adapt to the need of modern maintenance and support. To solve the problem of over maintenance and insufficient maintenance, Condition-based

X. Zan (✉) • S.-x. Zhang
Department of Technical Support Engineering, Academy of Armored Force Engineering, Beijing 100072, China
e-mail: zan_xiang.student@sina.com; zsxzh@sohu.com

H. Gao
Teach Room of Voluntary Artillery, Academy of Nanjing Artillery, Nanjing 102205, China
e-mail: 356165891@qq.com

Y. Zhang
Military Deputy Office of PLA in 674 Factory, Harbin 150056, China
e-mail: 147267728@qq.com

C.-s. Han
Troop No. 63960 of PLA, Beijing 102205, China
e-mail: 929134680@qq.com

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_31

Maintenance is raised and developed. Condition-based Maintenance judges difference in different individuals through fault detection and fault diagnosis. It is the base tone of Autonomic Logistics System and equipment efficient support.

Fault diagnosis of equipment Functionally Significant Instrument is the key step to Condition-based Maintenance. The effective information from all kinds of condition information of Functionally Significant Instrument is used as diagnosis evidence. The information is used for comprehensive evaluation on condition, judgment on degradation degree and judgment on fault location to support to maintenance decision-making.

There are many fault diagnosis methods at home and abroad. Stephen and cooperators researched fault prediction and diagnosis system based on artificial neural network theory, the function of which included fault prediction, pattern recognition and condition prediction [1]. Cheng designed and simulated diagnosis system of satellite attitude control system based on nonlinear unknown input observer [2]. Xu applied a new fast support vector algorithm into diagnosis system of aero-engine to solve disadvantages of support vector training part [3].

Due to high technology and complex structure of modern equipment Functionally Significant Instrument, condition signs and information are numerous [4]. So fault diagnosis must establish effective mapping relation between condition information and diagnosis result and pick up effective information from all kinds and numerous condition information.

2 Notes on Fault Diagnosis of Functionally Significant Instrument

2.1 Notes on Functionally Significant Instrument

Failure consequence of Functionally Significant Instrument may be one of conditions below (GJB1378-92) [5].

1. May impact the service safety of equipment,
2. May impact the complete of tasks,
3. May lead to heavy economic losses,
4. Hidden function failure and other failure may together lead to one or some consequences above or quadratic effect may lead to the same consequences.

Based on related theory of Condition-based Maintenance, FSI is the object of equipment Condition-based Maintenance.

3 Fault Diagnosis Method Based on BP Neural Network

3.1 Structure of BP Neural Network

BP (Back Propagation) neural network is a back propagation neural network algorithm [6]. The study process of BP Neural Network includes two processes,

which are forward propagation of signal and back propagation of error. In the study process, if forward output signal is mismatching with real result, weight and threshold will be revised through back propagation. Weight is continually adjusted by forward and back propagation to train the best neural network [7].

Basic BP neural network constitutes of input layer, output layer and hidden layer, and there may be many hidden layers [8]. There is a network including M layers, P nodes and N samples (x_k,d_k) (k = 1,2,...,N). If $I_{jk}{}^m$ shows input total of j-th note in m-th layer, output is $O_{jk}{}^m$ and W_{ij} shows weight between i-th note in m-1-th layer and j-th note in m-th layer, result can be acquired below.

$$I_{jk}^{l} = \sum_{i=1}^{n_l} W_{ij} O_{jk}^{l-1} \tag{1}$$

Because the number of notes in every hidden layer is uncertain, assume the number is n_l.

$$O_{jk}^{l} = f\left(I_{jk}^{l}\right) \tag{2}$$

where $f()$ shows recursion relation.

In the process of back propagation, quadratic sum of error between desire output and real output is seen as target function below.

$$E_k = \frac{1}{2}\sum_{j=1}^{m}\left(d_{jk} - y_{jk}\right)^2 \tag{3}$$

So sum error of N samples shows below.

$$E = \frac{1}{2S}\sum_{k=1}^{N} E_k \tag{4}$$

The target on study process of BP neural network is decreasing sum error E through adjusting weight W to weight change on the side of negative gradient of error function. The expression shows below.

$$W_{ij}(t+1) = W_{ij}(t) - \eta\frac{\partial E}{\partial W_{ij}} \tag{5}$$

In expression, t shows iterated times, η shows step size.

There is only one hidden layer of BP neural network showing as in Fig. 1.

In Fig. 1, $x_i(i = 1,2,3,...)$ shows input signal, $y_p(p = 1,2,3,...)$ shows output signal, $h_j(j = 1,2,3,...)$ shows hidden layer, v_{ij} shows weight between input layer and hidden layer, w_{jp} weight between hidden layer and output layer, a_j threshold of hidden layer, b_p threshold of output layer.

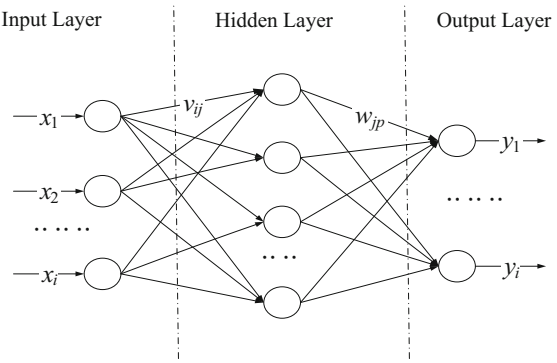


Fig. 1 BP neural network

3.2 *Microscopic Analysis of Fault Diagnosis Based on BP Neural Network*

1. Microscopic analysis of fault diagnosis

   System is seen as a whole. In the running process of system, all kinds of condition information, which exist difference. Diagnostic information is the effective information from multi-dimensional condition information space in the running process of system. If C shows ensemble of different conditions, P shows condition information space, S shows diagnostic information space, the steps of fault diagnosis shows below.

   - Process of diagnostic information extraction

     Effective information is extracted form condition information space, some of which characterizing Diagnostic condition is diagnostic information.
   - Process of fault diagnosis

     Fault diagnosis is $f{:}S{\rightarrow}C$ mapping process, which is usually nonlinear mapping. f is a multitude to one mapping relation. It means one condition can show through multi-characteristics information, but one characteristic information is corresponding to one condition.

2. Process of fault diagnosis based on BP neural network

   The characteristic of neural network is not necessary to establish related model. Neural network can acquire corresponding mapping ability through interior study and process of error revising. Neural network owns well applicability. Then there are two processes including forward study and back revising in the train of BP neural network. The best weight and threshold can be acquired by revising error, which owns nonlinear mapping ability in the process of fault diagnosis. The process of fault diagnosis can show as in Fig. 2. Diagnostic information of diagnostic space is input signal. Weights are revised by train process. Diagnostic information is recombined fused through information fusion ability of neural network. At last, output is mapped to condition space and diagnosis result can be acquired.

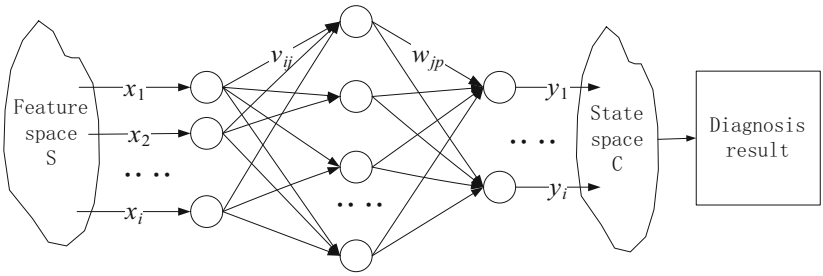


Fig. 2 Process of fault diagnosis based on BP neural network

3.3 Disadvantages of BP Neural Network

BP neural network is a simple method to diagnosis and algorithm is easy to realize. But disadvantages of BP neural network shows below.

1. Convergent speed of BP neural network is slow. The model is easy to fall into local optimum. It the process of convergent "climbing" ability is weak.
2. Structure, initial weights and threshold of BP neural network have great influence on train of network. But accurate initial value is difficult to acquire, which results that diagnosis result is inaccurate.

Due to BP neural network existing disadvantages, it should be optimized to improve precision of diagnosis.

4 Optimization of BP Neural Network Based on Genetic Algorithm

4.1 Design of Genetic Algorithm

Genetic Algorithm is an agglutination algorithm that can searching optimum solution in feasible solution space [9]. It can divide feasible solution space into different grades based on fitness function through choice, crossover and mutation to acquire the best solution gradually.

1. Encoding methods

   Based on situation, 0-1 Encode is used. Every chromosome constitutes of four parts which are v_{ij} between input layer X and hidden layer H, w_{ij} between hidden layer X and output layer Y, threshold a_j of hidden layer H and threshold b_p of output layer Y ($i,j,p = 1,2,3,\ldots$). Coding of four parts are connected to form a chromosome.

2. Initialization populations
   The binary initial populations is produced random, whose initial length is confirmed by the length of v_{ij},w_{ij},a_j,b_p (i,j,p = 1,2,3,...).
3. Crossover
   Single cutting-point crossover is used. Selecting strategy of directly proportion is applied to parent-chromosomes. Middle parts of elder chromosomes are exchanged. Inherit-chromosomes can be acquired.
   Mutation is changing location value of some codes selected chromosomes
4. Fitness function
   Because the target of Genetic Algorithm is reducing error of BP neural network output, error between output values of samples and real values is used as fitness function.
5. Termination condition
   When the time of iteration can reach to setting times, operation is over.

4.2 The Optimization Process of BP Neural Network Based on Genetic Algorithm

The Optimization Process of BP Neural Network based on Genetic Algorithm shows in Fig. 3.

Explaining of optimization process shows below.

