University of Alberta

Equipment Degradation Diagnostics and Prognostics Under a Multistate Deterioration Process

by

Ramin Moghaddass

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of **Doctor of Philosophy**.

in

Engineering Management

Department of Mechanical Engineering

Edmonton, Alberta
Fall 2013

This thesis is dedicated to my parents
for their love, endless support, and encouragement.

Abstract

The increasing level of system complexity in the current competitive market implies that efficient asset management is of paramount importance, particularly for systems with costly downtime and failure. Timely detection of faults and failures through an efficient reliability and health management framework allows for appropriate maintenance actions to be scheduled proactively to avoid catastrophic failures and minimize unnecessary maintenance actions. This thesis employs a general stochastic process - the Nonhomogeneous Continuous-Time Hidden Semi-Markov Process - to model a condition-monitored degradation process with hidden states. This thesis also proposes an unsupervised learning process, which can be used to estimate the characteristic parameters of the degradation and observation processes. It then develops dynamic diagnostic and prognostic measures for online health monitoring. Finally, it introduces a condition-based replacement policy that can be used as an online tool to determine when to replace a degraded device under condition monitoring.

Acknowledgements

I would like to express my heartfelt gratitude with sincere respect to my supervisor Dr. Ming J. Zuo for his direction, inspiration, and support through all my studies and research. His enthusiasm and encouragement made me eager to succeed. My sincere thanks to my Ph.D. examining committee members, Dr. John Doucette, Dr. Armann Ingolfsson, Dr. Faisal Khan, and Dr. Jie Han, for their helps in improving this thesis. It would not have been possible to write this doctoral thesis without the help and support from all members in the Reliability Research Lab, especially Mr. Mayank Pandey and Dr. Mohammad Hosseini. Working with them was an exciting and beneficial experience. Finally, I would like to thank my wife for her love, support, and understanding during my PhD study.

Contents

List of Tables

List of Figures

Acronyms

BIC Bayesian Information Criterion

CBM Condition-Based Maintenance

CDF Cumulative Distribution Function

CM Condition Monitoring

CTAMP Continuous-Time Aging Markovian Process

CTMP Continuous-Time Markov Process

DTAMP Discrete-Time Aging Markov Process

DTMP Discrete-Time Markov Process

EDCTSMP Explicit-Duration Continuous-Time Semi-Markov Process

EDDTSMP Explicit-Duration Discrete-Time Semi-Markov Process

EM Expectation-Maximization

EMC Embedded Markov Chain

HCTSMP Homogeneous Continuous-Time Semi-Markov Process

HDTHSMP Homogeneous Discrete-Time Semi-Markov Process

HMM Hidden Markov Model

HSMM Hidden Semi-Markov Model

MDP Markov Decision Process

MRL Mean Residual (Remaining) Life

MSE Mean Squared Error

MTTF Mean Time To Failure

NHCTHSMP Nonhomogeneous Continuous-Time Hidden Semi-Markov Process

NHDTHSMP Nonhomogeneous Discrete-Time Hidden Semi-Markov Process

OPM Observation Probability Matrix

PCA Principal Component Analysis

PDF Probability Distribution Function

PHM Prognostic and Health Management

RBM Risk-Based Maintenance

RCM Reliability Centered Maintenance

RUL Remaining Useful Life

SMDP Semi-Markov Decision Process

TBM Time-Based Maintenance

Notation List

M Multistate structure ($M = (\zeta, \theta)$)

$\boldsymbol{\zeta}$ Multistate topology ($\zeta = (N, \Omega, \xi, \lambda, I, V)$)

N The total number of health states

$\boldsymbol{\Omega}$ Transition diagram (describes the connectivity between states)

$\boldsymbol{\xi}$ Types of the transitions between states

$\boldsymbol{\lambda}$ Statistical form of transition rate functions

I Final condition monitoring feature used for health monitoring

V=(v$_1, v_2, ..., v_m$) The condition monitoring indicator space

$\mathbf{v}_i$ The ith output of the observation process

m Number of discrete possible outputs of the observation process

$\boldsymbol{\theta}$ The set of unknown characteristic parameters that characterize $\zeta(\theta = (\Gamma, B))$

$\boldsymbol{\Gamma}$ The set of characteristic parameters associated with the degradation process

B The set of characteristic parameters associated with the observation process

$\mathbf{b}_j(k)$ The probability of observing the kth value of the condition monitoring indicator when the device is in state j (an element of B)

L Total age of the device

$\mathbf{T}_n$ The time of the nth transition

$\mathbf{X}_n$ The state of the device after the nth transition

$\mathbf{U}_n$ The condition monitoring indicator at the nth observation point

$\mathbf{T}_n^0$ The age of the device at the nth observation point

$\mathbf{X}_n^0$ The state of the device at the nth observation point

$\mathbf{Z}_t$ The state of the device at time t

$\mathbf{Y}_t$ The condition monitoring indicator value at time t

$\mathbf{FS}_i$ The set of states to which a transition may occur from state i

$\mathbf{BS}_i$ The set of states from which a transition may occur to state i

$\mathbf{K}$ The number of historical (real-time) sequences of the observation process available for training a NHCTHSMP

$\mathbf{d}_k$ The number of observation points for the kth sequence of the observation process

$\mathbf{t}_1^{(k)}, t_2^{(k)}, ..., t_{d_k}^{(k)}$ The observation time-points of the kth sequence of the observation process

$\mathbf{O}^{(k)}$ The kth sequence of the observation process

$\mathbf{O}_p^{(k)}$ The condition monitoring indicator value at the pth CM point (time $t_p^{(k)}$) of the kth sequence of the observation process

$\mathbf{Q}_p^{(k)}$ The unobservable state of the device at the pth CM point(time $t_p^{(k)}$) of the kth sequence of the observation process

Chapter 1

Introduction

1.1 Background and Motivation

The increasing level of system complexity in the current competitive market implies that efficient asset management is of paramount importance, particularly for systems with costly downtime and failure. It is reported in [2] that 15-40 % of manufacturing costs across many industries are attributable to maintenance. Particularly in capital and energy-intensive industries such as petroleum, the economic loss of downtime and failure is huge. Because of the huge cost of maintenance, the need for a low-cost health monitoring system and integrated predictive maintenance framework has increased significantly over the years. Timely detection of faults and failures through an efficient reliability framework allows for scheduling appropriate maintenance actions to avoid catastrophic failures, such as the 2010 Gulf of Mexico oil spill [1, 3] (see Figure 1.1).

Reliability, which is the ability of a system to successfully operate on a satisfactory level of operation or performance, has always been an important aspect of efficient management of engineering assets. In the current competitive market, reliability analysis is playing a significant role on equipment life cycle cost minimization. Reliability analysis includes broad range of aspects from optimal design of complex systems to development of integrated maintenance strategies. Reliability analysis of mechanical devices has been studied extensively over the past decades and as a result a considerable amount of research results has been published. Many of the theoretical aspects of reliability, such as optimal design of systems, important diagnostic and prognostic measure calculation, and optimal maintenance strategies, have been implemented in commercial software packages (such as Weibull ++ for life data analysis [4], EXAKT for condition-based maintenance [5], etc.), which are

Figure 1.1: The Gulf of Mexico oil spill in 2010 [1]

currently used in various industries and businesses. For systems with expensive downtime and failure costs, an efficient reliability framework not only can significantly decrease the overall operation and maintenance costs, but also can prevent catastrophic and unexpected failures and minimize safety issues.

Because most mechanical devices operate under some sort of stress, load, static, and dynamic forces, they tend to deteriorate or degrade over time. In real-world systems, this gradual deterioration process eventually causes systems to become unable to operate at their desired level of performance, reliability, and availability. For mechanical devices, the overall health status can deteriorate over time due to one or multiple degradation processes. When a degradation indicator or factor, exceeds a certain threshold, then the device is considered to have failed. Finding the structure of the degradation process and developing cost-effective maintenance strategies are the topics of numerous research works in the reliability domain.

1.2 Multistate Degradation Analysis

In conventional reliability models, at any time point, equipment or systems are assumed to be in either of two possible health states, which are referred to as working state and failure state. This type of binary reliability analysis has been studied for many years and hundreds of high quality journal papers and monographs have been published accordingly. A comprehensive review of models for reliability and availability analysis of such systems can be found in [6]. Although, binary reliability analysis has constructed most of the fundamental theories in the reliability domain, it is subject to practical shortcomings that reveal the need for a more general and advanced type of reliability analysis called multistate reliability analysis.

Recent research studies and reported practical experiences have verified that most mechanical devices have more than two health states and conditions, in which they have not only different operational performances and outputs, but also different physical properties. Therefore, binary degradation models cannot fully capture the real degradation behavior of most mechanical systems. Over time, a device with multistate health states performs at different intermediate health states between working perfectly and complete failure. It is to be noted that the advantage of considering multistate modeling is that binary modeling can always be considered as the simplest case of multistate modeling. Therefore, results on multistate modeling can be easily applied to binary modeling. Researchers have investigated multistate modeling and developed mathematical models to evaluate the reliability and availability of many systems under multistate degradation. A comprehensive literature review of multistate reliability modeling can be found in [7].

The multistate deterioration process can be divided into two categories as: (1) continuous-state space [8] and (2) discrete-state space [9]. In a continuous-state degradation process, the overall degradation process is modeled as a continuous variable. The device is considered failed when this degradation process exceeds a predefined threshold. The major difficulty in implementing continuous-state reliability analysis is its mathematical complexity [10]. This challenge is the motivation for multistate reliability analysis in the discrete-state space, in which the overall status of the degradation process is divided into discrete levels with certain properties ranging from perfect functioning to complete failure. It should be pointed out that when the number of states is very large, then discrete-state space and continuous-state space become equivalent to each other.

Another classification of multistate degradation can be done with respect to the time domain considered for the degradation process. From this viewpoint, the multistate degradation process can evolve according to a (1) discrete-time process or (2) a continuous-time process. In a discrete-time degradation process, transitions are allowed only at discrete points while in a continuous-time degradation process, transitions are possible at any time point. The continuous-time degradation process is more compatible with mechanical devices, where degradation transitions evolve in a continuous time domain, that is, the degradation transitions can happen at any time point. This thesis focuses only on multistate degradation processes with discrete-state space and continuous-time domain.

1.3 Condition Monitoring of the Degradation Process

Monitoring the equipment behaviour while it is operating is a key step in implementing any reliability and maintenance framework. Most maintenance decisions are made based on the trend of the degradation process over a certain period of time. Due to the complexity of degradation processes and other practical limitations, such as the cost and the time required for inspection, the health status (degradation level) of most mechanical devices is not directly or continuously observable. In such cases, the degradation states may be directly observable only at limited points referred to as inspection points [11], totally unobservable [12], and/or indirectly observable through certain methods, such as condition monitoring [13]. The latter is also called degradation process with incomplete information [14] or partially-observed degradation process.

For three reasons, the degradation process with incomplete information has received much attention over the past years. First, in many practical cases, due to the huge cost of visual inspection and technical issues, obtaining the actual health status of the system while in operation mode is not possible. Second, for many degradation processes, no single or multiple observable degradation indicators having a direct and definite relationship with the actual health state can be found. Finally, with the advancement of sensor technology, it is very likely that single or multiple indicators with indirect relationships with the actual level of degradation can be used for health monitoring. Therefore, instead of costly and time-consuming inspections, the health status can be indirectly monitored without having to terminate the operation of the device. The process of monitoring certain aspects of health conditions in a device with the purpose of identifying the development of a failure is referred to as condition monitoring (CM).

Using condition monitoring data, certain indicators extracted from multiple sensors that have stochastic relationships with the actual degradation states of the system can be employed for diagnostic and prognostic purposes. This stochastic relationship reflects the non-deterministic relationship between the condition monitoring features and degradation levels, which can be represented by certain random variables. With condition monitoring, at each observation point, we can monitor directly or indirectly some important features (indicators) extracted from a single or multiple sensors that present at least partial information regarding the actual degradation process of the equipment. The process of extracting useful information from raw condition monitoring data is referred to as feature extraction. Each feature or alternatively called indicator is expected to represent specific characteristic information regarding the degradation and observation processes. These features can be directly obtained from measurements of multiple sensors (like vibration data, temperature, and pressure) or can be the result of a feature combination or a feature fusion procedure (such as principal component analysis (PCA)) that can transform a set of features to a single feature with more useful information. Based on the physical characteristics of the equipment under consideration and its degradation mechanism, different types of features can be calculated.

The indirect relationship between condition monitoring indicators extracted from multiple sensors and actual health states of mechanical devices motivates one to use condition monitoring techniques for degradation assessment and maintenance decision making. Condition monitoring is an efficient way to track the health state (condition) of the equipment, when it is not directly observable. Condition monitoring has been widely used in the literature for reliability analysis of devices under multistate degradation. Based on the above discussion, multistate degradation models with partially observable states involve two types of stochastic processes, namely, the degradation process and the observation process. The degradation process refers to the characteristics of the degradation transitions between different levels of health states and the observation process refers to the stochastic relationship between the degradation process and the condition monitoring indicators.

This PhD thesis focuses only on devices under multistate degradation processes with partially (indirectly) observable states, which is an appropriate representative of the overall degradation mechanism of many mechanical devices. Many applications of devices with degradation and observation processes are reported in the literature, such as in [15] for a hydraulic pump with four levels of health states and in [16] for a gearbox with five degradation states, both under vibration monitoring.

1.4 Structure Modeling and Parameter Estimation

In order to develop diagnostic and prognostic models for condition-monitored devices under multistate degradation with partially observable states, stochastic models can be employed to formulate the associated degradation and observation processes. In real-world systems, applying diagnostic and prognostic models requires determining the structure of the degradation and observation processes and estimating the characteristic parameters of the associated stochastic models from real-time data. Therefore, the structure determination and model selection techniques for the associated degradation and observation processes and parameter estimation models to estimate the corresponding characteristic parameters of the selected stochastic models are of high interest in diagnostic and prognostic analysis of mechanical devices under condition monitoring.

The process of structure modeling determines the configuration of the degradation and observation processes and the process of parameter estimation determines the set of characteristic parameters that characterize those selected stochastic processes. These characteristic parameters may vary, depending on the types (statistical form) of the degradation and observation processes employed for different devices. Model selection and parameter estimation play key roles in developing a prognostic and health management framework for mechanical systems. Failure to appropriately define the structure of the failure and degradation mechanisms and inaccurate estimation of the associated characteristic parameters may result in inaccurate residual life estimation, leading to either unnecessary maintenance actions or catastrophic failures. Although numerous models and algorithms have been developed for model selection and parameter estimation, this area of research deals with some important challenges as will be described in Section 1.7. This thesis aims to provide maintenance decision makers a tool that can assist them to select and train a reasonable mathematical model for degradation and observation processes associated with different types of systems under gradual degradation.

1.5 Diagnostics and Prognostics

Diagnostics and prognostics are known as two important aspects of condition-based maintenance (CBM). Diagnostics refers to activities that are done when a fault or degradation occurs. Fault detection, fault isolation, and fault identification are three main elements of fault diagnosis. Fault detection is a task to identify an abnormal condition in a monitored system, fault isolation is a task to identify the location of the faulty element, and fault identification is a task to identify the nature of the detected fault [17]. When condition monitoring is employed, degradation diagnosis becomes an online process that aims to monitor and detect abnormality in a system. Finding the current health status and the degradation level of the device using condition monitoring data is one of the key activities involved in degradation diagnosis. On the other hand, degradation prognosis refers to activities that help predict future behaviour of the degradation process. Examples are estimating the remaining useful life, its confidence limits, and the probability of a failure within a time interval in the future. It is desired that the result of degradation diagnosis and prognosis ultimately assists maintenance decision making.

