Springer Series in Reliability Engineering

Yu Liu
Hong-Zhong Huang
Tao Jiang

Selective Maintenance Modelling and Optimization

Basic Methods and Some Recent Advances

Springer

Springer Series in Reliability Engineering

Series Editor

Hoang Pham, Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ, USA

Today's modern systems have become increasingly complex to design and build, while the demand for reliability and cost effective development continues. Reliability is one of the most important attributes in all these systems, including aerospace applications, real-time control, medical applications, defense systems, human decision-making, and home-security products. Growing international competition has increased the need for all designers, managers, practitioners, scientists and engineers to ensure a level of reliability of their product before release at the lowest cost. The interest in reliability has been growing in recent years and this trend will continue during the next decade and beyond.

The Springer Series in Reliability Engineering publishes books, monographs and edited volumes in important areas of current theoretical research development in reliability and in areas that attempt to bridge the gap between theory and application in areas of interest to practitioners in industry, laboratories, business, and government.

Now with 100 volumes!

Indexed in Scopus and EI Compendex

Interested authors should contact the series editor, Hoang Pham, Department of Industrial and Systems Engineering, Rutgers University, Piscataway, NJ 08854, USA. Email: hopham@rci.rutgers.edu, or Anthony Doyle, Executive Editor, Springer, London. Email: anthony.doyle@springer.com.

Yu Liu · Hong-Zhong Huang · Tao Jiang

Selective Maintenance Modelling and Optimization

Basic Methods and Some Recent Advances

Springer

Yu Liu
Center for System Reliability and Safety
University of Electronic Science
and Technology of China
Chengdu, Sichuan, China

Hong-Zhong Huang
Center for System Reliability and Safety
University of Electronic Science
and Technology of China
Chengdu, Sichuan, China

Tao Jiang
Center for System Reliability and Safety
University of Electronic Science
and Technology of China
Chengdu, Sichuan, China

ISSN 1614-7839 ISSN 2196-999X (electronic)
Springer Series in Reliability Engineering
ISBN 978-3-031-17322-6 ISBN 978-3-031-17323-3 (eBook)
https://doi.org/10.1007/978-3-031-17323-3

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

Preface

Reliability has become a crucial characteristic of advance engineered systems as systems with poor reliability suffer from a great amount of lifecycle cost and potential risk of failures. Maintenance, involving both corrective and preventive actions, is an effective way to retain a system in or restore it to an acceptable operating condition, and it has been extensively implemented in industrial applications. Examples of maintenance activities for engineered systems are oil change of rotating systems, rotor balance of mechanical systems, shaft/coupling alignment, filter replacement, corroded components coating, and so forth. It is noted that inappropriate maintenance scheme may not guarantee the reduction of operation cost and the fulfillment of reliability target, and systems are over-maintained or under-maintained during its operation stage. The maintenance models and optimization have been, therefore, intensively studied in the past decade with the purpose of minimizing maintenance cost and/or maximizing reliability or availability of a specific system. The paradigm of maintenance strategy has shifted from corrective maintenance to preventive maintenance, and then to condition-based maintenance and predictive maintenance nowadays.

Selective maintenance optimization, as a specific condition-based maintenance problem, was firstly presented by Rice, Cassady, and Nachlas in the 7th Industrial Engineering Research Conference. In selective maintenance optimization, a system intends to perform successive assigned missions with a break between two adjacent missions, and maintenance actions can be executed in breaks to ensure the success of the subsequent missions. However, due to the limited maintenance resources, such as time and budget, it is impossible to carry out all the desired maintenance actions for aged and failed components. Alternatively, a subset of maintenance actions has to be selected from the set of all the optional maintenance actions, so as to maximize the success of the future missions. The preliminary model for selective maintenance of series–parallel systems with independent and identical copies of a component was generalized to a basic framework by Cassady and co-authors in the follow-up research works. As selective maintenance optimization exactly matches up with the many industrial and military scenarios where only a limited amount of maintenance resources have to be allocated among components of a system, it has been extensively studied in the past decade from various angles and implemented in a diversity of industrial applications. Research articles on this subject are continuously being published in journals and conference proceedings. Nevertheless, to the best of our knowledge, the subject has never been adequately or systematically reported in reliability book. The increased and sustained interest in this subject drives us to publish this book.