1. In the process of crossover, crossover rate Pc needs to be set. In the process of operation, ζ is taken from 0 to 1. If ζ>Pc, parent individuals will be kept. If not, single cutting-point crossover is done.
2. In the process of mutation, mutation rate Pm needs to be set. In the process of operation, ζ is taken from 0 to 1. If ζ>Pm, parent individuals will be kept. If not, mutation is done.
3. In the process of population-mixing, if K is the scale of population, parent-chromosomes are replaced by inherit-chromosomes based on fitness values. The law of process is keeping high-value chromosomes through sorting the chromosomes based on fitness values. At last the former K chromosomes is used for new population.

5 Case Analysis

5.1 Case Preparation

1. Case data
   Fault Diagnosis of equipment transmission is the case. Based on text [10], six parameters are chosen as feature information in Fig. 4.

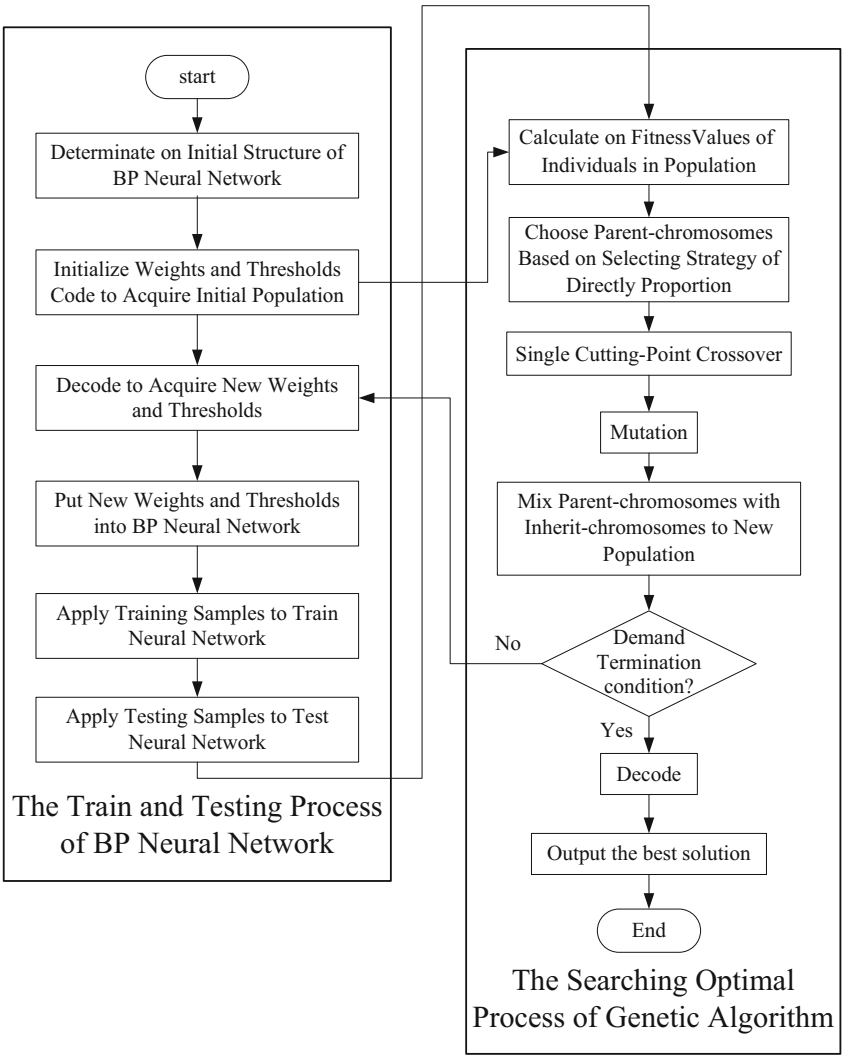


Fig. 3 The optimization process of BP neural network based on genetic algorithm

One group detecting data of transmission is pretreated and normalized. The result shows as in Table 1.

The group data is used as Training Sample Data. Because Transmission owns three conditions that Normal, Wear and Crack. Coding the three conditions can

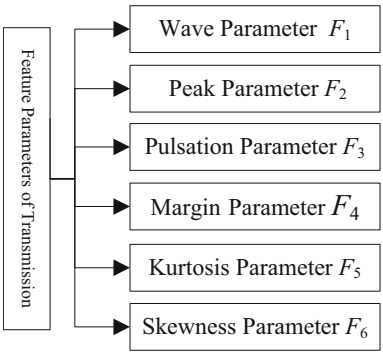


Fig. 4 Index about fault diagnosis of equipment transmission

Table 1 Training sample data of equipment transmission

Sample	F_1	F_2	F_3	F_4	F_5	F_6	Fault diagnosis result
1	0.2371	0.6461	0.8431	1	0.5203	0.5531	Normal
2	0.2041	0.6709	0.8311	1	0.4843	0.5723	Normal
3	0.1881	0.6496	0.8452	1	0.5184	0.5601	Wear
4	0.1456	0.6013	0.8107	1	0.5089	0.8703	Wear
5	0.1373	0.6069	0.8163	1	0.5049	0.8314	Crack
6	0.0432	0.5011	0.8076	1	0.4697	0.9003	Crack

Table 2 Training sample data

Sample	Input arrow	Output arrow
1	(0.2371 0.6461 0.8431 1.000 0.5203 0.5531)	(1 0 0)
2	(0.2041 0.6709 0.8311 1.000 0.4843 0.5723)	(1 0 0)
3	(0.1881 0.6496 0.8452 1.000 0.5184 0.5601)	(0 1 0)
4	(0.1456 0.6013 0.8107 1.000 0.5089 0.8703)	(0 1 0)
5	(0.1373 0.6069 0.8163 1.000 0.5049 0.8314)	(0 0 1)
6	(0.0432 0.5011 0.8076 1.000 0.4697 0.9003)	(0 0 1)

acquire ideal output showing Normal-001, Wear-010 and Crack-001. After coded, Training Sample Data shows in Table 2.

Other detecting data is used as Testing Sample Data. After disposed as above, Testing Sample Data shows in Table 3.

2. Design of Neural Network

Three layers of BP neural network are designed based on the question, which owns six nodes in input layer, three nodes in output layer. Based on experience, there is a proximity relation between the number of hidden layer n_2 and the number of input layer n_1 showing below [11].

Table 3 Testing sample data

Sample	Input arrow	Output arrow
1	(0.2113 0.6503 0.8259 1.000 0.4902 0.5063)	(1 0 0)
2	(0.2171 0.6489 0.8292 1.000 0.5011 0.5498)	(1 0 0)
3	(0.1793 0.6602 0.8432 1.000 0.5201 0.5610)	(0 1 0)
4	(0.1738 0.6601 0.8453 1.000 0.5181 0.5011)	(0 1 0)
5	(0.1373 0.6069 0.8163 1.000 0.5049 0.8306)	(0 0 1)
6	(0.0574 0.5804 0.8172 1.000 0.4927 0.9796)	(0 0 1)

Table 4 Parameters of Genetic Algorithm

Population scale	Generation of genetic	Crossover probability	Mutation probability	Generation gap
40	50	0.9	0.01	0.9

$$n_2 = 2 \times n_1 + 1 \tag{6}$$

So the number of hidden layer is 13=6×2.

The training process of neural network is adjusting weights and thresholds to reduce error. The number training time is 1000. The number training target is $1.0 \times e^{-6}$. The number study rate is 0.1.

5.2 *Operation Result*

1. Optimization of BP neural network based on genetic algorithm

   First BP neural network is optimized through genetic algorithm. Related parameters are set in Table 4.

Through training samples BP neural network is trained. The result shows in Fig. 5.

2. Fault diagnosis based on BP neural network

   After optimized through genetic algorithm, BP neural network is used to diagnose fault. The result shows in Table 5.

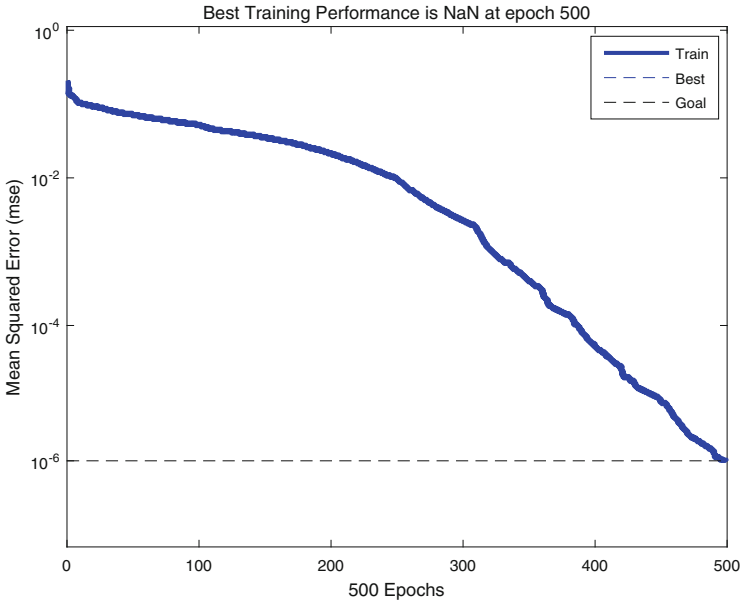


Fig. 5 Error evolution curve index

Table 5 Diagnosis result comparison

Sample	Output arrow after optimization	Output arrow before optimization
1	(0.9994 0.0010 0.0000)	(0.9999 0.0000 0.1784)
2	(0.9999 0.0000 0.0000)	(0.9991 0.0017 0.0000)
3	(0.0001 0.9994 0.0063)	(0.0010 0.9985 0.0005)
4	(0.0000 0.9998 0.0016)	(0.0003 00000 0.2589)
5	(0.0000 0.0005 0.9988)	(0.0000 0.0017 0.9986)

5.3 Analysis of Result

Through diagnosis result comparison, error of the testing results from optimized BP neural network reduces from 1.0339 to 0.5933. And error of the training samples from optimized BP neural network reduces from 1.0339 to 0.5933. Optimization effect is obvious. Through optimized BP neural network to fault diagnosis of Functionally Significant Instrument, the result fits the real result and veracity can be confirmed.