Numerous research papers have been published on machinery diagnostics and prognostics. A review on diagnostic and prognostic methods using condition monitoring can be found in [18, 19, 20]. In the domain of reliability and maintenance, it is very common to extract certain so-called reliability measures, which can directly represent certain aspects of the degradation process or the observation process. Reliability measures are sometimes referred to as performance measures. Finding accurate reliability measures, such as hazard rate and remaining useful life (RUL), plays an important role in minimizing the overall maintenance cost of systems. Therefore, it is important to investigate how condition monitoring data can be used to generate important reliability measures in order to conduct reliability evaluation of a device under a multistate degradation process with unobservable states.

As will be discussed in Chapter 5, reliability measures can be classified into static and dynamic measures. Static reliability measures are independent of the actual degradation process while the device is operating. For example mean time to failure is an important reliability measure that is usually calculated by using some historical knowledge about the degradation process. However, dynamic reliability measures are calculated using some information on the actual degradation process while the device is in operation mode. For example, conditional remaining useful life is a dynamic reliability measure, which is calculated based on the available information extracted from the degradation process. This thesis focuses on important dynamic reliability measures conditional to the history of the observation process, which can be used for online degradation diagnosis and prognosis.

As will be discussed in Chapter 5, reliability measures can be either diagnostic reliability measures or prognostic reliability measures. The objective of diagnostic reliability measures is to provide some information on the current health status of the device, while prognostic reliability measures aim to provide information on the future health status of the device. Although both diagnostic and prognostic reliability measures are important, over the past years, there has been a significant increase in research work regarding the prognostic reliability evaluation of mechanical systems. The main challenges of implementing machine prognostics are degradation assessment and remaining useful life prediction [21]. A comprehensive literature review of recent research work in this area and advantages and disadvantages of available models can be found in [22]. In addition, the strengths and weaknesses of current prognostic models and summary on how each can be applied to engineering prognostics have been discussed in [23].

1.6 Maintenance Decision Making Using Condition Monitoring Data

Conventional maintenance decision making is based on either corrective maintenance actions or preventive maintenance actions. The corrective maintenance is a set of activities that are performed only after a failure occurs. Corrective maintenance includes repair, restoration or replacement of components to restore the system to a working state (condition). Preventive maintenance or alternatively called schedule-based maintenance is a set of maintenance actions, which are done on some predefined intervals irrespective of the condition of the device. Periodic oil changes in a car, even though a significant portion of oil's life is still remaining is an example of a schedule-based maintenance [24].

The shortcomings of corrective and preventive maintenance strategies, such as independence to the actual level of degradation, unnecessary maintenance actions, and unexpected failures, have increased the need for another type of maintenance strategies referred to as condition-based maintenance (CBM). The concept of CBM was first introduced in late 1940s by the Rio Grande Railway Company and was initially called predictive maintenance [24]. CBM deals with the current condition of the device, which is derived from a real-time assessment of different aspects of the device′s condition obtained from embedded sensors or particular measurements. Condition-based maintenance (CBM) provides maintenance decision support information based on the data obtained from condition monitoring.

The ultimate task in a condition-based maintenance framework is to provide decision support information that can facilitate decision makers to determine, in a timely and efficient way, when to replace or maintain a degraded device. Employing condition monitoring data for dynamic maintenance decision making depends on the quality of the condition monitoring data, the structure of the maintenance decision rule, the dynamic characteristics of health state evolution, as well as the maintenance costs [25]. Efficient maintenance decision making very much depends on the selected structure of the degradation and observation processes, the accuracy of the unknown parameters, and the effectiveness of important performance measures. This reveals the need for an integrated framework, which takes into account all important aspects of condition-based maintenance. This thesis will develop a dynamic condition-based maintenance model, which can employ online condition monitoring data to determine when to replace a degraded device, which is subject to a multistate degradation process with unobservable states.

1.7 Research Scope and Objectives

This thesis aims to provide maintenance decision makers a comprehensive tool that can facilitate online health monitoring and timely and precise maintenance decision making for a degraded device under condition monitoring. The focus of this thesis is on devices that are subject to multistate degradation processes and are considered as good as new after replacement. This includes the vast majority of mechanical devices used in industries where a device is replaced at failure or a time point before failure. The result of this thesis can also be applied to repairable mechanical devices for which the repair process can bring the health status to as good as new. Condition monitoring information extracted from multiple sensors will be used to estimate the parameters of the model and then diagnostic and prognostic measures for the purpose of health state recognition will be developed.

As discussed earlier, condition-based maintenance has been applied extensively in machinery diagnostics and prognostics, however, the need for an integrated condition-based maintenance framework for devices under multistate degradation processes and unobservable states that can address practical limitations of reported models is still a challenging topic. Surprisingly, very limited research work has been devoted to multistate degradation processes with unobservable states. In this thesis, a condition-based maintenance framework is proposed in the sense that the main steps that need to be done for health monitoring of a device under degradation, are covered. The proposed CBM framework in this thesis includes three major phases, which are (1) modeling, (2) training, and (3) implementation. A schematic view of this framework, which presents the interaction among the important steps of the CBM, is shown in Figure 1.2. Modeling includes steps that contribute to the development of the mathematical model for a multistate degradation structure with unobservable states. Training includes steps that need to be done to characterize the structure of the multistate degradation model to be used for health monitoring. Implementation includes steps that are related to the actual online diagnostic and prognostic monitoring of the device as well as maintenance decision making.

Several research topics are defined, in the sense that the challenges in the three phases of modeling, training, and implementation of the proposed condition-based maintenance framework are addressed. Each topic, which relates directly to a phase in the proposed CBM framework, is thoroughly investigated in a separate chapter of this thesis, as described in Section 1.8. The following research topics are specifically investigated in this thesis:

I. Modeling Phase (will be discussed in Chapter 3)

- Developing a general mathematical framework for the degradation and observation processes associated with a condition-monitored device under multistate degradation processes with unobservable states.

II. Training Phase (will be discussed in Chapter 4)

- Developing a framework for training a multistate degradation structure with unobservable states including a general model selection and a parameter estimation method using condition monitoring data

III. Implementation Phase (will be discussed in Chapters 5 and 6)

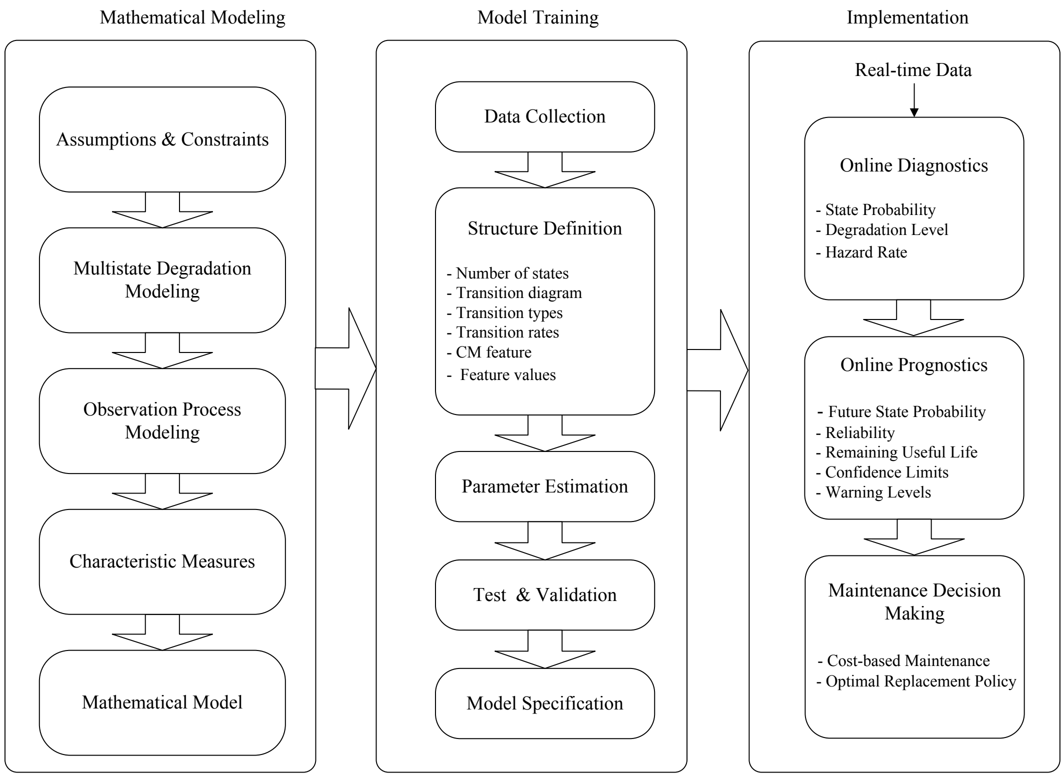

Figure 1.2: The proposed framework for condition-based maintenance

- Developing several important dynamic diagnostic and prognostic measures, which can be used for online health monitoring of a device under multistate degradation using condition monitoring data
- Developing a dynamic condition-based model, which can employ online condition monitoring data for maintenance decision making

It is expected that the results of this thesis will advance the state of the art for practical diagnostic management of mechanical systems. The main application of this thesis is on the operation management of condition-monitored machinery in capital and energy-intensive oil, gas, and wind industries, which deal with energy resources where (1) the actual degradation status is not directly observable, and (2) increasing the system reliability is a top priority due to the high costs of plant downtime and failure. The proposed CBM framework aims to make significant contributions to such industries around the world by providing decision-making tools for online health monitoring of assets to prevent unexpected failures, reduce the operation and maintenance costs, and therefore make them more competitive in their respective markets.

1.8 Thesis Organization

This thesis is composed of 8 chapters. Chapters are defined in the sense that (1) the identified limitations of available work in the literature are covered and (2) each chapter has definite deliverables contributing to the domain of reliability analysis of degrading systems under condition monitoring. The list of chapters and main contents of each chapter are as follows:

- Chapter 1 presents a brief introduction to the challenging topics in multistate degradation modeling and condition-based maintenance using condition monitoring data and illustrates the structure and outline of this thesis.

- Chapter 2 reviews relevant work in multistate degradation modeling, parameter estimation, diagnostics and prognostics, and maintenance decision making, and highlights the limitations of reported models and the contributions to be made in this thesis. Several research topics are then defined based on these limitations.

- Chapter 3, which deals with the modeling phase of the proposed CBM framework given in Figure 1.2, provides fundamentals of a general and flexible stochastic process called nonhomogeneous continuous-time hidden semi-Markov process, which can be used for modeling a multistate degradation structure with unobservable states. Also, the assumptions made throughout the paper are clearly described in this chapter. This chapter is the fundamental chapter of this thesis and its results are directly used in the rest of the thesis. Some of the results of this chapter are published in [26].

- Chapter 4, which deals with the model training phase of the proposed CBM framework given in Figure 1.2, introduces a CBM training approach, which can be used to determine the structure of the degradation and observation processes associated with a condition-monitored device under multistate degradation. In other words, how to utilize historical condition monitoring data for training a multistate structure is discussed in this chapter. The result of this chapter gives a trained stochastic model that can be used for degradation diagnosis and prognosis and maintenance decision making. The results of this chapter are published in [9, 26].

- Chapter 5 introduces the definition and mathematical derivations of important diagnostic and prognostic measures, which can employ condition monitoring data for online health monitoring of a device under gradual multistate degradation. Some of the results of this chapter are reported in [27].

- Chapter 6 presents a cost-effective and dynamic condition-based maintenance model, which can provide decision support information on when to replace a degraded device. Chapters 5 and 6 deal with the implementation phase of the proposed CBM framework given in Figure 1.2.

- In Chapter 7, the application of the proposed CBM framework on a benchmark prognostic database is presented. The proposed multistate structure is first employed to model the degradation process of turbofan engines and then diagnostic and prognostic measures are calculated for online health monitoring. Finally, the proposed condition-based maintenance model is employed to determine when to replace a degraded engine.

- Chapter 8 summarizes the contributions made throughout the thesis and discusses the results obtained in each chapter. It finally introduces the possible directions for moving forward in future work.

Chapter 2

Literature Review

This chapter is devoted to reviewing available research works on reliability analysis and maintenance decision making for devices under gradual degradation using the concept of multistate modeling. This chapter is organized as follows: Section 2.1 reviews current work on multistate degradation models including multistate degradation models with observable states and with unobservable states. Section 2.2 reviews available work on model selection and parameter estimation for multistate degradation models with unobservable states. Section 2.3 reviews work on diagnostic and prognostic health monitoring of multistate degradation models. Section 2.4 reviews condition-based maintenance models available for multistate degradation processes. Finally in Section 2.5, the literature review is summarized, the main limitations of the available methods are listed, and contributions made in this thesis are highlighted.

2.1 Multistate Degradation

As most mechanical devices operate under some sort of stress, load, and static and dynamic forces, they tend to deteriorate or degrade over time. In real-world systems, this gradual deterioration process eventually causes the systems to be unable to operate at their desired level of performance, reliability, and/or availability. Therefore, the overall health status of most mechanical systems gradually deteriorates over time. When a degradation indicator or a degradation factor, which reflects the level of degradation, exceeds a certain threshold, then the device is considered to be failed. Finding the structure of the degradation process associated with a certain device and developing cost-effective maintenance strategies are the topics of numerous research works in the reliability domain.

In conventional reliability analyses, systems are often assumed to be in either of two possible health states, namely, the working state and the failure state. The probability distribution of the Time To Failure (TTF) plays the key role in binary reliability modeling. A comprehensive literature review on the reliability of such binary systems has been conducted in [6]. Numerous reliability and maintenance models have been developed for these binary systems [28, 29, 30, 31]. However, most mechanical devices operating under a stress or a load condition deteriorate or degrade over time. These devices may perform at several intermediate health states ranging between working perfectly and complete failure [9]. Each state level may reflect certain operational performance, efficiency, and physical property of the device.

Compared to the binary degradation process, the multistate degradation process is a more realistic representative of a gradual degradation process. It is to be noted that the advantage of considering multistate degradation modeling is that binary degradation modeling can always be considered as the simplest case of multistate modeling. Therefore, results on multistate degradation modeling can be easily applied to binary degradation modeling. Researchers have investigated multistate degradation modeling and developed mathematical models to evaluate the reliability and availability of systems under multistate degradation. For example in [32], it is assumed that a friction drilling device has five levels of degradation based on the size of flank wear referred to as sharp, normal wear, micro fracture, macro wear, and breakage. A comprehensive literature review and fundamentals on multistate reliability modeling can be found in [7, 33].

Multistate degradation models can be categorized into certain types from three different aspects. In the following, major types of multistate models used in the literature for degradation modeling are briefly reviewed.