This book systematically introduces the basic selective maintenance optimization model. It is, to a large extent, a collection of our recent research advances on selective maintenance optimization from the Center for System Reliability and Safety at the University of Electronic Science and Technology of China. The layout of this book is as following:

Chapter 1 introduces the role of maintenance optimization in lifecycle management of engineering assets and gives an overview picture of research topics in maintenance optimization. It is followed by a systematical literature review on the existing research efforts on selective maintenance optimization.

Chapter 2 offers an introduction to the basic mathematical model of selective maintenance problem, in which both a system and its components are assumed to be binary-state. Three selective maintenance models with distinct objectives functions and constraints are formulated.

Chapter 3 discusses the selective maintenance optimization for multi-state systems with binary-capacitated components. The Kijima type II age reduction model, serving as a specific imperfect maintenance model, is incorporated in the selective maintenance optimization. The universal generating function is utilized to evaluate the probability of a system successfully completing the next mission. The genetic algorithm is introduced to resolve the resulting optimization model.

Chapter 4 focuses on the selective maintenance optimization for multi-state systems with the load sharing mechanism. A joint optimization model is formulated to simultaneously optimize the load distribution and the allocation of the limited maintenance budget among components. The genetic algorithm is employed to solve the optimization problem.

Chapter 5 discusses selective maintenance optimization under stochastic time durations of breaks and maintenance actions. The distribution of the number of completed maintenance actions in a break with time duration uncertainty is evaluated by using the saddlepoint approximation. A tailored ant colony optimization algorithm is developed to solve the resulting combinational optimization problem in the cases of large-scale systems.

Chapter 6 presents a robust selective maintenance optimization model to treat the uncertainty produced by imperfect observations. A multi-objective optimization model is formulated with the aims of maximizing the expectation and simultaneously minimizing the variance of a system successfully completing the next mission.

Chapter 7 takes account of uncertainties from both maintenance and inspection and introduces a joint selective maintenance and inspection optimization model. A finite-horizon mixed observability Markov decision process is formulated as the remaining resource is fully observable and the component states are partially observable. The dynamic programing and the deep reinforcement learning algorithm are implemented to resolve small-scale problems and large-scale problems, respectively.

Chapter 8 discusses selective maintenance optimization for systems executing multiple consecutive missions. The uncertainties associated with the time duration of each future mission and the working time of each component in each future mission are addressed. The selective maintenance problem is formulated as a max-min optimization model, and it is resolved by a customized simulated annealing-based genetic algorithm.

Chapter 9 introduces a dynamic selective maintenance for multi-state systems operating multiple consecutive missions. The resulting sequential decision problem is formulated as a Markov decision process with a mixed integer-discrete-continuous state space. A deep reinforcement learning method is customized based on the actor-critic framework, and a postprocess is utilized to search for the optimal maintenance actions in a constrained large-scale action space.

The target audience of this book is undergraduate and graduate students, reliability practitioners, and researchers. The readers should have background in basic probability theory, stochastic models, and optimization algorithms. The book offers a great mount of knowledge and insights on system maintenance modelling methods and optimization algorithms, with which readers can deal with many real-world engineering cases.