6 Conclusion

Fault diagnosis of equipment Functionally Significant Instrument is the key technology to Condition-based Maintenance and the key element of Autonomic Logistics. No matter maintenance in peacetime or wartime, fault diagnosis is an important link. In the text, fault diagnosis method based on BP neural network is researched. To disadvantages of BP neural network, Genetic Algorithm is used to optimize the structure of BP neural network and acquire the best weights and thresholds. At last, through a case, effectiveness of the method is verified.

References

1. Onk S, Maldonado FJ et al (2012) Predictive fault diagnosis system for intelligent and robust health monitoring. J Aerosp Comput Inf Commun 9(4):125–143
2. Cheng Y, Hou Q, Jiang B (2012) Design and simulation of fault diagnosis based on NUIO/LMI for satellite attitude control systems. J Syst Eng Electron 23(4):581–587
3. Xu Q-H, Geng S, Shi J (2012) Aero-engine fault diagnosis applying new fast support vector algorithm. J Aerosp Power 27(7):1605–1611
4. Zhang Y-H, Wang S-H, Han X-H (2013) Research status and prospect of condition-based maintenance decision-making. J Acad Armored Force Eng 2:6–13
5. GJB1378-92 (1992) Requirements and procedure of developing preventive maintenance program for material
6. Nazri Mohd N, Abdullah K, Mohammad ZR (2013) A new back-propagation neural network optimized with cuckoo search algorithm. Lect Notes Comput Sci 7971(1):427–437
7. Zhu D, Shi H (2006) Artificial neural network and application. Press of Science, Beijing
8. Tian L, Luo Y, Wang Y (2013) Prediction model of TIG welding seam size based on BP neural network optimized by genetic algorithm. J Shanghai Jiaotong Univ 47(11):1691–1701
9. Li J, Chen H, Zhong Z et al (2014) Method for electromagnetic detection satellites scheduling based on genetic algorithm with alterable penalty coefficient. J Syst Eng Electron 25 (5):822–832
10. Li S (2007) Research into the fault diagnosis expert system based on neural network. Northeast Normal University
11. Feng S, Wang H, Yu L (2011) 30 cases study of MATLAB metaheuristic. Press of Beijing University of Aeronautics and Astronautics, Beijing

Gearbox Fault Diagnosis Based on Fast Empirical Mode Decomposition and Correlated Kurtosis

Xinghui Zhang, Jianshe Kang, Rusmir Bajrić, and Tongdan Jin

Abstract Gear and bearing are widely used in a variety of rotating machinery. They are considered as the most critical components of gearbox system. Vibration-based feature extraction is an effective approach to fault diagnosis of gear and bearing units. Fault diagnosis plays an important role in assuring equipment availability and reducing operational costs. This paper proposes a new fault diagnosis method that synthesizes time synchronous technique, fast empirical mode decomposition and correlated kurtosis. Our study shows that an improved fault feature extraction method can be obtained from sampled vibration signals. Energy of gear wheel rotational frequency and its harmonics are designated as the degradation indicators. The effectiveness of proposed method is verified and validated using the vibration data from gearbox test rigs and commercial wind turbines.

Keywords Gearbox • Empirical mode decomposition • Correlated kurtosis • Fault diagnosis • Wind turbine

1 Introduction

Gearbox is an important part of transmission systems used in helicopters, wind turbines and ground vehicles. Unexpected or unplanned failures usually result in large production losses and maintenance costs. Many vibration techniques have been developed for gearbox fault diagnosis and condition monitoring with different levels of efficiency. However, gearbox faults by their nature are time dependent and

X. Zhang (✉) • J. Kang
Mechanical Engineering College, Shijiazhuang, Hebei 050003, China
e-mail: dynamicbnt@gmail.com; jskang201206@126.com

R. Bajrić
Public Enterprise Elektro Privreda BiH, Coal Mines Kreka, Tuzla 75000, Bosnia and Herzegovina
e-mail: rusmir.bajric@kreka.ba

T. Jin
Ingram School of Engineering, Texas State University, San Marcos, TX 78666, USA
e-mail: tj17@txstate.edu

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_32

non-stationary phenomena, belonging to localized transient events. To deal with the non-stationary signals generated by gearbox, empirical mode decomposition (EMD) has been widely used for mechanical fault diagnosis domain [1]. EMD is more suitable to handle non-stationary and non-linear mechanical fault signals than wavelet analysis, S-transform, and Fourier transform methods. To effectively identify gearbox faults, the use of EMD and its variants or the combination with other signal processing methods became an important research topic.

Extensions have been made on EMD to overcome the mode mixing phenomenon, including ensemble empirical mode decomposition (EEMD) and the noise assisted multivariate empirical mode decomposition (MEMD) [2, 3]. These techniques are increasingly adopted for fault diagnosis and prognosis of mechanical systems. Wu et al. [4] use EEMD and autoregressive model to detect mechanical looseness faults. Ibrahim and Albarbar [5]compare the gear fault diagnosis effect of Wigner-Ville distribution and EMD. Experimental results show that EMD is able to detect the early faults more effectively than other methods. Guo et al. [6] propose an impulsive signal recovery method based on spectral kurtosis (SK) and EEMD. Later, Lei et al. [7] review the application of EMD in fault diagnosis of rotating machinery. Its applications in fault diagnosis of bearings, gears and rotors have been discussed extensively and the main challenges are highlighted as well. In this paper, we propose an approach to selecting the optimal frequency band containing fault impulse signal using spectral kurtosis. Then, correlated coefficients are used to determine the best intrinsic mode function (IMF) decomposed by EEMD for post processing.

Recently, many researchers attempt to diagnose mechanical faults using EMD or its variants combined with other signal processing methods, such as machine learning. Signal processing methods consist of wavelet transform, wavelet packet decomposition, minimum entropy deconvolution, morphological filter, spectral coherence, Bi-spectrum, and high order spectrum. For statistical-based machine learning methods, there are self-zero space projection analysis, high order cumulant, artificial neural network, fault leading algorithm, least square support vector machine, support vector machine, ensemble optimal extreme learning machine, Hidden Markov model and singular value decomposition. Besides these methods, Feng et al. [8] investigate the problem of planetary gearbox fault diagnosis using EEMD. Zheng et al. [9] propose a generalized empirical mode decomposition (GEMD) method for bearing fault diagnosis. Li et al. [10] develop a differential based empirical mode decomposition to analyze a multi-fault issue. Dybała and Zimroz [11] construct a new method by integrating all IMFs into three combined mode functions. Hence the fault signal can be divided into three components: noise-only part, signal-only part and trend-only part. To resolve the capacity problem in data transmission through wireless communication, Guo and Tse [12] propose a new bearing fault signal compression method based on optimal EEMD. Georgoulas et al. [13] develop an anomaly detection method based on EMD and three machine learning methods, namely Gaussian mixed model, nearest neighbor model, and principal component analysis.

However, the major drawback of EMD and EEMD is the low processing speed when the sample size of signals becomes large. If the number of samples used for

decomposition is too small, it results in poor resolution for spectral analysis. Another trade-off of EMD is the calculation of IMFs using cubic spline function. It is effective to extract the harmonic feature, but fail to extract the transient feature. These issues are very inconvenient for fault diagnosis. In order to overcome this problem, Wang et al. [14] redesign the algorithm and the code such that EMD and EEMD have fast processing speed even if the signal under decomposition is very long.

Previous researchers usually select IMFs for post process using kurtosis and correlation coefficient [6] method. The proposed new criterion in this paper is named correlated kurtosis (CK) and will be used to select the best IMF. CK was proposed by McDonald et al. [15]. They develop a maximum CK deconvolution to replace the minimum entropy deconvolution because of its good effect. To mitigate specified problems, we propose hybrid fault diagnosis method based on time synchronous technique, fast EMD and CK. In particular, time synchronous technique is used to eliminate the rotating speed variation. Hence it guarantees the feasibility of spectral analysis of IMFs without spectral dispersion phenomenon. CK is used to select the best IMF through the comparison with other two indicators (i.e. kurtosis and correlation coefficient). Two data sets are used to verify the proposed method. One is the gear fault data with different tooth wear levels generated by test rigs in the laboratory setting. The other is the bearing fault data of a commercial wind turbine. The results show that the proposed method can effectively diagnose gear wear fault and gear wheel rotational frequency energy with first and second harmonic outperform root mean square (RMS) for degradation tracking. For bearing fault, IMF determined by CK contains obvious fault information.

This remainder of the paper is organized as follows. Section 2 introduces the algorithm of Fast EMD. Section 3 proposes the fault diagnosis framework of gear and bearing. Section 4 discusses the results of gear fault diagnosis and degradation track as well as the outcome of bearing fault diagnosis, and Sect. 5 concludes the work.

2 Fast Empirical Mode Decomposition

EMD algorithm is able to decompose a signal into finite IMFs which are extracted through an iterative sifting process. Generally, IMFs satisfy two conditions: First, in the whole time domain, the number of extrema and the number of zero-crossings are either equal or differ by only one. Second, the mean value of the envelopes defined by local maxima and local minima is zero. Given a signal $y_0(t)$, $t \in [1, n]$, the upper and the lower envelops can be acquired by cubic splines. The average of two envelops is obtained by subtracting from the original signal. First IMF is obtained through repeating the sifting process several times. Usually, the first IMF contains relatively higher frequencies compared to the residual signal. Next, this residual

signal can be continually decomposed into second IMF and new residual signal. If this process is repeated, a series of IMFs and residual signal $r(t)$ are obtained. The whole decomposition process can be represented using the following equation:

$$y_o(t) = \sum_{m=1}^{M} c_m(t) + r(t) \tag{1}$$

Where M is the number of IMFs. Detailed algorithm can be referred to [1].