2.1.1 Types of Multistate Degradation Models

Multistate degradation processes can be divided into several categories with respect to (1) levels of degradation, (2) time domain of the degradation process, and (3) observability of the degradation process. With respect to the levels of degradation, the multistate deterioration process can be divided into two categories as: (1) continuous-state space [8] and (2) discrete-state space [9]. In a continuous-state space degradation process, the overall degradation process is considered to be a continuous variable. The device is considered failed when this degradation process exceeds a predefined threshold. The major difficulty in implementing continuous-state reliability analysis is its mathematical complexity [10]. This challenge is the motivation for multistate reliability analysis in discrete-state space, in which the overall status of the degradation process is divided into discrete levels with certain properties ranging from perfect functioning to complete failure. This complies with most real cases where discrete levels of health states are considered for reliability analysis of a device. It should be pointed out that when the number of states approaches infinity, then discrete-state space and continuous-state space become equivalent to each other.

Another possible classification of multistate degradation models can be done with respect to the time domain considered for the degradation process. From this viewpoint, the multistate degradation process can evolve according to a (1) discrete-time stochastic process or (2) a continuous-time stochastic process. The discrete time can be used when one is interested in the number of cycles in a system or the number of times (hours, days, etc.) that a particular event occurs [34]. Although modeling and calculus in the discrete time domain are less expensive than continuous time domain, a continuous-time degradation process is compatible with more mechanical devices, where degradation transitions evolve in a continuous time domain, that is, transitions can occur at any time point.

Monitoring the equipment behaviour while it is operating is a key step in implementing reliability and maintenance frameworks. Most maintenance decisions are made based on the trend of the degradation process over a certain period of time. Usually, due to the complexity of degradation processes and other practical limitations, such as the cost and the time required for inspection, the health status (degradation level) of most mechanical devices is not directly or continuously observable. With respect to the possibility of observing the actual level of a degradation process while the device is operating, multistate degradation processes may be divided into four main types as: (1) continuously and directly observable, (2) directly observable only at limited points referred to as inspection points [11, 35, 36, 37], (3) totally unobservable [12], and (4) indirectly observable through certain methods, such as condition monitoring [13, 38, 39, 40]. The indirectly observed degradation process is also called degradation process with incomplete information [14] or partially-observed degradation process [41].

Due to the following three reasons, a degradation process with incomplete information has received more attention than others over the past years. First, in many practical cases, due to the huge cost of visual inspection and technical issues, obtaining the actual health status of the system while in operation is not possible. Second, for many degradation processes, no single or multiple observable degradation indicators having a direct and definite relationship with the actual health state can be found. Finally, with the advancement of sensor technology, it is very likely that a single or multiple indicators having an indirect relationship with the actual level of degradation can be employed for health monitoring. The process of monitoring certain aspects of health conditions in a device with the purpose of identifying the development of a failure is referred to as condition monitoring.

This thesis focuses only on multistate degradation processes with discrete-state space and continuous-time domain, and unobservable degradation process where states are only indirectly observable through condition monitoring. Such types of devices comply with many practical cases. Based on the above discussion, multistate degradation models with partially observable states involve two types of stochastic processes, namely, the degradation process and the observation process. The degradation process, which is a discrete-space continuous-time stochastic process, refers to the characteristics of the degradation transitions between different levels of health states and the observation process refers to the stochastic relationship between the degradation process and the condition monitoring indicators. In the following two subsections, available types of degradation and observation processes and their applications in multistate degradation modeling are reviewed.

2.1.2 Degradation Process in Multistate Degradation Models

As discussed earlier, the multistate degradation process deals with the characteristics of the degradation transitions between different levels of health states. In other words, how fast the device is deteriorating over time is defined based on this degradation process. When a stochastic process is employed to represent this degradation process, the future development of the degradation process is governed based on a random process. Therefore, the multistate degradation process can be defined as a continuous-time stochastic process $X(t)$ with a finite state space $E = \{1, \ldots, N\}$, where $X(t)$ represent the state occupied at time t. With respect to the dependency of degradation transitions to the history of the degradation process, the multistate degradation process may be divided into two major types of stochastic processes, namely, Markovian degradation process and non-Markovian degradation process.

In a multistate degradation process with a Markovian structure, a degradation transition between two states depends only on the current states involved in the transition, that is, the degradation process is independent of the history of the process. However, in a multistate degradation process with a non-Markovian structure, the above-mentioned assumption is relaxed, that is, the degradation process may depend also on other factors. It is known that inference under Markovian structures is much simpler than non-Markovian structures. As will be reviewed later in this chapter, most of the literature on multistate degradation has focused on cases with the simple Markovian structure. The Markov model is clearly the most popular one, as it assumes that the sojourn time at each state is exponentially distributed [42]. Discrete-time and continuous time multistate degradation models with Markovian structures are reviewed later in this subsection.

Different types of non-Markovian structures are considered in the literature for multistate degradation modeling. One of the most commonly used non-Markovian structures in the literature to represent the multistate degradation process is the semi-Markovian structure in which the one-step transition between two states follows a Markovian structure and the sojourn time at each state follows an arbitrary distribution [43]. Another example of non-Markovian degradation models is the multistate regression model, where degradation transition may also depend on a covariate vector Z, possibly time-dependent [44]. Proportional hazard model (PHM) with multistate structure is one of the most typical forms of multistate regression models [41]. Later in this subsection, discrete-time and continuous-time multistate degradation structures in the literature with semi-Markovian structures as well as the proportional hazard model as the most common types of a non-Markovian structure are reviewed. In addition, the main characteristics as well as the major limitations of each structure will also be briefly discussed.

It should be pointed out here that in the reliability literature, the keyword hidden is commonly used to represent an unobservable or indirectly observable stochastic process (see details in Section 2.1.3). For example, a multistate degradation process with indirectly observable healths states evolving according to a Markovian structure is usually referred to as a multistate process with a hidden Markovian structure. Markov and semi-Markov models with partially observable states (hidden states) are alternatively called hidden Markov models (HMM) and hidden semi-Markov models (HSMM), respectively. HMM and HSMM are well-studied methods and have been successfully applied in many areas such as speech recognition [45], health care [46], software reliability [47], and condition monitoring [48]. Review on models that are used to represent a hidden degradation process with an observation process is discussed in Section 2.1.3.

In a multistate degradation process, the degradation transition may depend on certain factors, such as (*a*) the two states involved in the transition (the actual level of degradation), (*b*) the time that the device reached the current state, (*c*) the time already spent (sojourn time) at the current state, (*d*) the total age of the device (sum of *b* and *c*), or (*e*) any other covariate that may or may not depend on the above factors. With respect to the above factors, multistate degradation models reported in the literature can be classified into five major categories: (1) discrete-time Markovian structure, (2) continuous-time Markovian structure, (3) discrete-time semi-Markovian structure, and (4) continuous-time semi-Markovian structure, and (5) proportional hazard model. Each of the above-mentioned categories are described below.

In a multistate degradation process with discrete-time Markovian structure, referred to as a discrete-time Markov process (DTMP) [49], after a fixed time interval ($\triangle$) or so-called step, the device transits from its current health state i to a degraded state j with a fixed probability of $P_{i,j}$. Based on this assumption, the probability of degradation depends only on the states involved in the transition and is affected neither by the total age of the equipment nor the time duration the device has spent at the current state. This means that the probability of transition from state i to state j at time t during the next interval is always $P_{i,j}$, that is the ignorance of aging. This results in a geometric sojourn time distribution for state i as $\Pr(D_i = d) = (1 - \sum_{j \neq i} P_{i,j})^{d-1} (\sum_{j \neq i} P_{i,j})$, where D_i is the sojourn time at state i ($d = 1\triangle, 2\triangle, ...$). Examples of the application of the discrete-time Markovian degradation can be found in [49, 50, 51, 52, 53]. For DTMP, the fixed transition probabilities and the geometric sojourn time distribution limit the use of this type of model in real deteriorating systems [54, 55, 56]. In other words, the fact that the probability of a transition to a degraded state increases as the device ages cannot be reflected by DTMP.

In a multistate degradation process with continuous-time Markovian structure, referred to as a continuous-time Markov process (CTMP) [57], the device can continuously (not necessarily at discrete time points) degrade from its current state i to a lower state j with a constant time-homogeneous (time-independent) transition rate $\lambda_{i,j}$. Based on this assumption, the sojourn time distribution at state i follows an exponential distribution as $f_i(d) = (\sum_{j \neq i} \lambda_{i,j}) \exp(-(\sum_{j \neq i} \lambda_{i,j})d), d \geq 0$. In a Markovian degradation structure, the transition between two states at time t depends only on the two states involved and is independent of the history of the process before time t (memoryless property). Examples of the application of the continuous-time Markovian degradation can be found in [57, 58, 59, 60, 61]. For CTMP, the constant transition rates (independent of the age of the device) and the exponential sojourn time distribution limit the use of this model in practical cases.

To overcome the above-described limitations of multistate degradation models with Markovian structures, researchers have considered semi-Markov structures, which take into account the history of the process and consider arbitrary sojourn time distributions at each state. With respect to the dependency of transitions to the states involved, the time that the device reached the current state, the time spent at the current state, and the total age of the device, the semi-Markov process can be in one of the four main forms, namely, discrete-time and continuous-time aging Markovian process (DTAMP and CTAMP), homogeneous discrete-time and continuous-time semi-Markov process (HDTSMP and HCTSMP), explicit-duration discrete-time and continuous-time semi-Markov process (EDDTSMP and EDCTSMP), and nonhomogeneous discrete-time and continuous-time semi-Markov process (NHDTSMP and NHCTSMP). In the following, the main characteristics of each of the above semi-Markov structures are briefly reviewed and their applications in multistate degradation modeling are described.

In multistate degradation with a discrete-time aging Markovian structure as reported in [56, 62, 63], the device degrades from any current state i to a degraded state j according to a transition probability, which changes with the total age of the device considering an aging factor. Chen and Wu [56] considered a fixed aging factor for a discrete-time Markov model in a way that the transition probabilities are updated after each fixed time interval. Peng and Dong [15] considered three types of aging factors for updating the probability transition matrix. With such aging factors, the probability of transition to a degraded state is increased after each observation interval $\triangle$. They developed estimation methods to find these aging factors. Another example of discrete-time aging Markovian degradation models can be found in [63], where a time-dependent transition probability matrix is used to describe the deteriorating mechanism of a production system. In [55, 62], a modified version of an aging Markovian process with explicit sojourn time distribution is introduced, where an aging factor discounts the probabilities of staying at the current state while increasing the probabilities of transitions to less healthy states. Although DTAMP aims to address the limitation of Markovian structures to some extent, finding an aging factor that can reasonably discount (over time) the probability of staying at each state and the independence of transition distributions to the age of the device at each state (sojourn time) and the total age at the same time are the main limitations of using DTAMP in multistate degradation modeling.

In multistate degradation with a continuous-time aging Markovian structure as reported in [32], the degradation between two states depends on the states involved in the transition and the total age of the device. Also, the degradation transition can happen at any time point (continuous-time stochastic process). This type of transition is also referred to as nonhomogeneous continuous-time Markov process (NHCTMP). The term nonhomogeneous refers to the dependency of the transition rate on the age of the device, that is, transition rate changes with the age of the device. Although this structure is more reasonable than a continuous-time Markov process for degradation modeling, the assumption of independency of transition distributions to the time that the device reached the last state (sojourn time) and the age of the device at each state is the main limitation of using NHCTMP in multistate degradation modeling [32, 64, 65]. Despite its flexible structure useful for degradation modeling, it is used only in few reported research studies on degradation modeling.

In multistate degradation models with an explicit-duration semi-Markov structure as reported in [54, 66] for EDDTSMP and in [67] for EDCTSMP, the device stays at its current state i following an arbitrary (but explicitly defined) distribution and then transits to a degraded state j with a constant probability of $p_{i,j}$. The terms discrete and continuous refer to the time-dependent sojourn time distribution at each state. For both EDDTSMP and EDCTSMP, the fixed transition probabilities (independent of the age of the device) limits the use of these models in practical cases [15, 55, 62]. Furthermore, finding an explicit type of sojourn time distribution for each state is not always a straightforward process.

In multistate degradation with a homogeneous semi-Markov structure as reported in [34, 68] for HDTSMP and in [7, 9, 69, 70] for HCTSMP, the device degrades according to a time-dependent transition rate (or transition probability for HDTSMP), which changes with the time spent at the current state. The term homogeneous here refers to the fact that the associated degradation process from state i to state j starts only at the time point the equipment reaches state i, that is, it is independent of the total age of the device (it depends on the sojourn time at state i). This type of transition rate has also been used in [71, 72]. Based on this structure, the sojourn time distribution at each state follows an arbitrary distribution. Although, compared to Markovian degradation, HDTSMP and HCTSMP comply with more realistic cases by taking the age of the device at each state into consideration, they are not applicable when the degradation process depends not only on the sojourn time at each state but also on the total age of the device.

In multistate degradation using conventional proportional hazard models, the degradation transition to the failure states (failure rate or hazard rate) is defined as a function of the total age of the device represented by a baseline hazard function (such as Weibull hazard rate) and a covariate process Z, which is multiplicatively related to the hazard rate. The covariates can be internal covariates, which are related to the actual degradation process, or external covariates, which are not related to the degradation process (such as environmental factors). As a result, this covariate process can also affect the time to failure. For example, in [41, 73, 74], a proportional hazard model was considered for the failure transition in a device with multiple health conditions where the covariate vector was defined based on the current state of the device. The degradation transitions between states were modeled with DTMP leading to fixed transition probabilities. The main limitation of using this type of proportional hazard models for multistate degradation modeling is that the probability of transition from state i to a degraded state j is independent of the age of the device, that is, it is constant $P_{i,j}$. Another limitation of this type of proportional hazard model used in reliability modeling is that hazard rate is defined in a continuous domain (through a hazard rate function), that is, the failure can happen at any time point. However, the intermediate transitions between degrading states (the states of the covariate process) are defined by a discrete-time Markov chain, which means transitions are allowed only at discrete time points.

The use of any of the above-reviewed stochastic models for multistate degradation modeling is subject to two common limitations. First, all degradation transitions should follow an identical structure, that is, degradation transition between states should have an identical statistical form. This is rarely the case as a degradation transition between health states may follow different structures. For example, for a four-state device, the transition rate from state 1 to state 2 might be a constant value (independent of the age of the device), but the transition rate from state 1 to state 4 might be age-dependent. Second, each degradation transition may depend only on some (not all) of the following factors, namely, the two states involved in transition (the actual level of degradation), the time point that the device reached the current state, the time spent at the current state, the total age of the device, or any other covariates. For example, for a four-state device, the degradation transition between state 2 and 3 at time t may depend on the time spent at state 2, the total age of the device, and the operational condition of the device at time t ($Z(t)$). In order to provide a more general and flexible degradation structure that can include more practical cases, the above limitations need to be relaxed, in the sense that: (1) degradation transitions between states in a single device can follow non-identical structures and (2) transitions between states can depend on any combination of the above-described factors, that is, a degradation transition rate as well as the overall degradation process may be affected by any of the above-mentioned factors. For example, for a single device, the overall degradation process may depend on the total age of the device as well as the time spent at each state [75].