This book is a collection of materials developed in the dissertations and journal/conference papers of several former and current graduate students from the Center for System Reliability and Safety at the University of Electronic Science and Technology of China. The majority of the chapters have been developed based on the dissertations and research works of Dr. Tao Jiang (M.Sc. student from 2014–2017 and Ph.D. student from 2017–2022), Dr. Yiming Chen (Ph.D. student from 2015–2022), Mr. Jian Gao (M.Sc. student from 2019–2022), Mr. Chujie Chen (M.Sc. student from 2012–2015), and Mr. Qin Zhang (M.Sc. student from 2020–2022). The book was edited with the additional assistance of Dr. Tangfan Xiahou (Ph.D. student from 2018–2022) and Dr. Mingang Yin (Post-doctoral research fellow from 2021–2023). We would like to express our sincere gratitude and appreciation to researchers and friends who have discussed with the concepts and models of this book, or have co-authored with us on some topics of this book. To name a few, Prof. Ming J. Zuo at University of Alberta, Prof. Wei Chen at Northwestern University, Prof. Min Xie at City University of Hong Kong, Prof. Lirong Cui at Qingdao University, Prof. Liudong Xing at University of Massachusetts-Dartmouth, Dr. Gregory Levitin at Israel Electric Corporation, Prof. Yi-Kuei Lin at Taiwan Yang Ming Chiao Tung, Prof. Tongdan Jin at Texas State University, Prof. Haitao Liao at University of Arkansas, Prof. Zhisheng Ye at National University of Singapore, Prof. Zhiguo Zeng at Centrale-Supélec—Université Paris-Saclay, Prof. Yisha Xiang at Texas Tech University, Prof. Yuchang Mo at Huaqiao University, and Prof. Hui Xiao at Southwestern University of Finance and Economics.

Last but not least, we would like to thank Prof. Hoang Pham at Rutgers University, who gave a great support to the publication of this book. It is also indeed our pleasure working with Mr. Kavitha Sathish and the Springer editorial team.

The research works in this book received financial support from the National Natural Science Foundation of China under contact numbers 71922006 and 71771039.

Chengdu, Sichuan, China
June 2022

Yu Liu
Hong-Zhong Huang
Tao Jiang

Contents

Chapter 1
Introduction

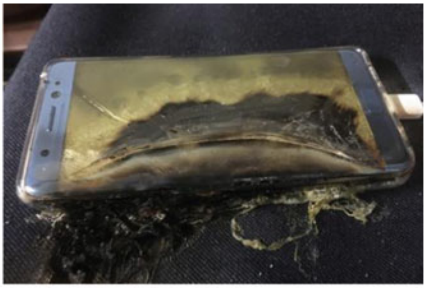
1.1 Overview of Maintenance Optimization

Reliability refers to the ability of a system or component to perform its intended functions under the stated conditions for a specified period of time [36]. Advanced engineered systems are designed for greater size, higher precision, and more complex functionality, along with the involvement of state-of-the-art artificial intelligent techniques. The reliability of these systems (e.g., nuclear plants, wind turbines, aircrafts, power systems, and machining centers) has been intensively studied across their entire lifecycle. Failure of these advanced systems often results in unexpected production delays and/or significant economic losses and can even cause severe threats to human life. Examples of recent major accidents, as shown in Fig. 1.1, include the Fukushima Daiichi nuclear power plant disaster in 2011, Samsung Galaxy Note 7 battery explosion in 2016, Italy Apulia train crash in 2016, and Boeing 737 Max crash in 2019. However, the causes of failures are diverse, ranging from the system design and manufacturing stages to the system operation and maintenance stages [121].

Industrial practitioners strive to minimize the occurrence and consequences of failures. Consequently, many attempts have been made by both industry and academia in the past decades to understand why and with which patterns system failures occur [143]. Figure 1.2 shows a typical pattern of the hazard rate function of engineered systems, which is well-known as a bathtub curve because of its shape. Systems with such a hazard rate function experience a decreasing failure rate at the early stage of their lifecycle (also called infant mortality), followed by a nearly constant failure rate stage (also called useful life) and by an increasing failure rate stage (also called wear-out). In the infant mortality region, the failure rate of systems is high because of defective components, manufacturing defects, and poor quality control. By removing defective components, the failure rate function of the system will continuously fall, and it will reach the useful life region. The failure rate function in the useful life region is fairly constant, and system failures are caused randomly by environmental loads, human errors, and chance events. Engineered systems are expected to operate for as long as