Traditionally, EMD algorithm is known as computationally intensive. It can only handle the decomposition of short signals, resulting in a low resolution of spectrum. Recently Wang et al. [14] proved that the time complexity of EMD or EEMD is $T = 41 \bullet NE \bullet NS \bullet n(\log_2 n) = \mathrm{O}(n \log n)$ where n is the signal length and NS and NE are the sifting and ensemble numbers, respectively. This is equal to the complexity of FFT. The space complexity of EMD is $M = (13 + \log_2 n) \bullet n = \mathrm{O}(n \bullet \log n)$. In addition, they optimized the program and a fast EMD algorithm was developed with the speed increased by 1000 times. Therefore this fast EMD algorithm is capable of processing the long vibration signals with high resolution quality.

3 The Fault Diagnosis Framework

Compared with laboratory experiments, load profile and rotating speed of rotating machinery in industrial applications generally are non-stationary. Take the wind turbine for example. A three percent of speed variation can cause 10 Hz difference of the bearing fault frequency [16]. If the raw vibration signal is used for EMD processing, the spectral analysis of selected IMF cannot accurately reveal the true fault information. This may lead to a miss-detection of faults. Therefore, time synchronous technique should be used for preprocessing the vibration signal collected from rotating machinery.

Presence of sideband frequencies around gear mesh frequency is the indication of amplitude modulation. For fault diagnosis, sideband frequencies represent the abnormality in the gearbox elements due to some faults or irregularity. The space between the sideband frequencies generally represents the rotating frequency of gear wheel. Therefore, one object of gear fault diagnosis is to find whether there exists an intense, equally spaced sideband frequency in frequency spectrum. Another objective could be the energy of gear mesh frequency and their associated harmonics. For bearing fault diagnosis, fault frequencies and their harmonics indicate the bearing fault presence. The object is to find the presence of these fault frequencies across the entire frequency spectrum. The detail diagnosis framework can be divided into four steps:

Step 1: Original vibration signal is processed using time synchronous technique. This technique guarantees the same sampling points for every revolution and overcomes the spectral spreading phenomenon. It should be noted that there is no need to perform the averaging operation, because the CK needs long signal to ensure the accuracy.

Step 2: Fast EMD is used to decompose the vibration signal into several IMFs containing different frequency fault information. The length of vibration signal should be long enough so that high spectral resolution of selected IMF is guaranteed. The processing can be done nearly in real time.

Step 3: CK is used to select the best IMF which contains the main fault information of gear or bearing. It can be expressed by following equation:

$$Correlated\ kurtosis\ of\ M-shift = CK_M(\tau) = \frac{\sum_{t=1}^{N}\left(\prod_{m=0}^{M} y(t-m\tau)\right)^2}{\left(\sum_{t=1}^{N} y(t)^2\right)^{M+1}} \quad (2)$$

where $y(t)$ is the input signal, τ is the interesting period of the fault, and N is the number of samples of $y(t)$. If $\tau = 0$ and $M = 1$, it gives the traditional kurtosis and can be used to detect the specific periodic impulse signals. For example, if the desired fault frequency is 50 Hz and the sampling frequency is 10,000 Hz, the value of τ could be 200 samples.

CK can detect the presence of some specific fault types. This characteristic can be used to find the best IMF that contains specific fault information and is very beneficial for degradation tracking when multi-fault exist.

Step 4: The selected IMF can be used for post processing. For gear fault diagnosis, gear wheel rotational frequency and the associated harmonics in envelope spectrum are important indicators to determine which gear wheel has fault. In addition, gear wheel rotational frequency energy and their associated harmonics can be used to track the fault degradation trend. For bearing fault diagnosis, specific fault frequencies of envelope spectrum are the indicators to identify faulty component. The fault diagnosis framework based on Fast EMD and CK is illustrated in Fig. 1.

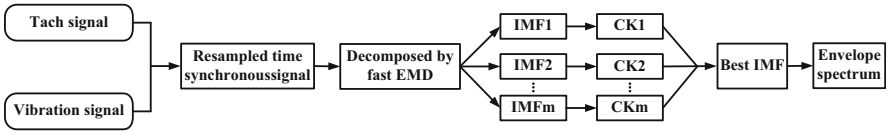


Fig. 1 Fault diagnosis framework based on fast EMD and CK

4 Experimental Validation

4.1 Experimental Setup

In this section, the proposed method is used to analyze the experimental vibration signal of a gearbox in a laboratory setting. The setup of the gearbox test rig is illustrated in Figs. 2 and 3. The test rig contains a 4 kW three phase asynchronous speed-adjustable motor for driving the gearbox, two-stage gearbox and a magnetic powder brake. The adjusted load is provided by the powder brake connected to the gearbox output shaft.

Vibration signals are measured by accelerometer secured by magnetic bases. The accelerometer is mounted on the top of low shaft speed bearing pedestal of the tested gearbox and the data are sampled at the frequency of 20 kHz. The tachometer

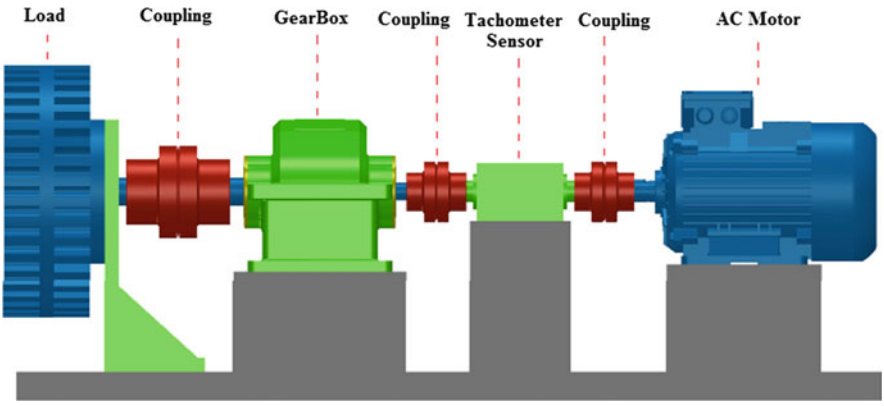


Fig. 2 Experimental gearbox test rig

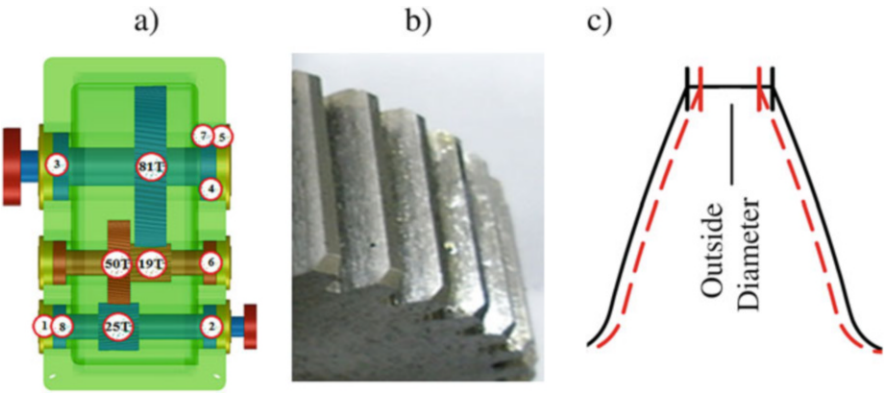


Fig. 3 (**a**) Gearbox schematic. (**b**) Partial gear wheel. (**c**) Tooth material removal

resolution is 60 impulses per revolution. The teeth number and rotational speed of high speed shaft are 25 and 13.33 Hz, respectively. Gear wheel on the high speed shaft meshes with gear wheel on intermediate speed shaft with 50 teeth. Gear wheel on the low speed shaft has 81 teeth and meshes with gear wheel on intermediate speed shaft with 19 teeth, which gives 1:8.53 gearbox transmission ratio and 1.56 Hz speed of low speed shaft. Gear wheel on low speed shaft is used for testing. The presented results refer to nominal low speed shaft torque of 5.54 Nm and the tests were carried out under constant load.

4.2 Processing Results

Distributed faults, as uniform wear around whole gear wheel generally produce high level side bands and its associated harmonics in the envelope spectrum. The goal of CK is to detect the periodic impulse with a period of about 25,479 samples (new sampling frequency after time synchronous processing) divided by gear wheel rotational frequency. However, the length of time synchronous averaging (TSA) generally is very short. It can only guarantee that there is one period in TSA data. Therefore, there is no need to perform the averaging operation. It only needs to resample the signal according to tachometer signal for assuring equal angle-interval sampling. That ensures the signal length is long enough to calculate the CK value. After time synchronous processing, the signal can be decomposed into ten IMFs using Fast EMD. For illustration purpose, the vibration data of gear wear state 2 is being analyzed. Figure 4 is the results of ten IMFs, and Fig. 5 is the corresponding frequency spectrum of FFT.

In this paper, all the computations including the plot of figures are performed in Matlab environment. Matlab is a very useful tool for scientific calculations in research. Certainly, other software applications can also be used to implement these functions also.