It is interesting to note that the only stochastic process that can directly address the above-mentioned limitations is the nonhomogeneous semi-Markov structure. Surprisingly, very limited amount of research work has been devoted to multi-state degradation modeling using either nonhomogeneous discrete-time semi-Markov process (NHDTSMP) or nonhomogeneous continuous-time semi-Markov process (NHCTSMP). Due to the flexibility, generality, and more practical applications of this structure, it is considered as the structure of the multistate degradation model in this thesis. In addition, as degradation transitions naturally evolve in the continuous-time domain and as the discrete-time case can be obtained from the continuous one, by considering counting measure for discrete time points [76], it is assumed in this thesis that the multistate degradation process follows a nonhomogeneous continuous-time semi-Markov process (NHCTSMP) evolving in a discrete-state space. To be able to describe transition rates in a more general form, we denote the transition rate between state i and j as $\lambda_{i,j}(s,t)$, where t is the sojourn time at state i and s is the time point that the equipment entered state i. With the above definition of transition rates, the overall degradation process can change with the level of degradation (states involved in transitions), the time spent at each state, the total age of the device, or any other dependent covariate ($Z(t)$). This general stochastic process covers the previously discussed stochastic processes used for degradation modeling (in continuous-time domain). The mathematical structure of NHCTSMP and its application in multistate degradation modeling are discussed in more details in Chapter 3.

It is important to note that if we consider the observation process, which is a stochastic process reflecting indirectly the unobservable degradation process, the term hidden can be used, so that above-described degradation models can be referred to as discrete-time hidden Markov process (DTHMP), continuous-time hidden Markov process (CTHMP), discrete-time hidden aging Markovian process (DTHAMP), continuous-time hidden aging Markovian process (CTHAMP) or alternatively called nonhomogeneous continuous-time hidden Markov process (NHCTHMP), explicit-duration discrete-time hidden semi-Markov process (ED-DTHSMP), explicit-duration continuous-time hidden semi-Markov process (ED-DTHSMP), homogeneous discrete-time hidden semi-Markov process (HDTHSMP), and homogeneous continuous-time hidden semi-Markov process (HCTHSMP), respectively. In the next section, the observation process and its role in indirectly reflecting the hidden degradation process are discussed. In Table 2.1, the references in which multistate degradation modeling is employed are reported.

Table 2.1: References for multistate degradation

Structure	Degradation Modeling
DTMP	[49, 50, 51, 52, 53, 77]
CTMP	[57, 58, 59, 60, 61, 78]
DTAMP	[56, 62, 63]
CTAMP (NHCTMP)	[32, 64, 65]
EDDTSMP	[54, 66]
EDCTSMP	[16, 67]
HDTSMP	[34, 68, 79]
HCTSMP	[7, 9, 32, 69, 70]

2.1.3 Observation Process for Multistate Degradation Models with Unobservable States

As discussed earlier, this thesis focuses only on multistate degradation processes with discrete-state space, continuous-time domain, and unobservable degradation process where states are only indirectly and partially observable through condition monitoring. This type of degradation process is referred to as a hidden process. Condition monitoring techniques include a wide range of applications from analysis of engine oil to vibration monitoring. The most common approach for data acquisition in real applications of condition monitoring frameworks is that the observations are taken at discrete, not necessarily equidistant time epochs, while the measurements extracted at observation points are in a continuous range [80]. The output of this condition monitoring framework is referred to as the observation process.

As discussed earlier, the observation process is generally not perfect due to the existence of noise and the nature of the observation process, that is, it does not directly reveal the actual degradation level of the device under operation. In other words, it is impossible to reveal the actual degradation state of the device just from the extracted condition monitoring data. This means that different condition monitoring indicator values may exist at the same degradation state (one-to-some relationship). Also, a certain observation value may be observed at different levels of health states (some-to-one relationship). That is why the hidden degradation process is also referred to as a partially observable degradation process through the observation process. Therefore, the relationship between the collected condition monitoring information (observation process) and the degradation level (degradation process) is some-to-some. This thesis assumes that the observation process has a stochastic relationship with the actual health status of the device (degradation process), that is, it can be used to indirectly monitor the degradation process. More details on the characteristic of this stochastic relationship are given in Chapter 3.

In a condition monitoring framework, the observation process may be able to indirectly reflect the health status of a device through monitoring a single or multiple indicators. These indicators or alternatively called features can be calculated from raw condition monitoring data extracted from single or multiple sensors. One of the primary objectives of condition monitoring is to find a stochastic relationship (some-to-some) between health states (degradation process) and the output of the observation process. With respect to the statistical form of the stochastic process used to represent the relationship between the degradation and observation processes, two types of observation processes may exist: (1) observation process with parametric distribution and (2) observation process with nonparametric distribution. In a parametric distribution, the relationship between actual health states (degradation process) and condition monitoring data (observation process) is represented by a well-known parametric distribution (such as multivariate Gaussian mixture distribution). In the case of multivariate Gaussian distribution, the observation process is assumed to be a random vector with the following distribution:

$$\Pr(Y_t = y | Z_t = i) = \sum_{k=1}^{M} W_{i,k} \mathcal{N}(y, \mu_{i,k}, \sum_{i,k}), \tag{2.1}$$

where Y_t and Z_t are respectively the output of the degradation and observation processes at time t, M is the number of Gaussian mixture, $W_{i,k}$ is the weight of the kth mixture at state i, and $\mu_{i,k}$ and $\sum_{i,k}$ are the mean and covariance matrix of the kth mixture at state i, respectively. The relationship between the degradation and observation processes is fully defined if the parameters of each Gaussian distribution are known. Examples can be found in [81, 82, 83, 84, 85].

On the other hand, in the case of nonparametric distribution, the relationship between the degradation and observation processes is represented in a nonparametric discrete form through a matrix called observation probability matrix (OPM). In other words, the observation process is assumed to take only certain discrete values and therefore a simple nonparametric probability mass function can be used to represent the relationship between each state and possible discrete outputs of the observation process. Here, observation probability matrix represents, in a nonparametric discrete form, the statistical relationship between the actual health state of the device and the condition monitoring data. The number of rows in this matrix is equal to the number of health states, and the number of columns is equal to the number of possible values of a condition monitoring feature. Each row of the observation probability matrix represents the distribution of the observation process at a certain level of health states. Many examples of such type of distribution are reported in [15, 54, 55, 73, 74, 86, 87]. More details regarding this type of observation process are discussed in Chapter 3.

The main challenge of considering a parametric distribution for the stochastic relationship between the degradation and observation processes is the complexities in finding a known parametric distribution that appropriately represents this relationship. On the other hand, the nonparametric representation of condition monitoring data entails the challenge of how to define discrete levels of outputs and how to superimpose original feature values to these discrete levels. In the literature, dimension reduction methods such as Vector Quantization (VQ) techniques are employed to convert original observation values to discrete groups or clusters [88]. The remaining challenge is how to evaluate possible ways to map a continuous or discrete distribution to a discrete distribution with a limited number of distinct values (symbols), which can efficiently represent the original observation distributions without losing much information. Also, for both parametric and non-parametric cases, finding the best set of condition monitoring features as the representative of the observation process is very challenging.

It can be concluded that finding an efficient way to select the best condition monitoring feature among possible features, determining the parametric form of the distribution to be used to represent the stochastic relationship between the degradation and observation processes, and determining the number of possible outputs to represent the observation process (in a nonparametric form) are the main challenges of available models used to represent this relationship. In this thesis, similar to many other work in the domain of condition monitoring, it is assumed that the relationship between the degradation and observation processes is represented in a nonparametric discrete form through an observation probability matrix. This nonparametric representation of the relationship between degradation and observation processes can reduce modeling bias by considering no specific model structure. In addition, it is often more useful when limited information is available or flexibility about the observation process is required. As will be discussed in Chapter 4, several methods such as clustering can be used to investigate a reasonable number of discrete points to be used to represent the observation process. However, developing an exact method to find the optimal value for the number of discrete outputs of the observation process is out of the scope of this thesis.

2.2 Structure Modeling and Parameter Estimation

The primary step before modeling degradation and observation processes associated with a device is to determine the structure of the multistate model. Structure determination or alternatively called model selection involves two steps, configuration (or topology) determination and parameter estimation. The purpose of the configuration selection step is to determine a reasonable topology for the associated multistate model with unobservable states. The purpose of the parameter estimation is to find a reasonable set of parameters for the selected topology using historical observation data. Configuration determination and parameter estimation are the primary steps before using multistate degradation models for diagnostics and prognostics.

The main elements that characterize a multistate topology are the number of states (N), transition diagram (Ω), transition types (ξ), statistical structure of transition rates (λ), condition monitoring feature (I), and the number of clusters to be used for feature representation (V). Therefore, the multistate topology can be denoted as $\zeta = \{N, \Omega, \xi, \lambda, I, V\}$. The detailed description of these elements will be illustrated later in Chapter 3. With regards to the parameter estimation step, parameters to be estimated are the ones that characterize the degradation process (Γ) and observation process (B). Therefore, $\theta = \{\Gamma, B\}$. These two sets of parameters will be further illustrated in this section.

Among the above two steps, the step of parameter estimation has received more attention in the domain of multistate modeling. There are three possible reasons for this: (1) finding a reasonable multistate topology usually requires large amount of data (6 elements need to be defined), (2) comparing possible topology alternatives is very time-consuming (due to the large number of alternatives), and (3) there is usually some prior knowledge on the elements of the multistate topology (such as historical information from similar devices). Due to the above reasons, most available research works have assumed that the topology of the multistate structure is known and they just focused on parameter estimation methods to estimate the unknown parameters of the selected topology. For example in [73], the number of states (N), transition diagram (Ω), transition types (ξ), statistical structure of transition rates (λ), condition monitoring feature (I), and the number of clusters to be used for final feature representation (V) are assumed to be known and only finding characteristic parameters for degradation and observation processes needs to be investigated. In real-world applications, finding an appropriate multistate topology, which can reasonably represent the actual degradation and observation processes, is very challenging. To avoid unnecessary maintenance actions and prevent catastrophic failures, it is crucial to understand the elements of the multistate structure and find a reasonable multistate model, which can be trusted to be used for online health monitoring.

As mentioned earlier, the literature on multistate topology selection is still at its infancy. There are very limited reported work that this challenge is addressed. An example can be found in [89], where the reasonable number of states for the hidden Markov model is studied. That is why the literature review in the following subsection is limited to the research works on parameter estimation of multistate degradation models. Later in Chapter 3, the elements of the multistate topology are clearly defined, and then in Chapter 4, a simple framework is proposed, which can be used to estimate the unknown parameters of a known multistate topology. It should be pointed out here that finding the best topology for the multistate model (the 1st step of model selection) is out of the scope of this thesis.

2.2.1 Parameter Estimation

In order to develop diagnostic and prognostic models for devices under degradation with partially observable states, stochastic models can be used to model the associated degradation and observation processes. Since in practical problems, applying diagnostic and prognostic models requires estimating the characteristic parameters of the associated stochastic models from real-time data, parameter estimation models and algorithms are of high interest in diagnostic and prognostic analyses of mechanical devices. It should be pointed out here that the parameter estimation step is done only after the topology of the multistate model is fully defined. The two sets of parameters of interest considered in parameter estimation of multistate degradation models with partially observable states are: (1) the set of parameters characterizing the degradation process and (2) the set of parameters characterizing the observation process.

These two sets of characteristic parameters and their elements may vary, depending on the topology selected for modeling the multistate degradation process. The procedure of parameter estimation for multistate degradation structures with unobservable states is usually an unsupervised estimation process, where the direct information on the degradation process is not available. However, if it is possible to obtain the actual degradation levels continuously or at some discrete points (such as inspection points, major shut downs, or overhaul times), then supervised estimation techniques may be employed. Laboratory experiments [15], simulation [90], and accelerated run-to-failure experiments [91] are helpful approaches that can provide run-to-failure data required in a supervised estimation method for a device with unobservable states. Such strategies are out of the scope of this thesis.

In this thesis, the focus is on the unsupervised estimation method, in which the actual degradation data is not available and instead only indirect information obtained from the observation process can be used for estimation. Several research works have considered the problem of parameter estimation for multistate structures with unobservable states. It should be pointed out here that most reported parameter estimation work in the domain of multistate modeling have employed the techniques used in the literature for training hidden Markov and hidden semi-Markov models as given in [92]. For example, the Expectation-Maximization (EM) technique, which is widely used in the training of hidden Markov (HMM) and hidden semi-Markov models (HSMM) has been also employed in the parameter estimation of multistate degradation models. Examples can be found in [93, 94].

Table 2.2 summarizes the references available on parameter estimation for different types of multistate structures. It can be observed from this table that there are less work available on parameter estimation for semi-Markovian multistate structures. In addition, there is no reported work on parameter estimation of homogeneous and nonhomogeneous continuous-time hidden semi-Markov degradation models. As NHCTHSMP is a more reasonable and flexible structure for degradation and observation process modeling, there is a need for developing an unsupervised parameter estimation method for estimating the parameters associated with a NHCTHSMP, which can use available historical condition monitoring data for parameter estimation. Such a parameter estimation procedure will be reported in Chapter 4 of this thesis.

Table 2.2: Summary of research work on parameter estimation

Structure	Parameter Estimation
DTHMM	[92, 95, 96, 97, 98, 99]
CTHMM	[57, 58, 80, 100]
DTAMP	[56, 62, 63]
CTAMP (NHCTMP)	-
EDDTHSMM	[54, 66, 101, 102, 103]
EDCTHSMM	[104]
HDTHSMM	[34]
HCTHSMM	-
NHCTHSMM	-

2.3 Diagnostics and Prognostics for Multistate Structures

With the development of advanced sensor technology over recent years, condition monitoring techniques have received a significant amount of attention. The ultimate goal of condition monitoring techniques is to explore the relationship between the actual level of degradation (degradation process) and the extracted condition monitoring data (observation process), which can finally enable decision makers to find the most efficient and cost-effective maintenance action, e.g. whether or not to replace a device at a certain condition.

Generally, the observed condition monitoring data can be used for two purposes: (1) diagnostic health monitoring and (2) prognostic health monitoring. Diagnosis and prognosis are known as the two important aspects of condition-based maintenance (CBM) [17]. Diagnosis refers to activities that are performed while the device is operating and the degradation process is evolving. The purpose of diagnostic health monitoring is to provide affirmative information on the current heath status (actual level of degradation) of the device before it fails. On the other hand, prognostics refers to activities, which help to predict future fault propagations or failures. The purpose of prognostic health monitoring is to predict the future trend of the fault propagation and degradation process in the sense that timely and cost-effective maintenance actions can be performed. There are many published research works available on machinery diagnostics and prognostics. A review on diagnostic and prognostic methods using condition monitoring data can be found in [18, 19, 20].

In real-world systems, diagnostic and prognostic health monitoring is done through monitoring several dynamic reliability measures reflecting certain aspects of the degradation process. These dynamic performance measures are calculated using the condition monitoring data extracted from the observation process. Here, dynamic means that the measure is calculated over time given the most updated profile of the condition monitoring data. Therefore, dynamic diagnostic performance measures employ all condition monitoring data up to the current time to provide information on the current actual level of degradation (health states) while dynamic prognostic reliability measures employ such data to provide information on the future stochastic behavior of the device with respect to degradation and failure. Although the theory of multistate modeling has been widely used in the literature for reliability analysis, its application for diagnostic and prognostic health monitoring of condition-monitored devices with an unobservable degradation process is still at its infancy and needs to be further investigated. In the following, relevant research studies where multistate degradation process is employed for diagnostic and prognostic health monitoring are reviewed.