Y. Liu et al., *Selective Maintenance Modelling and Optimization*,
Springer Series in Reliability Engineering,
https://doi.org/10.1007/978-3-031-17323-3_1

(a) Fukushima nuclear plant disaster [89]



(b) Galaxy Note 7 battery explosion [35]

(c) Italy Apulia train crash [7]

(d) Boeing 737 Max crash [98]

Fig. 1.1 Examples of major accidents in recent years

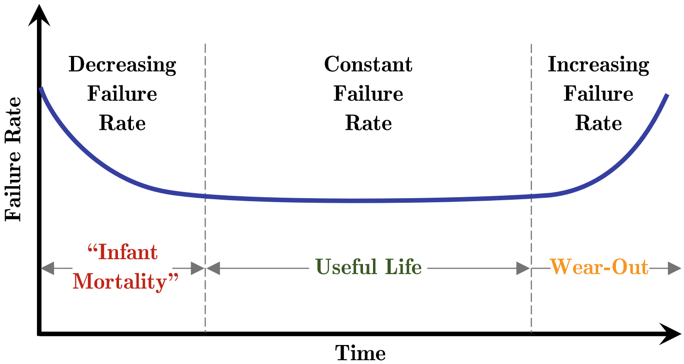


Fig. 1.2 Bathtub curve

possible in this region. Eventually, the failure rate function of the systems rises again in the wear-out region as the components in the systems start to deteriorate. Typical causes of component deteriorations include fatigue, corrosion, friction, and aging. By replacing or recovering the wear-out components in a timely manner, the failure rate function of the system will decline immediately, leading to a lower probability of failure in the future than that of the case without replacement or recovery. Such a system that can be repaired in its lifecycle is called a repairable system.

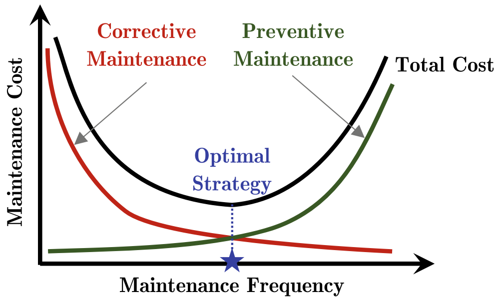


Fig. 1.3 Maintenance decision-making

Component replacement is a typical maintenance activity in industrial practice. Further, all activities aimed at maintaining a system in or restoring it to the physical state considered necessary for the fulfillment of its production function can be regarded as maintenance activities [43]. Examples of maintenance activities for engineered systems include oil change in rotating systems, rotor balance of mechanical systems, shaft/coupling alignment, filter replacement, and coating of corroded components. Despite that maintenance activities can prolong the usage lifetime and maintain the performance capacity of systems, unoptimized maintenance planning may not guarantee a reduction in operation costs and the fulfillment of reliability targets, and systems are over-maintained or under-maintained during the operation stage. By constructing a proper maintenance decision model, the maintenance plan for engineered systems can be optimized to minimize maintenance costs and/or maximize the reliability or availability of a specific system. An example of maintenance decision making is delineated in Fig. 1.3. With an increase in the preventive maintenance (PM) frequency, the cost associated with corrective maintenance (CM) declines, whereas that of PM increases. From the perspective of the total maintenance cost, an optimal PM frequency that possesses a minimal total maintenance cost is readily found. Such an optimal maintenance strategy can be identified by formulating a maintenance decision model and resolving it using an appropriate optimization algorithm.