5 Discussions

From Fig. 5, we can see that IMF 2 contains more harmonics of low gear wheel rotational frequency and higher energy. The values of three indicators of these IMFs are depicted in Figs. 6, 7, and 8. They correspond to correlation coefficient (CC), kurtosis and correlated kurtosis (CK), respectively. The approximately best IMF can be determined from these indicators. Root mean square values (RMS) of original vibration signals prior to EMD processing, and optimal IMF selected by CK value can be acquired for five wear states of gear. In addition, energy of low

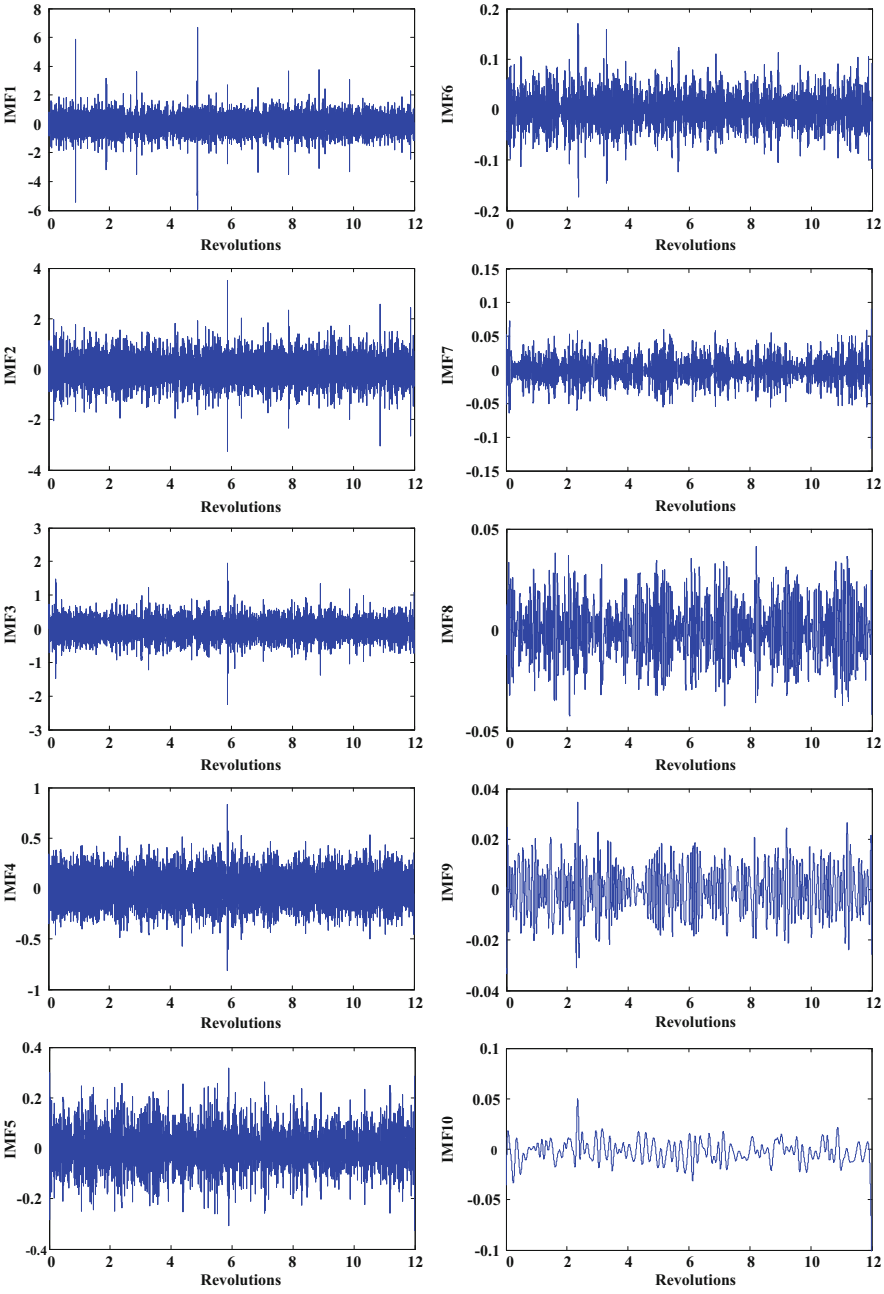


Fig. 4 IMFs decomposed from gear wear level 2

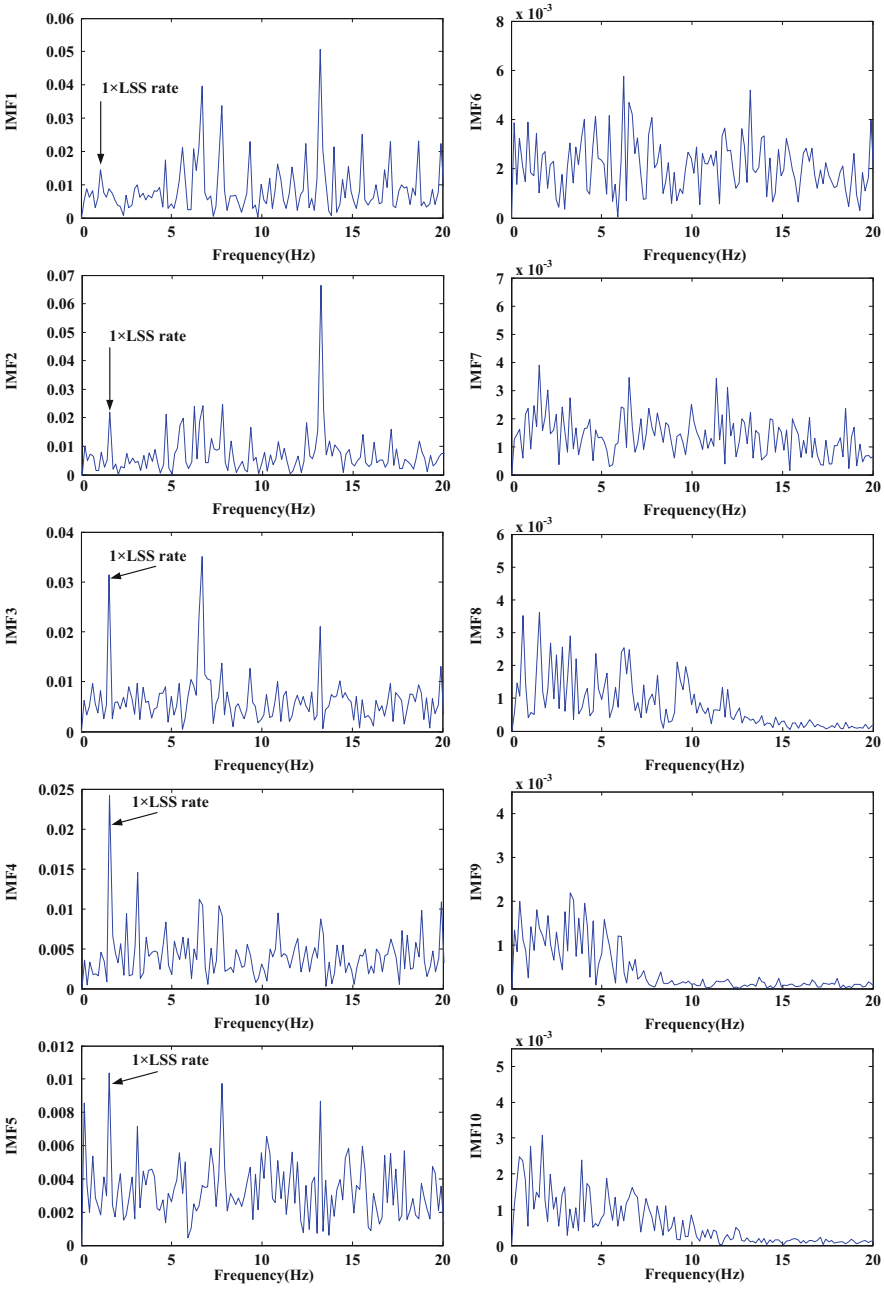


Fig. 5 Frequency spectrum of different IMF of gear wear state

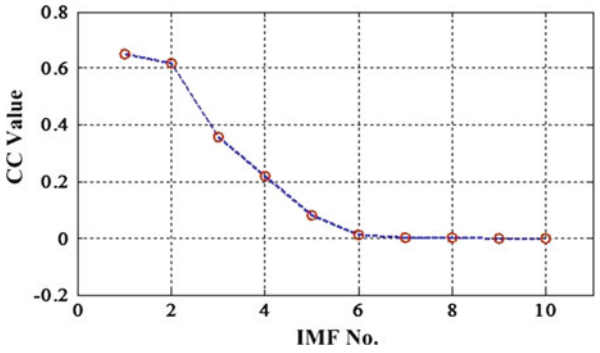


Fig. 6 CC values of ten IMFs

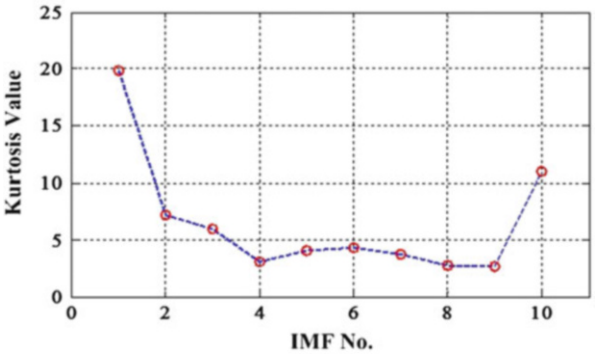


Fig. 7 Kurtosis values of ten IMFs

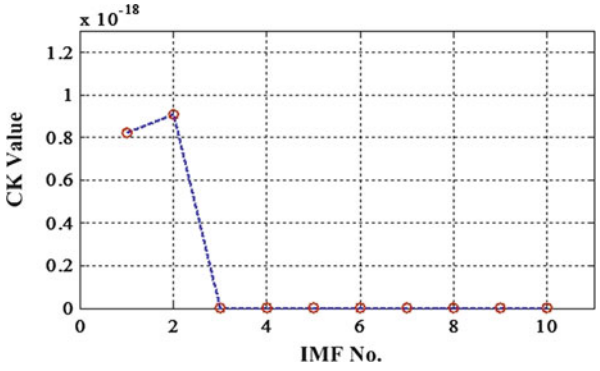


Fig. 8 CK values of ten IMFs

gear wheel rotational frequency and its second and third harmonics in the envelope spectrum can also be obtained. Figure 9 represents the normalized RMS and energy values. It shows that gear wheel rotational frequency energy is the best indicator to track the gear degradation, because they have systematic increasing trend compared with RMS values of other two types.