2.3.1 Diagnostic Health Monitoring for Multistate Devices

Using condition monitoring data for diagnostic health monitoring has been reported frequently in the literature. The most commonly used measure for this purpose is the current degradation level of the device (state level), which can be calculated using methods such as Viterbi [92]. The Viterbi algorithm is a dynamic programming algorithm for finding the most likely sequence of states from a sequence of observed data. In [60], several measures, such as the probability that the system will be found in a functioning state at time t, expectation of performance output at time t, and time to reach the failure state for a multistate Markovian degradation process are proposed. Shu et al. [70] developed dynamic performance measures for machine tools with multistate degradation processes in the sense that the total experience of the manufacturer over the target life of the tool is employed to calculate performance measures. In [50], HMM and neural network methods are used to estimate the current level of tool wear as a diagnostic measure. Tobon-Mejia et al. [105] used a mixture of Gaussian hidden Markov model for failure diagnostics and prognostics. They estimated the asset's current health state, its remaining useful life, and the associated confidence degree using condition monitoring data.

Our review concludes that most available research studies are subject to at least one of the following three shortcomings. First, many reported studies such as [106] consider only important reliability measures calculation for Markovian structure. This means that important reliability measures applicable for non-Markovian degradation structures need to be developed. Second, traditional reliability research studies are limited to calculating the static reliability measures, such as unconditional reliability function, that is, the dynamic behavior of the degradation process is ignored ([60]). Finally, many reported research studies assumed that the degradation process is directly observable ([107]). As a result, a practical and efficient approach for diagnostic health monitoring of devices under unobservable degradation and condition monitoring through developing dynamic reliability measures obtainable from condition monitoring data needs to be developed. In other words, important dynamic performance measures, which have the ability to reflect the dynamic behavior of the degradation process, need to be developed for a multistate degradation process under NHCTHSMP. Using efficient diagnostic performance measures enables maintenance decision making to detect the degradation progression and therefore, the failure of system could be prevented. That is why diagnostic measures can play important roles in the life cycle cost minimization of devices under gradual degradation. In Chapter 5, the mathematical derivations for important diagnostic measures, such as the conditional state probability, average degradation level, its associated confidence interval, and hazard rate are introduced for a degradation process under NHCTHSMP.

2.3.2 Prognostic Health Monitoring for Multistate Devices

Accurate estimation of prognostic reliability measures, such as reliability and remaining useful life (RUL), is playing an important role in minimizing the overall maintenance cost of systems. Prognostic reliability measures can help maintenance decision makers to determine the time to initiate maintenance setup and the time to replace a degraded device and therefore can decrease unnecessary maintenance actions and unexpected failures. As the ultimate goal of estimating prognostic performance measures, such as remaining useful life, is to prevent failures, they are of significant interest to maintenance decision makers. Over the past years, there has been a significant increase on methodologies regarding prognostic reliability evaluation of mechanical systems. A comprehensive literature review of recent research work in this area and advantage and disadvantage of several available models can be found in [22]. In addition, the strengths and weaknesses of the main prognostic models and summary on how each can be applied to engineering prognostics have been discussed in [23].

The idea of using condition monitoring data to predict future behaviour of a device has been investigated widely in the literature. Ghasemi et al. [74] considered a proportional hazard model to estimate the remaining useful life for a discrete multistate system with specific state transitions where a transition was possible from all levels of health states to the failure state. Dong and He [108] introduced diagnostic and prognostic analysis of an explicit duration discrete-time hidden semi-Markov degradation process and developed a modified forward-backward and a re-estimation method to estimate the unknown parameters of the model. For the estimation of the sojourn time (the probability of duration at each state), they considered a discrete-time nonparametric distribution.

The main limitation of current studies in prognostic analysis in the multistate domain is the lack of mathematical derivations for important condition-based (dynamic) prognostic measures that (1) can be calculated from real-time condition monitoring data (observation process) and (2) provide useful information to maintenance decision makers that can help determine whether or not to initiate maintenance setup or whether or not to replace a degraded device. Such prognostic measures have not been developed for a degradation process evolving according to NHCTHSMP. In Chapter 5, the mathematical derivations for well-known dynamic prognostic reliability measures, such as conditional reliability, conditional remaining useful life, and warning levels are developed for a degradation process under NHCTHSMP.

2.4 Maintenance Decision Making

Conventional reliability and maintenance frameworks rarely consider the actual health status of the device while operating in decision making. They are usually either a corrective maintenance framework or a time-based preventive maintenance framework [109]. In corrective maintenance of non-repairable systems, a device is replaced only after failure. In a time-based preventive maintenance, a predefined time threshold, irrespective to the actual degradation process, is considered to determine when to replace a device. With the development of sensor technologies and condition monitoring techniques, advanced reliability studies have focused more on the development of methodologies and frameworks that can somehow employ actual information regarding the degradation process for efficient online health monitoring of systems under gradual degradation. It has been proved that better maintenance actions can be made if actual health status (degradation status) of the device over time is employed for maintenance decision making [110]. This type of maintenance decision making is also referred to as condition-based maintenance (CBM). By taking maintenance actions only when there is an evidence of abnormal behaviors of a physical asset, CBM can also avoid unnecessary maintenance actions [17].

The concept of condition-based maintenance has been applied extensively in the domain of reliability and maintenance. Comprehensive reviews of reported models on machine condition-based maintenance can be found in [17, 20, 111, 112, 113]. There are also reported research works on the application of unobservable multistate degradation structures for condition-based maintenance decision making. Makis and Jiang [114] developed an optimal replacement strategy for a multistate system considering continuous-time hidden Markov model with known exponential transition rates. The cost of maintenance in their work was assumed to depend on the state of the system. Maillart [115] introduced a heuristic solution for a condition-based maintenance optimization problem. They considered three choices for the types of maintenance, which are do nothing, replace, and collect observations. The objective function in that work was to minimize the long run unit average cost of the system, where failure state is observable. The simple discrete-time hidden Markov model is considered in [116] for a multistate system. The observation probability matrix was assumed to follow a binomial distribution. They investigated replacement policy (replace or do nothing) and introduced an approximation method for their work. The objective was to minimize the total cost of the system.

Chen and Wu [56] developed a dynamic maintenance policy (preventive maintenance or do nothing) where the objective function was to minimize maintenance and downtime costs for a system under a discrete semi-Markov structure with aging Markovian deterioration. Rosenfield [117] developed an optimal maintenance policy (repair, no action, and inspection) for a multistate system under a special discrete-time Markov structure with incomplete information. A replacement policy and control limits for a multistate system were considered in [118] where a dynamic programming model for the optimization problem was proposed.

Ivy and Pollock [119] analyzed an optimal maintenance policy for a multistate system with deteriorating states and considered a silent failure where the system can continue working in the failure state. It was assumed that the effect of imperfect maintenance depends on the state of the system. They used the concept of partially observed Markov decision process for their analysis. Transitions from states to states were assumed to follow a geometric distribution. Tomasevicz and Asgarpoor [120] developed a method for optimal preventive maintenance of a multistate system under repair. They used continuous-time Markov structure for their system.

There are also published case studies where condition-based maintenance was applied and associated results were presented. Byon and Ding [121] examined optimal repair strategies for wind turbines operated under stochastic weather conditions. Their objective was to derive an optimal preventive maintenance policy that minimizes the expected average cost over an infinite horizon using information obtained from multiple sensors. A case study on a helicopter gearbox based on HMM is presented in [122]. They proposed a method for parameter estimation and condition-based maintenance. In [41], a condition-based replacement policy for a system under degradation was developed and then a dynamic programming approach was employed to solve the problem.

It should be pointed out here that the above review excludes other types of maintenance management frameworks, such as reliability centered maintenance (RCM) and risk-based maintenance (RBM). Interested readers may refer to [123, 124]. As reviewed in this subsection, the concept of condition-based maintenance has been applied widely in theory and application. However, there is no reported work on developing CBM techniques applicable in unobservable multistate degradation model evolving according to a NHCTHSMP. Also, a method that can determine the optimal time of maintenance setup initiation needs to be developed.

2.5 Concluding Remarks

Based on the conducted literature review, important shortcomings and challenges of available research studies for multistate degradation modeling with unobservable states are summarized in this section. These challenges are classified into four different categories. For each category, the contribution made in this thesis to address a particular challenge is discussed.

Challenge I) Current stochastic models used for multistate degradation modeling have the limitation that transitions between states do not depend simultaneously on the states involved in transitions, the time spent at the current state, the total age of the device, and other covariates that depend on these factors. Therefore, there is a need for a more general stochastic process model to be used for multistate degradation modeling that can address the above-mentioned shortcoming. This general stochastic process, referred to as the nonhomogeneous continuous-time hidden semi-Markov process (NHCTHSMP), is employed throughout this thesis to represent degradation and observation processes of condition-monitored devices. Using such a general and flexible structure, the overall degradation process may depend on the actual level of degradation, the time spent at each state, the total age of the device, and any other covariate dependent on the aforementioned factors. Also, degradation transitions between states will have the flexibility to follow different structures. The fundamental properties of a NHCTHSMP will be described in Chapter 3 of this thesis. In addition, how this flexible structure can be used to model the degradation and observation processes associated with a condition-monitored device will be demonstrated.

Challenge II) The problem of employing historical condition monitoring data for the purpose of model selection and parameter estimation procedure has not been studied for a multistate degradation model with unobservable states evolving according to NHCTHSMP. Such procedures can provide maintenance decision makers a guideline on how to model a condition-monitored degradation process with a multistate structure using historical data. In Chapter 4 of this thesis, a framework that can employ condition monitoring data for training a multistate degradation structure with unobservable states will be proposed.

Challenge III) Although there are several diagnostic and prognostic methods available for devices with multiple health states, to the best of our knowledge, efficient dynamic diagnostic and prognostic health monitoring measures that can use condition monitoring data for monitoring a multistate degradation process evolving according to NHCTHSMP do not exist. Such performance measures can enable maintenance decision makers to be able to continuously monitor the degradation process. In Chapter 5 of this thesis, the mathematical formula for important dynamic diagnostic and prognostic performance measures are introduced, which can employ online condition monitoring data to provide useful information on the actual level as well as the future behaviour of the degradation process.

Challenge IV) A dynamic condition-based replacement model, which can employ online condition monitoring data to determine when to initiate maintenance setup and when to replace a degraded device, needs to be investigated. Unlike artificial neural networks or data-driven classification methods with the black box modeling structure, multistate degradation models based on a NHCTHSMP can be used as a reasonable tool to determine the optimum time to replace a degraded device. In Chapter 6 of this thesis, a dynamic replacement policy is introduced, which can employ condition monitoring data for online maintenance decision making.

In Chapter 7, a case study on turbofan engines are analyzed and finally the conclusion of this thesis and directions for future work are reported in Chapter 8. It is expected that the result of this thesis will advance the state of the art in the practical health management of mechanical systems. The structure of this thesis is designed in the sense that each of the above challenges is addressed through a particular research topic. Chapter 3 is devoted to address challenge I, which is multistate degradation modeling using NHCTHSMP. In Chapter 4, challenge II is investigated. Chapter 5 investigates important dynamic diagnostic and prognostic measures for health monitoring a degradation process and attempts to address challenge III. In Chapter 6, challenge IV is investigated and a dynamic condition-based model is proposed, which can employ condition monitoring data to determine when to replace a degraded device.

Chapter 3

Multistate Degradation Modeling Using NHCTHSMP

3.1 Introduction

This chapter reviews the basic structure of a general stochastic process, namely, the nonhomogeneous continuous-time semi-Markov process (NHCTSMP) and then illustrates how this stochastic process can be used for multistate degradation modeling when states are hidden (not directly observable). In addition, the main elements of a NHCTHSMP are illustrated in details. The fundamental properties of the NHCTHSMP described in this chapter will be directly used in the remainder of this thesis. This chapter is organized as follows. Section 3.2 illustrates the assumptions made in this thesis for the device under study, the degradation, and the observation processes. The elements of a NHCTHSMP are described in Section 3.3. The mathematical structure of the NHCTHSMP is described in Section 3.4. The fundamental describers of the degradation process are illustrated in Section 3.5. Some important characteristic measures of NHCTHSMP are defined in Section 3.6. Some of the materials of this chapter are published in [26].

3.2 Assumptions

The main assumptions made in this thesis for the device under study and the degradation and observation processes are described in this subsection. For the complete list of notation used in this thesis, see Notation List. These assumptions also set the scope of this thesis with regards to practical applications. It should be pointed out here that the term device used throughout this thesis can refer to: (1) a single element (component) of a mechanical system, such as an impeller in a pump

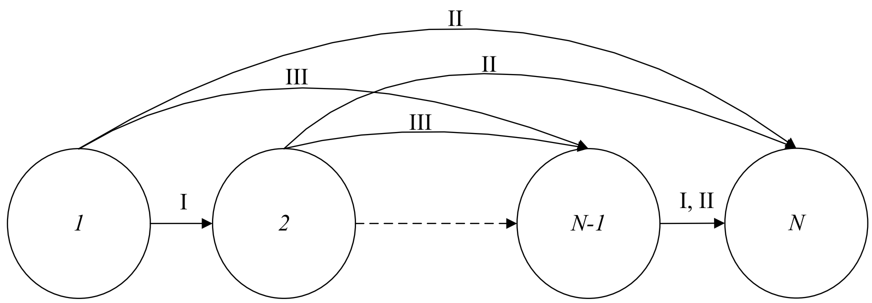

Figure 3.1: Three types of transitions

or a gear in a planetary gearbox, or (2) a piece of equipment with multiple elements (components), such as a pump or a compressor, for which only the overall degradation process is under study (not the individual degradation processes of each component).

I. The device has N known possible discrete levels of degradation states ranging from the perfect functioning state (state 1) to the complete failure state (state N). Each health state may reflect a certain level of damage, operational performance, efficiency, and/or physical properties.

II. Except the failure state, which is assumed to be self-announcing, the state of the device is only indirectly observable through the observation process.

III. At any level of health states, the device can degrade according to three types of transitions, which are: (I) transition to the neighbor state (progressive degradation or soft degradation), (II) transition to the failure state (sudden degradation or hard failure), and (III) transition to any intermediate states (multi-step degradation). Therefore, at each state, the device is subject to multiple competing deterioration processes. See Figure 3.1 for the schematic view of these three types of transitions for a device with N states.

IV. Transition rate functions are used as the main describer of the degradation process. Each degradation transition can follow an arbitrary distribution. Degradation transitions between two states may depend on the states involved in the transitions (level of degradation), the time spent at the current state, the total age of the device, or any combination of these factors. In addition, transitions can depend on any covariate that depends on the above elements. As will be shown in Section 3.2, the degradation transitions are represented by transition rate functions using the general formula given in Eq. (3.3). NHCTSMP structure is used to model the degradation process.

V. The device is under condition monitoring. The condition monitoring process (observation process) does not directly reflect the actual level of degradation. A single condition monitoring indicator is used for health monitoring. This single indicator can be the output of a feature fusion (feature combination) process, which transforms a set of indicators to a single indicator. This indicator is calculated at certain discrete points referred to as observation points (condition monitoring points).

VI. The time between two observation points is small enough, so that at most one transition may occur in the interval between two observation points.

VII. The condition monitoring indicator can take one of the m possible outcomes denoted by v_1 ,v_2 ,...,v_m.

VIII. The condition monitoring indicator (observation process) is stochastically related to the actual levels of degradation (degradation process). This relationship is represented by a nonparametric discrete probability distribution referred to as the observation probability distribution (B) with discrete values $(v_1, v_2, ..., v_m)$. In a general form, the probability that v_k is observed when the device is in states j is defined as $b_j(k) = \Pr(Y_t = v_k | Z_t = j), \forall t, 1 \leq k \leq m$.