1.1.1 Paradigms of Maintenance Optimization

Looking back on the history of maintenance optimization, the paradigm of maintenance strategy has shifted from CM to PM and then to condition-based and predictive maintenance. The following are the definitions of these maintenance paradigms:

CM is also called reactive maintenance or run-to-failure strategy; that is, maintenance activities will be confined to reactive tasks of repair actions or component replacement upon the failure of components or systems. Examples of systems with CM strategies include light bulb replacement and computer repair. Non-essential items and systems in which failure causes marginal economic loss can be replaced or repaired when they run to failure. The advantages of CM include low cost associated with monitoring, minimal planning requirements, and simplicity of processing. Conversely, the disadvantages of CM are unpredictable failures and unscheduled downtime, shorter lifecycle, and potential for high long-term cost.

PM (also known as preventative maintenance) is conducted regularly on a population of identical engineered systems to reduce their likelihood of failure. It can be considered a proactive maintenance paradigm and further divided into two types: time-based PM and usage-based PM. In the former case, PM is executed at a fixed calendar time instant. For example, this could be a weekly or monthly maintenance routine for the production lines. In the latter case, PM is triggered after a set number of production cycles, hours in use, or even distance travelled. For example, the oil filter of a vehicle is periodically replaced every 10,000 mileages. Improved reliability, reduced cost associated with unexpected failures, and less disruption or unscheduled downtime are key features of PM. However, such a maintenance paradigm incurs additional costs and personnel for executing PM actions, and a poorly prepared PM plan may lead to a specific individual system from a population being over-maintained or under-maintained.

Condition-based maintenance (CBM) is also a proactive maintenance paradigm. Using advanced sensing techniques, the actual health status of engineered systems can be monitored to facilitate a timely and cost-efficient proactive maintenance plan for each individual system. CBM is triggered when certain monitored indicators show signs of decreasing performance or upcoming failure. The indicators could be the vibration signals of rotating machines, temperatures, and debris of lubricating oil. In specific engineering applications, condition monitoring data can be collected continuously or periodically at certain time intervals [49]. As CBM leverages the monitoring data of each individual system, unnecessary maintenance tasks can be avoided unlike either time-based or usage-based PM. Meanwhile, the repair cost and sudden downtime caused by random failures can be reduced further. However, equipping condition monitoring incurs additional upfront system setup costs. In most cases, these costs are offset by the potential costs of unnecessary maintenance or unexpected failures.

Predictive maintenance (PdM) is an emerging maintenance paradigm in recent years, and it is flourishing very fast with the development of predictive techniques. Slightly different from CBM, PdM not only monitors the actual health status of engineered systems using sensors or inspection instruments, but also predicts the evolution of the health status and remaining useful life of each individual system using historical monitoring data and future mission profiles [49]. Such a prediction activity that facilitates PdM is termed a prognosis in the reliability community [113]. Physical model-based and data-driven methods are two major research lines of prognosis. Recently, machine learning and artificial intelligence techniques, such as support vector machines, deep neural networks, and generative adversarial networks, have been intensively implemented as data-driven tools for prognosis [70, 144]. With progressively updated prognosis results, the PdM strategy is dynamically scheduled for each individual system. Because maintenance is only executed as required for each system when failure is imminent, PdM is often more cost-effective than the other maintenance paradigms. Similar to CBM, PdM requires condition monitoring devices for data acquisition. However, its effectiveness relies heavily on prognosis accuracy [8].

1.1.2 System Degradation Characteristics

The health status of an engineered system inevitably deteriorates over time. To schedule effective maintenance planning, decision makers should first mathematically characterize the deterioration behaviors of a system. Throughout the entire history of maintenance optimization, the most straightforward and simplest method of characterizing the failure behavior of a system is to assume an appropriate lifetime distribution (e.g., exponential and Weibull distributions). However, the lifetime distribution of a repairable system is based on a coarse probabilistic model. It is more suitable to consider an underlying stochastic process to manifest a set of stages that the system experienced before completely failing.