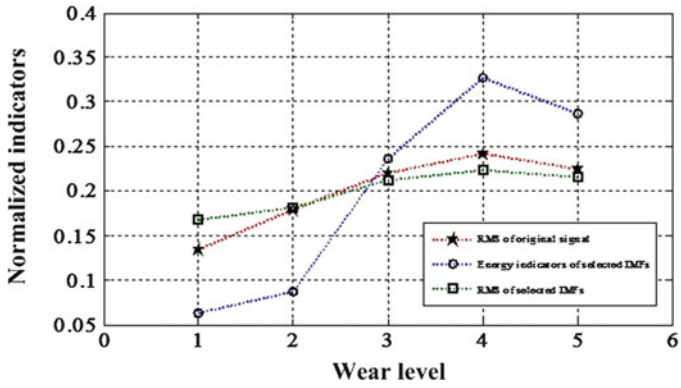


Fig. 9 Degradation indicators comparison results

6 Conclusions

This paper proposes a hybrid gear and bearing fault diagnosis method based on three indicators, namely, time synchronous technique, fast EMD and correlated kurtosis. To demonstrate the effectiveness in real applications, the method is tested on the gear wear fault identification of fixed-axis gearbox; and on the bearing faults in drive train gearbox of wind turbines. Fast EMD algorithm can decompose a long signal into several IMFs in real time manner, enabling us to obtain high resolution frequency spectrum. Correlated kurtosis can determine the best IMF used for post processing. The length of time synchronous averaging is very short, which guarantees a single period of TSA data. Therefore, the averaging operation is not required; instead it only resamples the signal by the tachometer through equal angle-interval waveform. This ensures that the signal is long enough for calculating the CK value. After time synchronous processing, the signal can be decomposed into ten IMFs using Fast EMD. Finally, gear and bearing fault data of test rigs and field wind turbines are used to verify the effectiveness of proposed method. The results show the rotational frequency energy with first and second harmonic frequency outperforms the root mean square for degradation tracking. In the future the proposed method will be extended to industrial applications where machines or systems are subject to random shocks or vibration signals of certain components are available.

References

1. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH (1998) The empirical mode decomposition and the Hilbert spectrum for non-linear and non-stationary time series analysis. Proc R Soc London Ser A 454:903–995
2. Mandic D (2011) Filter bank property of multivariate empirical mode decomposition. IEEE Trans Signal Process 59:2421–2426

3. Wu Z, Huang NE (2009) Ensemble empirical mode decomposition: a noise assisted data analysis method. Adv Adapt Data Anal 1:1–41
4. Wu TY, Hong HC, Chung YL (2011) A looseness identification approach for rotating machinery based on post-processing of ensemble empirical mode decomposition and autoregressive modeling. J Vib Control 18:796–807
5. Ibrahim GR, Albarbar A (2011) Comparison between Wigner-Ville distribution and empirical mode decomposition vibration-based techniques for helical gearbox monitoring. Proc IME C J Mech Eng Sci 225:1833–1846
6. Guo W, Tse PW, Djordjevich A (2012) Faulty bearing signal recovery from large noise using a hybrid method based on spectral kurtosis and ensemble empirical mode decomposition. Measurement 45:1308–1322
7. Lei YG, Lin J, He ZJ, Zuo MJ (2013) A review on empirical mode decomposition in fault diagnosis of rotating machinery. Mech Syst Signal Process 35:108–126
8. Feng ZP, Liang M, Zhang Y, Hou SM (2012) Fault diagnosis for wind turbine planetary gearboxes via demodulation analysis based on ensemble empirical mode decomposition and energy separation. Renew Energy 47:112–126
9. Zheng JD, Cheng JS, Yang Y (2013) Generalized empirical mode decomposition and its applications to rolling element bearing fault diagnosis. Mech Syst Signal Process 40:136–153
10. Li M, Li FC, Jing BB, Bai HY, Li HG, Meng G (2013) Multi-fault diagnosis of rotor system based on differential based empirical mode decomposition. J Vib Control 21(9):1821–1837
11. Dybała J, Zimroz R (2014) Rolling bearing diagnosing method based on empirical mode decomposition of machine vibration signal. Appl Acoust 77:195–203
12. Guo W, Tse PW (2013) A novel signal compression method based on optimal ensemble empirical mode decomposition for bearing vibration signals. J Sound Vib 332:423–441
13. Georgoulas G, Loutas T, Stylios CD, Kostopoulos V (2013) Bearing fault detection based on hybrid ensemble detector and empirical mode decomposition. Mech Syst Signal Process 41:510–525
14. Wang YH, Yeh CH, Young HWV, Hu K, Lo MT (2014) On the computational complexity of the empirical mode decomposition algorithm. Phys A 400:159–167
15. McDonald GL, Zhao Q, Zuo MJ (2012) Maximum correlated Kurtosis deconvolution and application on gear tooth chip fault detection. Mech Syst Signal Process 33:237–255
16. Bechhoefer E, Hecke BV, He D (2013) Processing for improved spectral analysis. In: Annual conference of prognostics and health management society

Bulge Deformation in the Narrow Side of the Slab During Continuous Casting

Qin Qin, Zhenglin Yang, Mingliang Tian, and Jinmiao Pu

Abstract Bulge deformation is one of the key factors that affect the slab quality in continuous casting involving heat transfer, phase transition and thermo-mechanical coupling problem. A systematical investigation into the bulge deformation rules of the narrow side was provided in the present study by using ABAQUS; And a three-dimensional thermo-mechanical coupling model has been established with taking into account the dynamic contact between slab and rolls. The result reveals that the cumulative bulge deformation without the constraint from the roller supporting is greater than the recovery count which compensates deformation in the wide side. The influences of various casting process parameters on bulge deformation of the slab have been also investigated in this paper. The consequence has been obtained that bulge deformation in the narrow side increases with the increase of casting speed and roll pitch; And the method of fixes-gap and variable-diameter was suggested to reduce the bulge deformation for manufacturing.

Keywords Slab quality • The narrow side • Thermo-mechanical coupled • Casting speed • Roll pitch

1 Introduction

Bulge deformation of the slab affects not only the quality of the billet, but also the stability of the production. However, continuous casting process involves heat transfer, phase transition and thermo-mechanical coupling problem. Therefore, the analysis and measurement of bulge deformation is a very difficult issue to be solved. In recent years, analytical methods and finite element methods have been generally applied to investigate bulge deformation of the slab. In the early stages of the bulging research, analytical methods were initially used due to their convenience. In these methods, casting slabs were usually simplified to continuous beam models or plate models according to the theory of bending beam or plate theory [1, 2]. However the

Q. Qin (✉) • Z. Yang • M. Tian • J. Pu
University of Science and Technology Beijing, Beijing, China
e-mail: qinqin@me.ustb.edu.cn; zhenglinyang@hotmail.com; tml817817@163.com; pujinmiao@126.com

M.J. Zuo et al. (eds.), *Engineering Asset Management 2016*, Lecture Notes in Mechanical Engineering, DOI 10.1007/978-3-319-62274-3_33

calculation accuracy of analytical methods is relatively low because there are some differences between the actual conditions and the oversimplified models. Therefore the finite element methods have been widely applied to study the bulging problems. Two-dimensional models were firstly adopted by the majority of previous researches [3, 4]. However, bulge deformation of the narrow side in these models cannot be investigated. Therefore, some three-dimensional models of bulge deformation have been developed by some researchers for improving the calculation accuracy. K. Okamura et al. have applied a three-dimensional elasto-plastic and creep model to investigate the effect of the narrow face shell on restraining the bulging deflection [5]. K. Feng et al. have established a three-dimensional coupling model and obtained the detailed external and internal distributions of stress and strain for slab [6]. Unfortunately, these models were based on the static contacts between the slab and the rolls. Q. Qin et al. have established a three-dimensional thermo-mechanical coupling model which was based on the dynamic contact between the slab and rolls [7]. But this model was merely used to discuss the differences between the 3D and 2D models in calculating the bulge deformation. The deformation distributions of the narrow side have not been analyzed in detail.

The objective of this paper is to provide a systematical investigation into the bulge deformation rules of the narrow side by using ABAQUS, and has established a three-dimensional thermo-mechanical coupling model with considering the dynamic contact between slab and rolls. Moreover, the influences of casting parameters on bulge deformation in the narrow side have also been discussed to design for manufacturing.

2 Establishment of Bulging Model

Bulge analysis involves the highly non-linear behaviors including multiple objects, complex thermo-mechanical coupling problems. Some simplified treatments were adopted in this model and shown as follows:

- A quarter bulging model was built due to the symmetry of heat transfer and model structure.
- Both the driving rolls and the driven rolls were assumed to be rigid bodies.
- The pressure was assumed to apply outward at the surface of the quarter bulging model.

Continuous casting is a high-temperature process, and material properties of the slab are sensitive to the temperature. Young's modulus and Poisson's ratio of Q235 steel were listed in Table 1, and the coefficient of thermal expansion of Q235 steel is shown in Fig. 1. Moreover, creep behavior of the slab was taken into consideration besides mechanics behavior in this paper. And the time hardening creep model was adopted to describe the viscoelastic behavior of the slab as Eq. (1) [7]:

Table 1 Young's modulus and Poisson's ratio of Q235 steel

Temperature (°C)	Young's modulus (Mpa)	Poisson's ratio
700	8422.4	0.336
800	7319.6	0.344
900	6005.9	0.352
1000	5055.2	0.360
1100	4367.4	0.369
1200	4215.5	0.377
1300	3021.3	0.385
1400	668.2	0.393

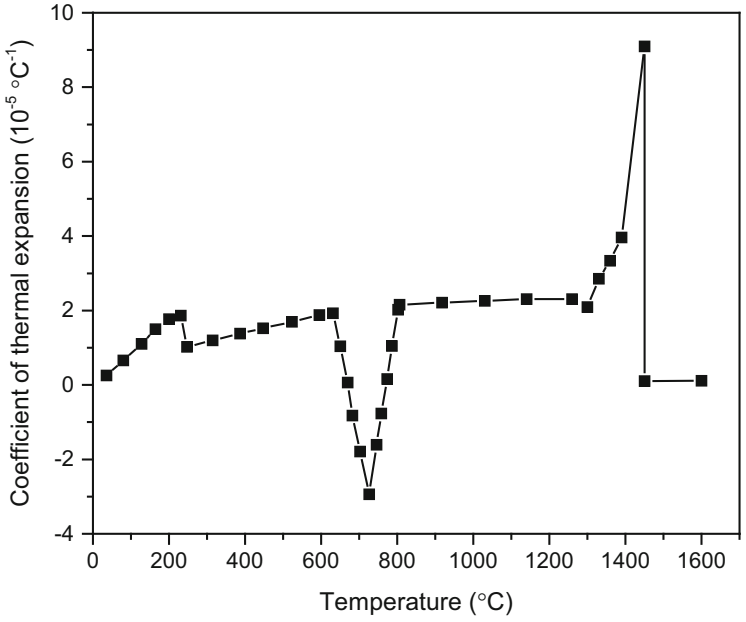


Fig. 1 The coefficient of thermal expansion

$$\dot{\varepsilon} = C\exp(-\mathrm{Q}/\mathrm{T})\sigma^n t^m \tag{1}$$

In order to fit the actual situation, boundary conditions and contact definition were taken into consideration as follows and the three-dimensional dynamic finite element model involving the slab and the rolls:

- The symmetrical displacement constraint was applied to the symmetrical surface of the slab. And both ends of the slab were controlled with translate and rotate constraint.
- The pressure caused by the liquid core was assumed to apply as the uniform force increasing linearly with distance below the liquid steel meniscus.