IX. The device is not repairable (transitions are left-to-right only) and it is as good as new after a failure replacement. This assumption is also applicable for devices for which the repair is costlier than replacement.

It should be pointed out that the above-described assumptions are the common assumptions made throughout this thesis. Additional assumptions for each chapter will be provided when needed.

3.3 Elements of a Multistate Structure

The primary step to use a multistate degradation process for diagnostics and prognostics is to determine a reasonable and efficient structure for the associated multistate model and subsequently find the best set of characteristic parameters for the unknown elements of the selected structure. Therefore, the elements under which the degradation and the observation processes associated with the device are characterized need to be clearly defined. These elements are also referred to as the configuration elements from which the structure of the degradation and observation processes is characterized. For notational convenience, the multistate structure is divided into different sub-elements, which are the multistate topology (ζ) and the set of characteristic parameters associated with the multistate topology (θ). The elements in the multistate topology describe the structure of the multistate model and the set of characteristic parameters characterize the stochastic processes associated with the degradation and the observation processes for a given topology.

In Chapter 4, a parameter estimation method is proposed, which can be used to estimate the characteristic parameters of a known topology. Then, a simple enumerative approach used in this thesis to find a reasonable configuration (topology) is introduced. In the following, the main elements of the multistate topology are described.

3.3.1 The Overall Number of Health States (N)

This element represents the number of discrete levels of degradation for a device under study. This value can be determined with respect to the actual levels of degradation that the device may experience while operating. In other words, it should reflect the evolution of the degradation process over time. Most available studies have assumed that this value is known. For example in [32], it is assumed that a friction drilling device has only five levels of degradation referred to as: sharp, normal wear, micro fracture, macro wear, and breakage. There are also reported work that N is treated as a decision variable [89]. The number of states is the most important elements of a multistate topology as it affects all other elements of the structure. Generally, the size of the multistate structure and the complexity associated with finding important measures are greatly affected by this element. Failure to determine a reasonable value for N may result in misrepresentation of the degradation process, which may lead to unnecessary maintenance actions and catastrophic failures. The two extreme cases for N are $N = 2$ (binary-state levels) and $N = \infty$ (continuous-state levels).

3.3.2 Transition Diagram (Connectivity between States (Ω))

This element defines the relationship (connectivity) between degradation states. As noted in the assumptions, the device is not repairable, that is, transitions are left-to-right only. This means that transitions are possible only to a degraded state (see Figure 3.1). These left-to-right transitions reflect the possible failure trends associated with the device. Typically, two main types of failures are considered in reliability modeling referred to as soft failure and hard failure. Soft failure occurs when the degradation process passes through all levels of health states before failure. Soft failure is the result of soft degradation (see Type I transition in Figure 3.1), that is, the device degrades only one-step to its neighbor state and transitions are always to the neighbor state (one-step progressive transition). Therefore, under soft failure, the device continuously degrades until it reaches a certain threshold, referred to as the failure state. Under such a failure, all states are visited before failure. For example, fatigue crack growth [125], which can be measured by crack length, is a common type of soft failure.

On the other hand, hard failure is defined as the event that the device suddenly stops performing its intended function [126]. In hard failure, the device does not pass through all degradation states. For example, hard failures can be the result of instantaneous stress from a shock process [127]. Under such a failure, direct transition to the failure state is possible from any intermediate state (see Type II transition in Figure 3.1). The most general type of degradation is the case where transition to any degraded state is possible. For example the device may transit from state 1 to state 4 without visiting states 2 and 3. It is worth mentioning that when the device has only soft and hard failures with Markovian transition rates, a well-known distribution called Coxian distribution [128] can be used to model the degradation process. In can be concluded that three possible types of transition may exist: (I) transition to the neighbor state, (II) transition to the failure state, and (III) transition to an intermediate state between the current state and the failure state. For any state i, these connections are denoted by FS_i, which is the set of states directly accessible from state i. See Figure 3.1 for a schematic view of these three types of transition for a N-state device.

3.3.3 Types of the Transitions (ξ)

Transition type refers to the dependency of degradation transitions to the actual level of degradation (states involved in transition), time that states are reached, sojourn time at each state, total age of the device, or any other covariate (dependent to the above factors). The main possible types of transitions considered in this thesis are Markovian transition (Type I), and non-Markovian transition, which may include aging Markovian degradation (Type II), explicit-duration semi-Markovian degradation (Type III), homogenous semi-Markov degradation (Type IV), and non-homogeneous semi-Markovian degradation (Type V). Each of the above types has certain distinguished statistical properties. The main stochastic properties of each of the above transition types are described in Section 2.1.1.

3.3.4 Statistical Form of Transition Rate Functions (λ)

This refers to the statistical form of transition rate functions, which characterizes the selected transition type. In this thesis, transition rates are treated at the fundamental describer of the degradation process. The detailed relationship between transition rates and the degradation process is given later in this chapter. Several explicit types of distributions have been used in the literature to represent transition rates. The most commonly used distributions are exponential, Gamma, Weibull, and Gaussian distribution. For example, when the structure of the multistate degradation process is a homogenous semi-Markov process (Type IV) and Weibull distribution is used to characterize each transition rate, then the transition rate between state i and j is represented as: $\lambda_{i,j}(t) = (\beta_{i,j}/\alpha_{i,j}) \times (t/\alpha_{i,j})^{\beta_{i,j}-1}$, where t is the sojourn time at state i and $\alpha_{i,j}$ and $\beta_{i,j}$ are characteristic elements of the Weibull distribution. It should be pointed out here that for Markovian transitions, transition rates are constant values, that is, transition type directly determines the statistical form of transition rates.

3.3.5 Condition Monitoring Feature Used for Health Monitoring (I)

In the multistate structure considered in this thesis, the final output of the observation process is assumed to be a single indicator having indirect information regarding the actual health status of the device. Here, the notation I is used to denote the selected feature for health monitoring. This indicator (feature) does not necessarily have a monotonic trend with the damage level; however, it is expected to have a stochastic relationship with the actual level of health states. The stochastic relationship between the health states and the selected condition monitoring indicator is shown by a nonparametric discrete probability distribution referred to as the observation probability distribution. This observation probability distribution is represented by an observation probability matrix (OPM). See assumption (VIII) on Section 3.2 for more details on each element of this matrix.

In real-world applications, finding such a single indicator having a stochastic relationship with the actual health levels is challenging. Usually, a set of measurements obtained from multiple sensors is available. Feature fusion techniques aim to combine different feature vectors in the sense that different characteristics of the pattern can be reflected [129]. Finding the best feature may require developing a feature selection process, which itself may need feature fusion (combination) techniques. Investigating such scenarios is out of the scope of this thesis.

3.3.6 The Condition Monitoring (CM) Indicator Space ($V = \{v_1, v_2, ..., v_m\}$)

As discussed earlier, in this thesis, the final CM feature is represented in a nonparametric discrete form. In other words, the set of original observation values needs to be converted to several discrete levels. The element V reflects the possible outputs of the observation process with respect to the selected condition monitoring indicator. The integer m reflects how the selected condition monitoring indicator I is finally represented (number of possible outputs of the observation process). Different methods, such as vector quantization and clustering can be used to determine this value [130]. The key point here is that this number should be selected in an effective way, so that the distribution of the condition monitoring indicator is realistically represented without having to lose important information regarding the observation process. It should be pointed out here that finding an effective way to convert original observation values to discrete levels is out of the scope of this paper, that is, the condition monitoring indicator space is assumed to be known.

Once the above-described elements are determined, the structures of the multi-state degradation process and its associated observation process are fully known.

3.4 Modeling Degradation and Observation Processes Using NHCTHSMP

This section summarizes the fundamental of the nonhomogeneous continuous-time hidden semi-Markov process and its application in multistate degradation modeling. As discussed earlier, a NHCTHSMP deals with two types of processes, which are referred to as the degradation process and the observation process. The degradation process is considered to evolve according to a continuous-time stochastic process with finite state space (E). It can be then represented by a couple (X_n,T_n), where X_n is the true unknown state of the device at the nth transition and T_n is the time (age of the device) at the nth transition.

In homogeneous Markov renewal process (HMRP), the inter-arrival times between two states are assumed to be i.i.d. random variables with an arbitrary distribution while in a nonhomogeneous Markov renewal process (NHMRP), the inter-arrival times between two states are independent random variables with an arbitrary distribution, not necessarily following an identical distribution. The classical Markov renewal model described in [131, 132, 133, 134] has a powerful and flexible mathematical structure, which can be employed for multistate degradation modeling. The most general and fundamental describer of a semi-Markov process is the kernel function, which completely describes the stochastic behavior of a semi-Markov process [7]. The process (X, T) is called a nonhomogeneous Markov renewal process, if its function has the following property [135]:

$$Q_{i,j}(s,t) = \Pr(X_{n+1} = j, T_{n+1} \leq t | X_n = i, T_n = s, (X_c, T_c), 0 \leq c < n) = \\ \Pr(X_{n+1} = j, T_{n+1} \leq t | X_n = i, T_n = s), \forall (i,j) \in E, \quad (3.1)$$

where X_n and T_n are respectively the state and the time at the nth transition and E is the state space. Now, the degradation process at time t denoted by Z_t $(t \geq 0)$ can be defined, where $Z_t = X_{N_t}$ and $N_t = sup\{n : T_n \leq t\}$. Here, N_t is the transition number at which the last state before time t is reached. This degradation process follows the nonhomogeneous continuous-time semi-Markov process.

Now, let $V = \{v_1, v_2, ..., v_m\}$ be the observation space with m possible values and let us also define the random variable $U_n : \Omega \rightarrow V$, where U_n is the output of the observation process at the nth observation point. The relationship between the degradation process and the observation process can be defined as follows:

$$\Pr(U_n = v_j|(U_c, X_c^0), 0 \leq c < n, X_n^0) = \Pr(U_n = v_j|X_n^0 = i) = b_i(j),$$
$$\forall i \in E, 1 \leq j \leq M, \quad (3.2)$$

where X_n^0 is the state of the device at the nth observation point and $b_i(j)$ is an elements of the observation probability matrix (B), which denotes the probability of observing the jth value of the CM indicator when the device is in state i. Let us define $Y_t\,(Y_t \in V, t > 0)$ as the output of the observation process at time t and T_n^0 as the corresponding time of the nth observation point. Then, we have $Y_t = U_{N_t^1}, N_t^1 = sup\{n : T_n^0 \leq t\}$, where N_t^1 is the number of observation points before time t and T_n^0 is the time of the nth observation point. Also, let $Z_t = X_{N_t}$, where $N_t = sup\{n : T_n \leq t\}$. Given that Z_t follows a nonhomogeneous continuous-time semi-Markov process, the (Z, Y) process is a non-homogeneous continuous-time hidden semi-Markov process (NHCTHSMP). This NHCTHSMP will be used in the remaining sections of this thesis as the basic tool for multistate degradation and observation modeling, which includes both degradation and observation processes. As discussed earlier, the process (Z) is the hidden degradation process which is indirectly observable through the observation process (Y). In Chapter 4, details on how to estimate the characteristic parameters of these two stochastic processes will be provided.

3.5 Fundamental Describers of the NHCTHSMP

The idea of using transition rate functions in modeling transitions between states in a semi-Markov process has been employed in research work such as [134, 136]. In this section, a general definition for the degradation transition rate is introduced, which has the flexibility to cover most of the previously studied transition rates in the literature. In this thesis, transition rate functions are treated as the fundamental describer of the degradation transition between states. Then, the relationship between the transition rate function and the kernel function as the other fundamental describer of the degradation process is defined. The stochastic behavior of the degradation process can be fully defined if transition rates between states or the associated kernel functions are known.

As mentioned in Chapter 2, transitions between two states (in continuous domain) can depend on the states involved in transitions, the time spent at each state, the time that the last state is reached, the total age of the device, or any combination of these factors. The main purpose here is to generalize the definition of the transition rate in the sense that for a single piece of device, different types of transitions could exist (see Section 3.3.3 for different types of transitions). For the stochastic process associated with the transition between state i and state j, given that the device is in state i at time u, the instantaneous probability rate that it transits to state j in an infinitesimal time interval $(u, u + du)$ is expressed as $\lambda_{i,j}(s, u)$, where $\lambda_{i,j}$ denotes the transition rate of the process at time u. The following general definition can be used as the transition rate function at time u between state i and state j, given that the device reached state i at time s:

$$\lambda_{i,j}(s,u) = \lim_{du \to 0} \frac{\Pr\{\{u \leq T_{n+1} - T_n \leq u + du\}, \{X_{n+1} = j\} | \{u \leq T_{n+1} - T_n\}, \{X_n = i\}, \{T_n = s\}\}}{du}. \tag{3.3}$$

The above general definition for transition rate has the flexibility to cover the other four main types of transitions described earlier. In the following, Eq. (3.3) is simplified for transition Types I-IV.

Type I - CTMP: As discussed earlier, this type of transition is represented by a constant transition rates $\lambda_{i,j}$, $(i,j) \in E$. Therefore,

$$\lambda_{i,j}(s,u) = \lambda_{i,j}, \quad (i,j) \in E, (s,u) \in [0, \infty). \tag{3.4}$$

Eq. (3.4) verifies that transitions are independent of the sojourn time at the current state and the process time (the total age of the equipment).

Type II - CTAMP: As discussed earlier, this type of transition can be represented by a time-dependent transition rate $\lambda_{i,j}(t)$, $(i,j) \in E$. Therefore,

$$\lambda_{i,j}(s,u) = \lambda_{i,j}(s+u), \quad (i,j) \in E, (s,u) \in [0, \infty). \tag{3.5}$$

Eq. (3.5) verifies that this type of transition deals with those that depend on the two states involved in transitions and the total age of the device.

Type III - EDCTSMP: As discussed earlier, this type of transition can be represented by a time-dependent transition rate $\lambda_i(t)$, $i \in E$, and one-step transition probabilities $p_{i,j}$, $(i,j) \in E$. Therefore,

$$\lambda_{i,j}(s,u) = p_{i,j}\lambda_i(u), \quad (i,j) \in E, (s,u) \in [0, \infty). \tag{3.6}$$

Eq. (3.6) verifies that transitions depend on the two state involved in transition and the time spent at the current state.

Type IV - HCTSMP: As discussed earlier, this type of transition can be represented by a time-dependent transition rate $\lambda_{i,j}(u)$, $(i,j) \in E$. Therefore,

$$\lambda_{i,j}(s,u) = \lambda_{i,j}(u), \quad (i,j) \in E, (s,u) \in [0,\infty). \tag{3.7}$$

Eq. (3.7) verifies that transitions are independent of the time that the equipment enters state i. This type of transition deals with those that depend on the two states involved in transitions and the time spent on the last state (the sojourn time at the last state). This means that transitions are not affected by the total age of the equipment. In other words, the degradation process from state i to state j is initiated only when the equipment reaches state i .

The relationship between the kernel function (Eq. (3.1)) as a describer of the degradation process and the transition rate functions can now be defined as follows:

$$Q_{i,j}(s,t) = \Pr(X_{n+1} = j, T_{n+1} \leq t | X_n = i, T_n = s) = \\ \int_o^{t-s} \lambda_{i,j}(s,u) \exp\left(-\int_0^u \sum_z \lambda_{i,z}(s,x)dx \right) du, \forall (i,j) \in E, (s,t) \in [0,\infty). \tag{3.8}$$

3.6 Important Characteristic Measures

In this section, important characteristic measures reflecting the stochastic behavior of the degradation process are reviewed. These characteristic measures are very useful to determine the type of transition between states [26]. These important measures are used in the remainder of this thesis to calculate diagnostic and prognostic measures (see Chapter 5). The original definitions of these measures are given in [135]. Here, these measures are described in terms of transition rate functions.