Owing to the development of advanced sensors and inspection techniques, the degradation trajectories of engineered systems can be recorded completely or partially. A plethora of degradation models have been developed over the past few decades to characterize deterioration behaviors in an effective and accurate manner. The degradation models of repairable systems can generally be categorized based on two categories of state spaces: discrete-state and continuous-state degradation models.

For engineered systems in which health status can be easily distinguished into different distinct states, discrete-state degradation models can be utilized. Discrete-state degradation models can be further categorized into binary-state, three-state, and generalized multi-state models. Essentially, the lifetime distribution of a repairable system is a binary-state degradation model (perfectly functioning and completely failed). The three-state degradation model herein refers to the delay-time model wherein the deterioration process of a system is defined as a two-stage process [125]. The first stage is the normal operating stage from the new stage to the point where a hidden defect has been identified, whereas the second stage is defined as the failure delay time from the time of defect identification to failure [125]. Moreover, many engineered systems can govern multiple (more than three) distinct states. For concreteness, power systems can operate at different performance capacities. Another typical example is that the health status of a cutting tool can be roughly classified into "normal," "moderately worn out," "seriously worn out," and "completely worn out." The Markov property is a conventional and basic hypothesis in multi-state degradation models. Markov models, such as discrete-time and continuous-time (homogenous) Markov chains, and semi-Markov models are effective tools for characterizing the multi-state degradation trajectories of degradation systems. Additionally, the hidden Markov and hidden semi-Markov models are implemented if the uncertainty associated with inspections is considered. Based on whether the system can transit from its current state to a non-adjacent state, multi-state degradation models can be distinguished into progressive and non-progressive models. The delay-time model can also be extended to a generalized multi-state degradation model by considering defect severity.

Typically, the deterioration of engineered systems is gradual over time and difficult to classify into multiple distinct states. In such cases, continuous-state degradation models should be used. Continuous-state degradation models in the context of maintenance optimization mainly include stochastic process-based models, typically Wiener, Gamma, and inverse Gaussian processes. The Wiener process is appropriate for characterizing non-monotonically increasing (or decreasing) degradation, such as the resistance of an electronic component and the capacity of batteries. Degradation in the form of cumulative damage, that is, monotonic increasing (or decreasing) degradation, can be modeled by both Gamma and inverse Gaussian processes, such as the wear process and fatigue crack propagation. Inverse Gaussian process-based degradation model has recently gained significant attention owing to their ability and flexibility in incorporating random effects and covariates [133].

In many real-world situations, degradation processes of engineered systems may be caused or indicated by various time-varying environmental factors and/or external shocks. Environmental factors include general external factors (e.g., temperature, vibration, load, and running speed) and internal factors (e.g., inherent deterioration mechanism and deterioration level). For example, the wear process of a cutting tool can be affected by running speed and material properties. Meanwhile, the proportional hazard model (PHM), as one of the most reported covariate-based models, has been extensively applied in reliability analysis and maintenance optimization owing to its flexibility and simplicity. The conventional PHM assumes that the hazard rate function comprises two multiplicative parts: a baseline hazard rate function and a function of covariates. Time-varying environmental factors per se can be easily incorporated into degradation processes. Various PHMs and their variants have been developed to accommodate various industrial scenarios. However, the influence of shocks can be versatile: (1) some parameters in a degradation model are related to the number of shocks, (2) an additional degradation increment is incurred directly by the occurrence of a shock, and (3) a shock can induce a system failure, which poses the case of competing failure.

1.1.3 Maintenance Efficiencies

In maintenance optimization problems, various maintenance actions are optional to retain a deteriorated system or recover it to a better condition. Repair refers to actions carried out upon the failure of a system and thus corresponds to CM. Before implementing a maintenance action on a deteriorated system, decision makers must determine the degree to which the system can be restored by the action. Maintenance actions can be classified into the following five categories:

Perfect maintenance or *prefect repair* can restore the condition of a system to “as good as new”; that is, the condition of aged or failed system is recovered to the status that is the same as a brand-new system. A typical example of perfect