- The penalty friction model was applied to the dynamic contact between the slab and the rolls. The friction coefficient of the driving rolls was considered as 0.3 and for the driven rolls was 0.001 according to the conversion of friction coefficient of the ball bearing.
- The driving rolls were applied with the angular velocity and the driven rolls were applied with constraints.

3 Bulging Analysis of the Narrow Side

Generally, bulge deformation of the wide side has been widely discussed, but that of the narrow side has been rarely investigated. Based on the establishment of the dynamic bulging model, two typical nodes were chosen to analyze the characteristics of the bulge deformation in the narrow side. As shown in Fig. 2, nodes A1 is directly over the rolls while nodes B1 is at the center of two successive rolls. With the results of the bulging simulation, the horizontal displacement curves along with time of the nodes show the same trend of decline (Fig. 3). Under the pressure caused by the liquid core, bulge deformation in the narrow side develops more freely without the constraint from the roller supporting.

The deformation increase in the narrow side is greater than the count of the recovery which compensates bulge deformation of the wide side. As a result, bulge deformation in the narrow side is cumulative. Moreover, the horizontal displacement situations of two nodes are on the contrary because they differ half of a roll pitch. In the first several casting cycles, bulge deformation of the narrow side is in the unstable state, and then becomes stable with the increase of distance along the casting direction. The difference between values of maximum and minimum deformation approaches certain in the steady deformation zone.

Figure 4 shows the difference between values of maximum and minimum deformation to accurately calculate deformation accumulation of the narrow side. The deformation cycle and the residual deformation have been marked as T and

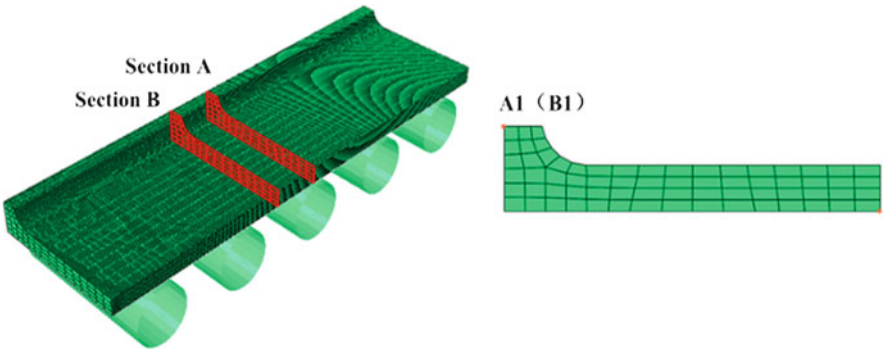


Fig. 2 Critical nodes in the narrow side

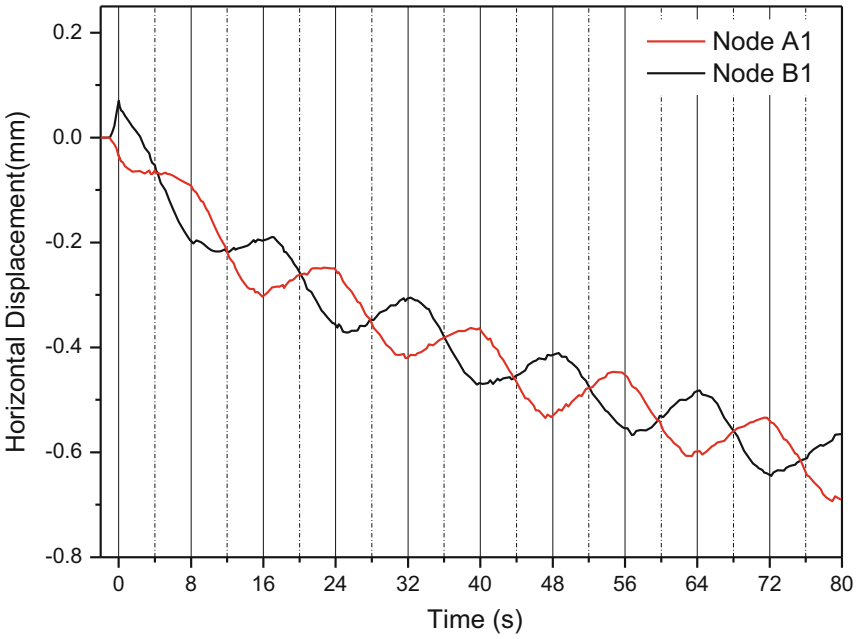


Fig. 3 Horizontal displacement curves along with time of two nodes

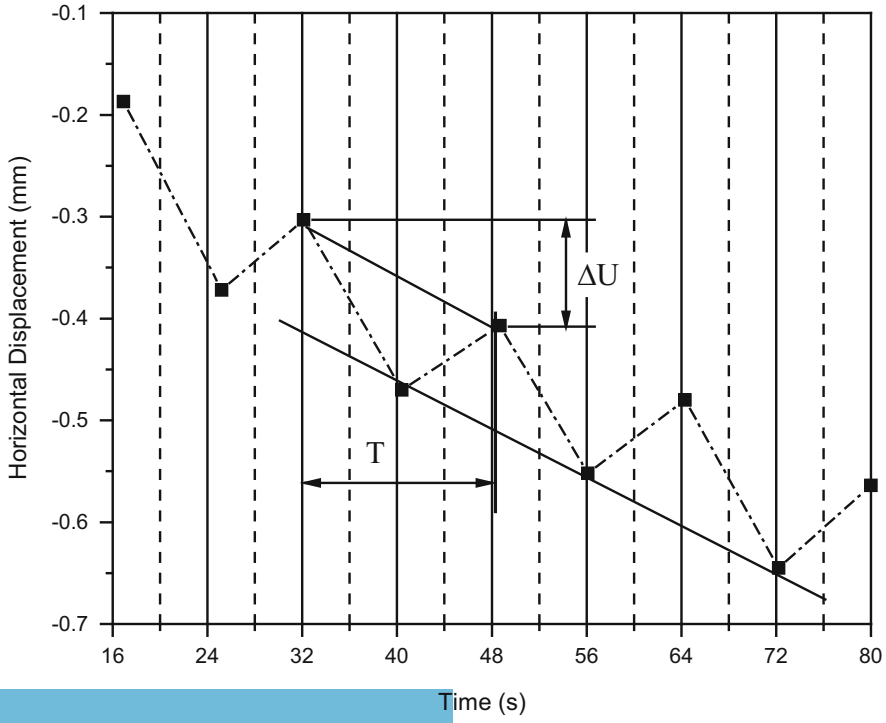


Fig. 4 Bulge deformation curve along with time in the narrow side

Table 2 Bluge accumulation varies with time

Times (s)	Bulge accumulation (mm)
16	0.07
32	0.14
48	0.211
64	0.281
80	0.351

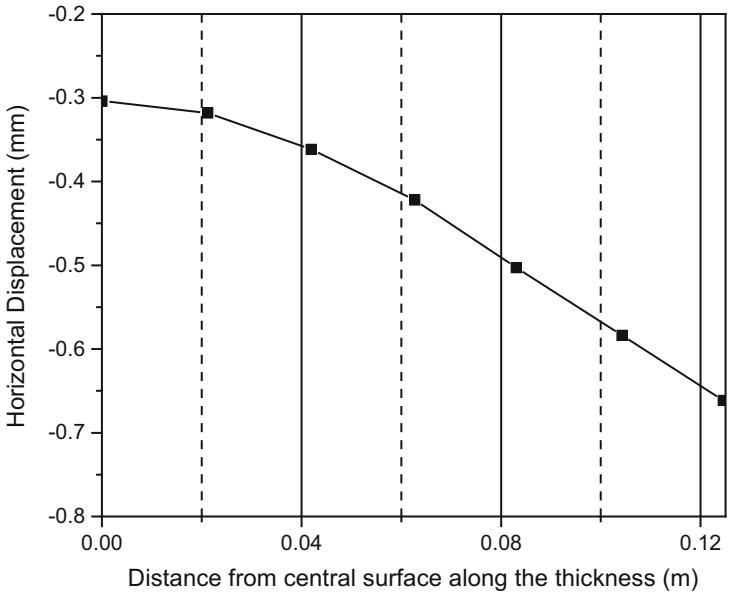


Fig. 5 Bulge deformation of the narrow side along the thickness direction

ΔU, respectively. The absolute value of the slope of the boundary was found to be $|K| = 4.39 \times 10^{-3}$, and then bulge accumulation of the narrow side was calculated as listed in Table 2.