The first measure is the embedded transition probability matrix ($P(s) = [p_{i,j}(s)]$), which provides the one step transition probabilities of the nonhomogeneous embedded Markov chain (EMC). An EMC is a discrete-time Markov chain or a jump process, which is the result of considering a Markov process only at the moments upon which the state of the system changes. The (i,j) element of this one-step transition probability matrix represents the conditional probability of transitioning from state i into state j, given that that state i is reached at time s.

$$p_{i,j}(s) = \Pr(X_{n+1} = j | X_n = i, T_n = s) = \lim_{t\to\infty} Q_{i,j}(s,t) = \int_o^{\infty} \lambda_{i,j}(s,u) \exp(-\int_0^u \sum_z \lambda_{i,z}(s,x)dx)du, \quad (i,j) \in E, s \in [0,\infty). \quad (3.9)$$

Now, given that state i is reached at time s, the probability of the state subsequently occupied can be calculated from Eq. (3.9). In other words, $p_{i,j}(s)$ shows the one-step probability of transition between state i and state j. It is worth mentioning that $p_{i,j}(s)$ is independent of s, only if the degradation transition between states i and j does not depend on the time that state i is reached. For example, for a multistate degradation under CTMP, we have $p_{i,j}(s) = \frac{\lambda_{i,j}}{\sum_j \lambda_{i,j}}$ for all $(i,j) \in E$, and $s \geq 0$.

The second measure is the sojourn time at state i, given that state i is reached at time s, which is a random variable with the following cumulative distribution function (CDF):

$$H_i(s,t) = \Pr(T_{n+1} - T_n \leq t | X_n = i, T_n = s) = \sum_{j\in E} Q_{i,j}(s, t+s) = 1 - \exp(-\int_0^t \sum_{z\neq i} \lambda_{i,z}(s,x)dx), \quad i \in E, (s,t) \in [0,\infty). \quad (3.10)$$

The above measure can be used to find the expected sojourn time distribution at each state. The third measure is the conditional sojourn time distribution given that the state subsequently occupied is state j and state i is reached at time s. This random variable has the following CDF:

$$G_{i,j}(s,t) = \Pr(T_{n+1} - T_n \leq t | X_n = i, X_{n+1} = j, T_n = s) = \begin{cases} Q_{i,j}(s, s+t)/p_{i,j}(s), & \text{if} \quad p_{i,j}(s) \neq 0 \\ 1, & \text{if} \quad p_{i,j}(s) = 0 \end{cases}, (i,j) \in E, (s,t) \in [0,\infty). \quad (3.11)$$

Eq. (3.11) verifies that the kernel function can be found if $p_{i,j}(s)$ and $G_{i,j}(s,t)$ are defined over their domains. This means that it is also possible to describe the stochastic behavior of a degradation transition when both the associated embedded transition probability matrix and the conditional sojourn time distributions are known. The following shows how transition rate functions can be calculated from the above-described characteristic measures as:

$$\lambda_{i,j}(s,t) = Q'_{i,j}(s, s+t)/(1 - H_i(s,t)) = (G'_{i,j}(s,t) \times p_{i,j}(s))/(1 - H_i(s,t)), \quad p_{i,j}(s) > 0, (i,j) \in E, (s,t) \in [0,\infty), \quad (3.12)$$

where $Q'_{i,j}(s, s+t) = \dfrac{\partial Q_{i,j}(s, s+t)}{\partial t}$ and $G'_{i,j}(s, s+t) = \dfrac{\partial G_{i,j}(s, s+t)}{\partial t}$.

3.7 Summary

This chapter reviews the basic structure of a general stochastic process, namely, the nonhomogeneous continuous-time semi-Markov process (NHCTSMP) and then illustrates how this stochastic process can be used for multistate degradation modeling when states are not directly observable or alternatively called hidden. In addition, the main elements of a NHCTHSMP are illustrated in details. The fundamental properties of the NHCTHSMP described in this chapter will be directly used in the remainder of this thesis.

Chapter 4

Multistate Structure and Parameter Estimation

4.1 Introduction

The primary step to use the multistate degradation process and the corresponding observation process described in Chapter 3 for diagnostics, prognostics, and maintenance decision making is to determine a suitable structure for the associated multistate model and subsequently find the best set of characteristic parameters for the selected structure. Therefore, structure determination or alternatively called model selection involves two steps, which are (1) configuration (topology) selection and (2) parameter estimation. The purpose of the configuration selection step is to determine a reasonable topology for the associated multistate model with unobservable states. The main elements that determine a multistate topology are the number of states (N), transition diagram (Ω), transition types (ξ), statistical structure of transition rates (λ), condition monitoring feature (I), and the number of clusters to be used for final feature representation (V). Therefore, the multistate topology can be denoted as $\zeta = \{N, \Omega, \xi, \lambda, I, V\}$. The detailed description of these elements is given in Section 3.3. With regards to the parameter estimation step, parameters to be estimated (θ) are the ones that characterize the degradation process (Γ) and the observation process (B). Therefore, $\theta = \{\Gamma, B\}$. These two sets of parameters are further illustrated in this section. The purpose of parameter estimation is to find the estimated values of the set of parameters for the selected topology using historical data. A complete multistate model M is defined if ζ and θ are known as $M = (\zeta, \theta)$.

Although our main focus of this chapter is on the parameter estimation for a selected topology, the configuration (topology) selection process is also discussed. This chapter is organized as follows. First, a parameter estimation method is proposed in Section 4.2, which can employ condition monitoring data to train a multistate model with a known structure. Then, a simple enumerative approach, which can be used to find a structure for the multistate model is presented in Section 4.3. Finally, in Section 4.4, a simulation-based numerical example is employed to demonstrate the correctness and application of the proposed estimation method. The result of this chapter is published in [9, 26].

4.2 Parameter Estimation

As discussed earlier, for a multistate structure considered in this thesis, two sets of parameters need to be estimated in the parameter estimation phase. Depending on the type of available historical condition monitoring information, estimation methods can be classified into supervised estimation methods and unsupervised estimation methods. The data required for estimation in a supervised estimation method includes trajectories of both the degradation and observation processes. To reveal the actual health states of a device over time (degradation process), an inspection process can be implemented. During the inspection time, the system is usually shut down or suspended, and depending on the type of the device being inspected, methods, such as visual inspection and measurement, can be used for health state identification. Examples of such an inspection policy are reported in [78] for periodic inspection and in [137] for continuous inspection.

Directly observing the health states may be too costly and technically complicated, and because of that, unsupervised estimation methods need to be developed for devices with unobservable states. In an unsupervised estimation method, the data required for estimation are only the observation process. The parameter estimation process described in this thesis is an unsupervised estimation method, which employs only the observation process for parameter estimation. It should be pointed out that the assumption here is that the configuration (topology) of the multistate model is known. For example, the number of degradation states, transition diagram, transition types, condition monitoring indicator, and its definition domain are known.

The first group of unknown parameters to be estimated (Γ) deals with parameters that characterize the distributions of transition rates between states. In other words, these parameters characterize the degradation process. It should be noted again that the assumption here is that the structures of the degradation and observation processes are known. Depending on the distribution of transition rate functions between states, the number of unknown parameters for each transition can vary. For example, if a Weibull distribution is used to represent a homogeneous semi-Markov transition between two states, two parameters (shape and scale) need to be estimated for this particular transition.

The second group of parameters to be estimated (B) represents the stochastic relationship between the health state of the equipment and the observation process. In this thesis, this relationship is represented in a nonparametric and discrete form by a matrix called the observation probability matrix (OPM). The entries of this matrix are the unknown parameters of the model. This matrix has N (number of states) rows and m columns (number of observation process output). The entry in the ith row and the jth column of this matrix represents the probability that the jth condition monitoring indicator is observed when the device is in state i.

Let us assume that there are K independent sequences of condition monitoring observations available to be used for parameter estimation. This set of data is referred to as the training set. Each sequence of the observation process includes temporal condition monitoring indicator values, which were extracted while the device was operating. The kth sequence of the observation process (denoted by $O^{(k)}$) has d_k observation points, which are recorded at time $t_1^{(k)}, t_2^{(k)}, ..., t_{d_k}^{(k)}$. We also denote $O_p^{(k)}$ as the condition monitoring indicator value at the pth CM point (time $t_p^{(k)}$) of the kth sequence of the observation process. The unobservable (hidden) state of the device at the pth monitoring point for the kth unobservable sequence of states (Q^k) is denoted by Q_p^k.

The maximum likelihood method (MLE) is used for parameter estimation, so that the estimation problem is formulated in a form of an optimization problem. In the corresponding maximum likelihood optimization problem, the product of the probability of these K observation sequences (joint likelihood of observations) is to be maximized:

$$L = \prod_{k=1}^{K} \Pr(O^{(k)}|\theta) \overset{L'=\log(L)}{\Rightarrow} L' = \log(\prod_{k=1}^{K} \Pr(O^{(k)}|\theta)) = \sum_{k=1}^{K} \log(\Pr(O^{(k)}|\theta)), \quad (4.1)$$

which yields $\theta^* = \arg\max_{\theta} \left(\sum_{k=1}^{K} \log(\Pr(O^{(k)}|\theta)) \right)$. The remainder of this chapter introduces the mathematical steps that need to be done to evaluate Eq. (4.1). In order to solve this equation, the Baum's auxiliary function [92] can be employed, under which the value of the likelihood function can be iteratively improved starting from an initial solution. The Baum's auxiliary function for multiple sequences of observations can be expressed as:

$$\omega(\theta_{old}, \theta) = \sum_{k=1}^{K} \sum_{Q^{(k)}} \log(\Pr(O^{(k)}, Q^{(k)}|\theta)) \times \Pr(Q^{(k)}|O^{(k)}, \theta_{old}), \tag{4.2}$$

where $Q^{(k)}$ is an arbitrary sequence of states with a same length as $O^{(k)}$. It has been shown that maximizing $\omega(\theta_{old}, \theta)$ leads to increasing the likelihood function as $(\theta_{new} = \arg\max_{\theta}[\omega(\theta_{old}, \theta)]) \Rightarrow \Pr(O|\theta_{new}) \geq \Pr(O|\theta_{old}))$ [92]. Now, instead of directly optimizing Eq. (4.1), by considering an initial estimate for θ as $\theta_{old} = \theta_0$, the joint likelihood function can be iteratively improved by maximizing Eq. (4.2). Then, a stopping criterion can be defined to terminate the iteration procedure, if there is no improvement in the maximum likelihood function. Now, the relationship between θ and Eq.(4.2) should be clearly defined. As will be shown below, Eq. (4.2) can be simplified in terms of θ. Since $\Pr(O^{(k)}, Q^{(k)}|\theta) = \Pr(Q^{(k)}|\theta) \times \Pr(O^{(k)}|Q^{(k)}, \theta)$ and $\Pr(O^{(k)}|Q^{(k)}, \theta) = \prod_{t=1}^{d_k} b_{q_t^{(k)}}(O_t^{(k)})$, and $\Pr(Q^{(k)}|O^{(k)}, \theta) = \dfrac{\Pr(Q^{(k)}, O^{(k)}|\theta)}{\Pr(O^{(k)}|\theta)}$, we have:

$$\omega(\theta_{old}, \theta) = \sum_{k=1}^{K} \left(\sum_{Q^{(k)}} \log(\Pr(Q^{(k)}|\theta)) \times \Pr(Q^{(k)}|O^{(k)}, \theta_{old}) \right) + \sum_{k=1}^{K} \left(\sum_{Q^{(k)}} \sum_{t=1}^{d_k} \log(b_{q_t^{(k)}}(O_t^{(k)})) \times \Pr(Q^{(k)}|O^{(k)}, \theta_{old}) \right), \tag{4.3}$$

where $q_t^{(k)}$ is the hidden state at the tth observation point of the k observation sequence. Eq. (4.3) can be divided into two parts as $\omega(\theta_{old}, \theta) = \omega_1(\theta_{old}, \theta) + \omega_2(\theta_{old}, \theta)$ in a way that the first term depends only on the elements of Γ and the second term depends only on the elements of B from $\theta = (\Gamma, B)$. This will enable us to independently estimate these two sets of unknown parameters at each step of the likelihood improvement. To simplify Eq. (4.3), it will be first shown how to calculate $\Pr(Q^{(k)}|\theta)$, which is the probability of a random sequence $Q^{(k)}$.

Let Q_n be an arbitrary sequence of states between the (n-1)th and the nth transitions in $Q^{(k)}$. We can characterize Q_n based on four elements, which are the time at the $(n-1)$th transition (T_{n-1}), the time at the nth transition (T_n), the state at the $(n-1)$th transition (X_{n-1}), and the state at the nth transition (X_n). Thus, the following relationship holds true between $Q^{(k)}$ and Q_n:

$$\Pr(Q^{(k)}|\theta) = \prod_{n=1}^{n_{Q^{(k)}}} \Pr(X_n, T_n | X_{n-1}, T_{n-1}, \theta) = \prod_{n=1}^{n_{Q^{(k)}}} \Pr(Q_n | Q_{n-1}, \theta), \tag{4.4}$$

where $n_{Q^{(k)}}$ is the number of state transitions in $Q^{(k)}$. Now, $\omega_1(\theta_{old}, \theta)$ can be expressed as:

$$\omega_1(\theta_{old}, \theta) = \sum_{k=1}^{K} \left(\sum_{Q^{(k)}} \log(\Pr(Q^{(k)}|\theta)) \times \Pr(Q^{(k)}|O^{(k)}, \theta_{old}) \right) =$$

$$\xrightarrow{\Pr(Q^{(k)}|\theta) = \prod_{n=1}^{n_{Q^{(k)}}} \Pr(Q_n|Q_{n-1},\theta)}$$

$$\sum_{k=1}^{K} \sum_{Q^{(k)}} \left(\sum_{n=1}^{n_{Q^{(k)}}} (\log(\Pr(Q_n|Q_{n-1},\theta))) \times \prod_{n=1}^{n_{Q^{(k)}}} \Pr(Q_n|Q_{n-1}, O^{(k)}, \theta_{old}) \right) =$$

$$\sum_{k=1}^{K} \sum_{Q^{(k)}} \left(\sum_{n=1}^{n_{Q^{(k)}}} \left(\log(\Pr(Q_n|Q_{n-1},\theta)) \times \prod_{m=1}^{n_{Q^{(k)}}} \Pr(Q_m|Q_{m-1}, O^{(k)}, \theta_{old}) \right) \right) =$$

$$\sum_{k=1}^{K} \sum_{Q^{(k)}} \sum_{n=1}^{n_{Q^{(k)}}} (\log(\Pr(Q_n|Q_{n-1},\theta)) \times \Pr(Q_n|Q_{n-1}, O^{(k)}, \theta_{old}) \times$$
$$\prod_{m=1, m\neq n}^{n_{Q^{(k)}}} \Pr(Q_m|Q_{m-1}, O^{(k)}, \theta_{old}). \tag{4.5}$$