Bulge deformation of the narrow side along the thickness direction was not uniform as shown in Fig. 5. Resultly, bulge deformation of the outer node in narrow side is 0.661 mm when the bulge deformation of the inner node is 0.306 mm, which accounts for 46.3% of that of the outer side. Figure 6 shows the displacement curves of the narrow side that are various with the increase of distance along the thickness direction. However, the relative positions corresponding to the displacement amplitudes of the inner and outer nodes in narrow side are on the contrary, where bulge deformation of the inner side is minimum when that of the outer side is maximum. Generally, bulge deformation of the narrow side develops more freely than the wide sides and is not uniform along the thickness direction.

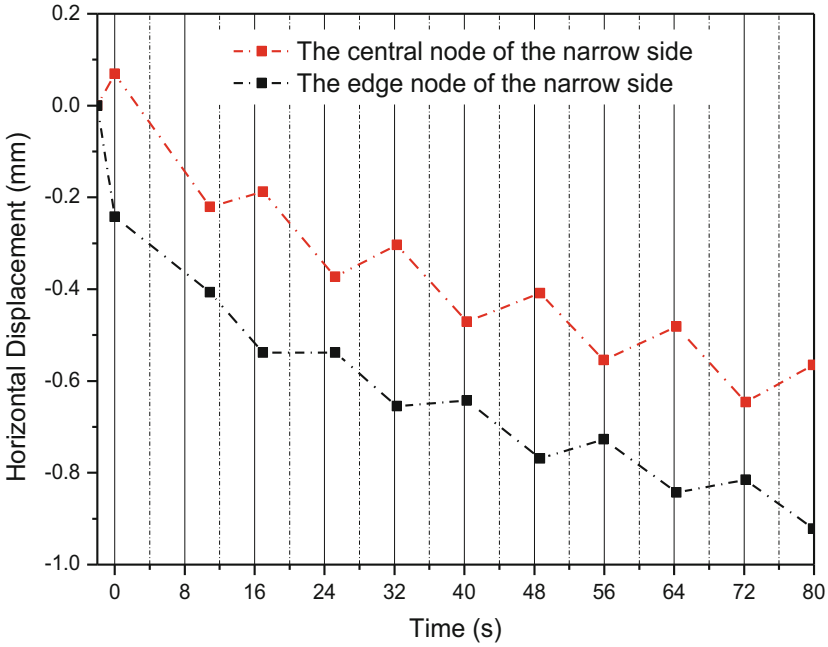


Fig. 6 Relative Positions corresponding to the displacement amplitudes in the narrow side

4 Influences of Casting Processing Parameters

As mentioned in the introduction, many influence parameters, such as roll pitch, the ferrostatic pressure, temperature fields of the slab, the thickness of the solidified shell and casting speed, are involved in the production conditions of continuous casting. However, casting processing parameters are often coupled with each other. For instance, the ferrostatic pressure, temperature fields and thickness of the slab have been found to be affected by casting speed. Therefore, their effects on bulge deformation could be taken into overall account for the influence of casting speed. Consequently, the influences of casting speed and roll pitch on bulge deformation were respectively investigated for manufacturing in this study.

The influence of casting speed was investigated under various casting speeds and constant roll pitch. The bulge deformation was obtained by using the dynamic bulging models. Figure 7 shows the bulge deformation of the slab under various casting speeds. Generally, bulge deformation of the narrow side increases about 0.23 mm with the increase each 0.1 m/min of casting speed. That is mainly due to an increase in temperature fields of the slab induced by the increase of casting speed, which consequently leads to the increase of the bulge deformation.

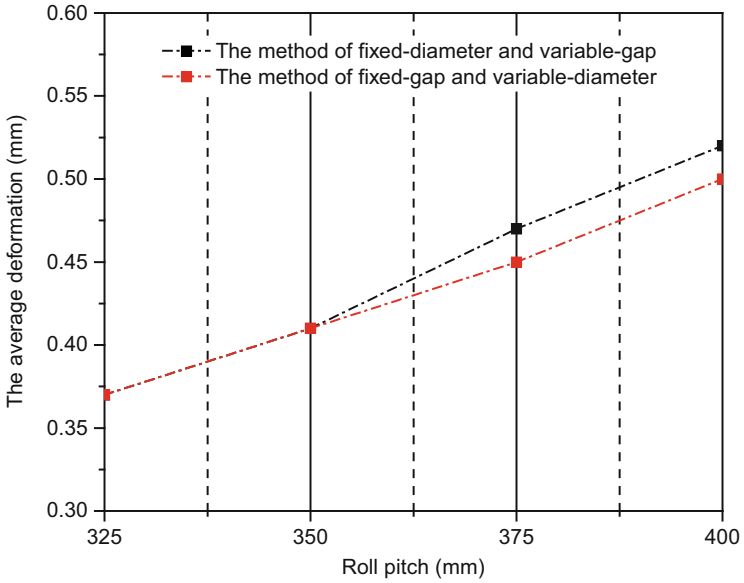


Fig. 7 Bulge deformation under various casting speeds

Similarly, the influence of roll pitch was conducted by using the dynamic bulging models under various roll pitches and constant casting speed. In this study, two methods were adopted to change the size of roll pitch. One method is fixed-gap variable-diameter, where roll gap is kept a constant value of $L_0 = 70$ mm and roll diameter is variable. Constrastly, another method is fixed-diameter and variable-gap that roll diameter is kept a constant value of $\Phi_0 = 230$ mm and roll gap is variable. Bulge deformation curves of the narrow side under various roll pitches are shown in Fig. 8. By using two methods, bulge deformation of the narrow side respectively increases about 0.037 mm and 0.043 mm when roll pitch increases each 25 mm. The increase of roll pitch directly induces an increase in creep time of the slab, which consequently results in the increase of the bulge deformation. Additionally, bulge deformation simulated with the method of fixes-gap and variable-diameter is smaller than that with the method of fixed-diameter and variable-gap. That is a result of the increase of the contact area between the slab and rolls caused by the larger roll diameter of fixed-gap and variable-diameter method, which then indirectly increases the stiffness of the slab.

In generally, bulge deformation of the narrow side increases with growth of casting speed and roll pitch. In addition, the method of fixes-gap and variable-diameter is better than the method of fixed-diameter and variable-gap to reduce the bulge deformation.

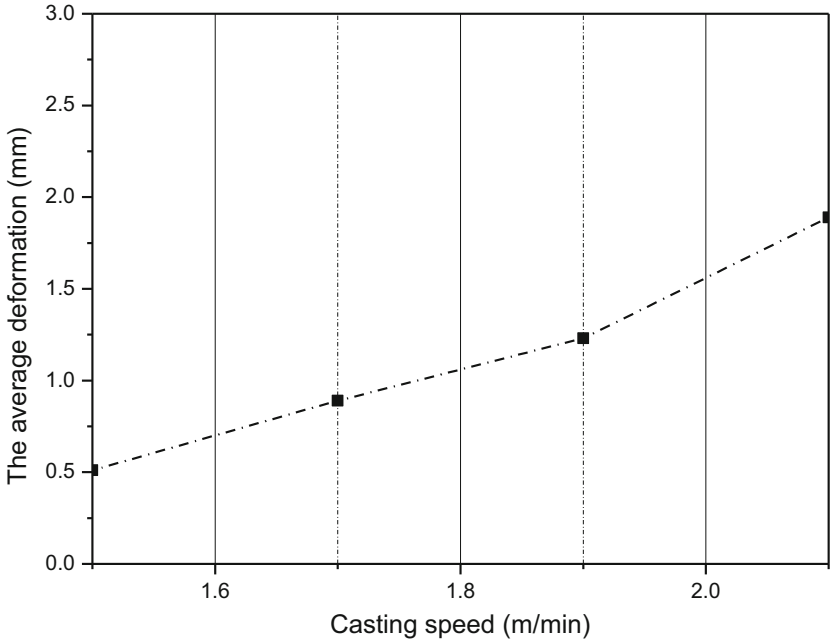


Fig. 8 Bulge deformations under various casting speeds

5 Summary

Bulge deformation of the narrow side is cumulative because the deformation increase of the narrow side is greater than the count of recovery which compensates bulge deformation of the wide side.

Bulge deformation of the narrow side is not uniform along the thickness direction. Bulge deformation of the outer node in narrow side is 0.661 mm when the bulge deformation of the inner node is 0.306 mm, which accounts for 46.3% of that of the outer side.

Bulge deformation of the narrow side increases about 0.23 mm when casting speed increases each 0.1 m/min.

By using two methods to change the size of roll pitch, bulge deformation of the narrow side respectively increases about 0.037 mm and 0.043 mm when roll pitch increases each 25 mm. The approach of fixes-gap and variable-diameter has been suggested to reduce the bugle deformation for manufacturing.

Acknowledgement This research was supported by National Natural Science Foundation of China (51375041).

References

1. Jiquan S, Yipin S, Xingzhon Z (1996) Analysis of bulging deformation and stress in continuous cast slabs. J Iron Steel Res Int 8:11–15
2. Yoshii A, Kihar S (1986) Analysis of bulging in continuously cast slabs by bending theory of continuous beam. Trans Iron Steel Inst Jpn 26:891–894
3. Janik M, Dyja H, Berski S, Banaszek G (2004) Two-dimensional thermomechanical analysis of continuous casting process. J Mater Process Technol 153:578–582
4. Koric S, Thomas BG (2007) Thermo-mechanical models of steel solidification based on two elastic visco-plastic constitutive laws. J Mater Process Technol 197:408–418
5. Okamura K, Kawashima H (1989) Three-dimensional elastoplastic and creep analysis of bulging in continuously cast slabs. ISIJ Int 29:666–672
6. Feng K, Chen D, Xu C et al (2004) Effect of main thermo-physical parameters of steel Q235 on accuracy of casting transport model. Spec Steel 25:28–31
7. Qin Q, Shang S, Wu DP, Zang Y (2014) Comparative analysis of bulge deformation between 2D and 3D finite element models. Mech Eng 154:14–17