Then we have:

$$\omega_{1,1}(\theta_{old}, \theta) = \sum_{k=1}^{K} \Pr(O^{(k)}|\theta_{old})^{-1} \times$$
$$\sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \sum_{t=0}^{d_k-1} \sum_{d=1}^{d_k-t} \Big(\log(Q_{t+d-1}^{(k)} = i, Q_{t+d}^{(k)} = j | Q_{t-1}^{(k)} \neq i, Q_t^{(k)} = i, \theta) \times$$
$$\Pr(Q_{t+d-1}^{(k)} = i, Q_{t+d}^{(k)} = j, Q_{t-1}^{(k)} \neq i, Q_t^{(k)} = i, O^{(k)}|\theta_{old}) \Big). \tag{4.6}$$

Now, we can rewrite Eq. (4.6) in an equivalent form (from optimization perspective) as follows:

$$\omega_{1,1}(\theta_{old}, \theta) =$$
$$\sum_{k=1}^{K} \Pr(O^{(k)}|\theta_{old})^{-1} \times \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} \sum_{a=0}^{d_k-1} \sum_{d=1}^{d_k-a} \log(\varepsilon_a^{(k)}(i,j,d|\theta)) \times \kappa_a^{(k)}(i,j,d,O^{(k)}|\theta_{old}), \tag{4.7}$$

where

$$\varepsilon_a^{(k)}(i,j,d|\theta) = \Pr(X_n = j, t_{a+d-1}^{(k)} < T_n \le t_{a+d}^{(k)} | X_{n-1} = i, t_{a-1}^{(k)} < T_{n-1} \le t_a^{(k)} | \theta), \tag{4.8}$$

and

$$\kappa_a^{(k)}(i,j,d,O^{(k)}|\theta_{old}) = \Pr(X_n = j, t_{a+d-1}^{(k)} < T_n \le t_{a+d}^{(k)}, X_{n-1} = i, t_{a-1}^{(k)} < T_{n-1} \le t_a^{(k)}, O^{(k)}|\theta_{old}). \tag{4.9}$$

In order to simplify the above results, all transitions out of state r can be considered as a group and therefore $\omega_{1,1}^r(\theta_{old},\theta)$ can be constructed as follows:

$$\omega_{1,1}^r(\theta_{old},\theta) = \sum_{k=1}^{K} \Pr(O^{(k)}|\theta_{old})^{-1} \times \sum_{j=1}^{N} \sum_{a=0}^{d_k-1} \sum_{d=1}^{d_k-a} \log(\varepsilon_a^{(k)}(r,j,d|\theta)) \times \kappa_a^{(k)}(r,j,d,O^{(k)}|\theta_{old}), \quad 1 \le r \le N-1. \tag{4.10}$$

It is important to note that each $\omega_{1,1}^r(\theta_{old},\theta)$ contains only transitions out of state r and Eq. (4.10) depends only on the elements of the degradation process (Γ) and is not affected by the elements of the observation process (B). The relationship between the elements of Eq. (4.10) and Γ is described later in this section. Now, the second term of Eq. (4.3), which involves only the elements of B can be expressed as:

$$\omega_{2,1}(\theta_{old},\theta) = \sum_{k=1}^{K} \Pr(O^{(k)}|\theta_{old})^{-1} \times \left(\sum_{j=1}^{N} \sum_{t=1}^{d_k} \Pr(Q_t^{(k)} = j, O^{(k)}|\theta_{old}) \times \log(b_{q_t^{(k)}}(O_t^{(k)})) \right), \tag{4.11}$$

where $\sum_{j=1}^{m} b_i(j) = 1, i \in E$. Eq. (4.11) can be maximized by adding the Lagrange multiplier $\vartheta(\sum_{j=1}^{m} b_i(j) - 1), i \in E$, and setting the associated derivative of $\omega_3(\theta_{old},\theta) = \omega_{2,1}(\theta_{old},\theta) + \vartheta(\sum_{j=1}^{m} b_i(j) - 1), i \in E$ with respect to each $b_i(\omega)$ and ϑ equal to zero. Thus, we set $\frac{\partial}{\partial b_i(\omega)}\omega_3(\theta_{old},\theta) = 0$ and $\frac{\partial}{\partial \vartheta}\omega_3(\theta_{old},\theta) = 0$, which yields to:

$$b_i(\omega) = \frac{\sum_{k=1}^{K}\left(\Pr(O^{(k)}|\theta_{old})^{-1} \times \sum_{t=1}^{d_k}\Pr(Q_t^{(k)} = i, O^{(k)}|\theta_{old}) \times \delta_{O_t^{(k)},\omega}\right)}{\sum_{k=1}^{K}\left(\Pr(O^{(k)}|\theta_{old})^{-1} \times \sum_{t=1}^{d_k}\Pr(Q_t^{(k)} = i, O^{(k)}|\theta_{old})\right)} = \frac{\sum_{k=1}^{K}\left(\Pr(O^{(k)}|\theta_{old})^{-1} \times \sum_{t=1}^{d_k}\gamma_t(i, O^{(k)}|\theta) \times \delta_{O_t^{(k)},\omega}\right)}{\sum_{k=1}^{K}\left(\Pr(O^{(k)}|\theta_{old})^{-1} \times \sum_{t=1}^{d_k}\gamma_t(i, O^{(k)}|\theta)\right)}, \quad (4.12)$$

where $\delta_{O_t^{(k)},\omega}$ is equal to 1 when the tth observation value of $O^{(k)}$ is equal to v_ω, and 0 otherwise. Now, at each step of the stepwise optimization problem, all entries of the observation probability matrix B can be directly estimated using Eq. (4.12). To be able to optimize Eq. (4.10) and use Eq. (4.12) to find θ_{new}, the concept of Expectation-Maximization (EM) [94] can be used. EM algorithm is a well-studied iterative method for finding maximum likelihood estimates of statistical models when the corresponding likelihood equation cannot be solved directly.

4.2.1 The Summary of the Parameter Estimation

The summary of all steps in the Expectation-Maximization method for the unsupervised estimation procedure in order to find the unknown parameters of a NHCTHSMP associated with the multistate device is illustrated in Algorithm 1.

Algorithm 1 : Summary of the Estimation Procedure

Step 1: Set initial estimates for Γ and B and let $\theta_{old} = (\Gamma_0, B_0)$.

Step 2: Use re-estimation formula given in Eq. (4.12) and update B_{new}.

Step 3: Optimize all N-1 equations in the form of Eq. (4.10) to find Γ_{new} (parameters of transition rate distributions). Update θ as $\theta_{new} = (\Gamma_{new}, B_{new})$.

Step 4: Find the average log-likelihood function using Eq. (4.1) as $\frac{L'}{K}$. If $\frac{L'_{(new)} - L'_{(old)}}{K} \leq \epsilon$, terminate the algorithm and output $\theta^* = \theta_{new}$, otherwise set $\theta_{old} = \theta_{new}$, and go back to step 2. Here, ϵ is the stoping threshold for the iteration process.

Finding the expected value of the likelihood function is the expectation step (E-step) and finding the parameters that maximize Eq. (4.3) is the maximization step

(M-step). It is expected that after each iteration of the above algorithm, the value of log-likelihood function increases. It is important to note that the EM algorithm cannot guarantee the global maximum value. To perform the above steps, the expressions of all elements used in Eqs. (4.1), (4.10), and (4.12) should be defined in terms of θ. For notational convenience, we show how to calculate each of these elements for a single observation sequence O with l observation points and constant observation interval Δ. However the results can be simply extended to sequences of observations with variable observation intervals and lifetimes.

4.2.2 Steps to Calculate $\Pr(O|\theta)$

The term $\Pr(O|\theta)$, which is the probability of a single observation sequence ($O = O_1, O_2, ..., O_l$), can be efficiently calculated by modifying the forward-backward procedure given in [92]. Here, O_i is the output of the observation process at time t_i corresponding to the ith observation point. The first forward variable is defined as: $\alpha_t(i, O|\theta) = \Pr(O_1, O_2, ..., O_t, Q_t = i|\theta)$, which is the joint probability of being at state i at the tth observation point and observing the partial sequence of $O_1, O_2, ..., O_t$. The second forward variable is $u_t(i, O|\theta) = \Pr(O_1, O_2, ..., O_t, Q_{t-1} \neq i, Q_t = i|\theta)$, the joint probability of reaching state i for the first time at the tth observation point and observing the partial sequence of $O_1, O_2, ...O_t$. All $\alpha_t(i, O|\theta)s$ and $u_t(i, O|\theta)s$ can be approximated iteratively using the two steps, namely, Initialization and Induction steps. In the Initialization step, the values of specific forward variables are determined and in the Induction step, remaining forward variables are iteratively calculated based on the previously calculated forward variables. These steps are based on the assumptions that the device is in state 1 at time zero and the condition monitoring observation interval is small enough, so that at most one transition may occur in each interval.

In order to perform these two steps, the element $G_{i-j}(t, t+d, d_0|\theta)$, which is the conditional probability of a transition from state i (at the tth observation point) to state j (at time $(t + d)$th observation point), given that state i is observed at the $(t - d_0)$th observation point and stayed there for $d_0 \times \Delta$ units of time, should be defined. It should be pointed out here that when state j is not immediately reachable from state i, then $G_{i-j}(t, t + d, d_0|\theta) = 0$. The element G can be defined as follows:

$$
\begin{aligned}
&G_{i-j}(t,t+d,d_0|\theta) = \\
&\Pr(Q_{t+1},...,Q_{t+d-1}=i,Q_{t+d}=j|Q_1,...,Q_{t-d_0-1}\neq i,Q_{t-d_0},...,Q_t=i,\theta). \quad (4.13)
\end{aligned}
$$

The element G can also be expressed in terms of kernel function (Q). For $1 \leq i \neq j \leq N$; we have:

$$
\begin{aligned}
&G_{i-j}(t,t+d,d_0|\theta) = \\
&\Pr\left(X_{n+1}=j,(t+d-1)\Delta < T_{n+1} \leq (t+d)\Delta\right| \\
&X_n=i,(t-d_0-1)\Delta < T_n \leq (t-d_0)\Delta, t\Delta \leq T_{n+1},\theta) = \\
&\frac{\Pr\left(X_{n+1}=j,(t+d-1)\Delta < T_{n+1} \leq (t+d)\Delta | X_n=i,(t-d_0-1)\Delta < T_n \leq (t-d_0)\Delta,\theta\right)}{\Pr(t\Delta \leq T_{n+1}|X_n=i,(t-d_0-1)\Delta < T_n \leq (t-d_0)\Delta,\theta)} \\
&\approx \frac{Q_{i,j}((t-d_0)\Delta,(t+d)\Delta) - Q_{i,j}((t-d_0)\Delta,(t+d-1)\Delta)}{1-H_i((t-d_0)\Delta,d_0\Delta)}, \\
&0 \leq t \leq l, 0 \leq d \leq l-t, 0 \leq d_0 \leq t, \quad (4.14)
\end{aligned}
$$

and for $1 \leq i = j \leq N$, we have:

$$
\begin{aligned}
&G_{i-j}(t,t+d,d_0|\theta) = \\
&\Pr((t+d)\Delta \leq T_{n-1}|X_n=i,(t-d_0-1)\Delta < T_n \leq (t-d_0)\Delta, t\Delta \leq T_{n+1},\theta) = \\
&\frac{\Pr((t+d)\Delta \leq T_{n+1}|X_n=i,(t-d_0-1)\Delta < T_n \leq (t-d_0)\Delta,\theta)}{\Pr(t\Delta \leq T_{n+1}|X_n=i,(t-d_0-1)\Delta < T_n \leq (t-d_0)\Delta,\theta)} \approx \\
&\frac{1-H_i((t-d_0)\Delta,(d_0+d)\Delta)}{1-H_i((t-d_0)\Delta,d_0\Delta)}, \quad 0 \leq t \leq l, 0 \leq d \leq l-t, 0 \leq d_0 \leq t. \quad (4.15)
\end{aligned}
$$

The details of the two steps of Initialization and Induction are shown in Algorithm 2. Now, consider a backward variable $\beta_t^d(i,O|\theta)$, which is the conditional probability of future observations ($O_{t+1},...,O_l$), given that state i is observed, $(d-1)\Delta$ units earlier than the current time (t). This backward variable can be defined as:

$$
\beta_t^d(i,O|\theta) = \Pr(O_{t+1},...,O_l|Q_1,...,Q_{t-d}\neq i,Q_{t-d+1},...,Q_t=i,\theta). \quad (4.20)
$$

Similar to forward variables, we can start from an Initialization step and proceed with the Induction step to find all possible $\beta_t^d(i,O)s$. The detailed steps of the backward procedure are given in Algorithm 3. In both Algorithms 2-3, the termination step can be used to obtain $\Pr(O,\theta)$.

4.2.3 Steps to Calculate $\varepsilon_t(i,j,d|\theta)$ and $\kappa_a(r,j,d,O^{(k)}|\theta_{old})$

The term $\kappa_t(i,j,d,O^{(k)}|\theta_{old})$ is the joint probability of reaching state i at the tth observation interval, observing state j at the $(t+d)$th observation point and observing the full sequence of observations $O_1,...,O_l$ and $\varepsilon_t(i,j,d|\theta)$ is the probability

Algorithm 2 : Forward Procedure to Find $\Pr(O, \theta)$

I) Initialization step:

$$\begin{cases} u_0(1, O|\theta) = 1 & \\ u_t(1, O|\theta) = 0 & 1 \leq t \leq l \\ u_0(i, O|\theta) = 0 & 2 \leq i \leq N \\ u_1(i, O|\theta) = G_{1-i}(0, 1, 0|\theta) \times b_i(O_1) & 2 \leq i \leq N \\ \alpha_1(i, O|\theta) = G_{1-i}(0, 1, 0|\theta) \times b_i(O_1) & 1 \leq i \leq N \end{cases}. \tag{4.16}$$

II) Induction step:

$$u_t(i, O|\theta) = \sum_j \sum_{z=0}^{t-1} u_z(j, O|\theta) \times G_{j-i}(z, t, 0|\theta) \times \prod_{w=z+1}^{t-1} b_j(O_w) \ \times b_i(O_t),$$
$$2 \leq i \leq N, 2 \leq t \leq l, \tag{4.17}$$

$$\alpha_t(i, O|\theta) = \sum_{z=0}^{t} \left(u_z(i, O|\theta) \times G_{i-i}(z, t, 0|\theta) \times \prod_{w=z+1}^{t} b_i(O_w) \right),$$
$$1 \leq i \leq N, 2 \leq t \leq l. \tag{4.18}$$

After the last forward variables $(\alpha_1(1, O|\theta), ..., \alpha_l(N, O|\theta))$ are calculated from Eq. (4.18), the following termination step can be employed to compute the joint probability of observation sequence O, as:

III) Termination step:

$$\Pr(O, \theta) = \sum_{i=1}^{N} \alpha_l(i, O). \tag{4.19}$$

---

of transition from state i to state j at the $(t+d)$th observation interval, given that state i is observed at the tth observation point. After α, u, and β are calculated, these two measures can be calculated as follows:

$$\varepsilon_t(i, j, d|\theta) = G_{i-j}(t, t+d, 0|\theta), (i, j) \in E, 1 \leq t \leq l, l \leq d \leq (l-t), \tag{4.24}$$

$$\kappa_t(i, j, d, O|\theta_{old}) =$$
$$u_t(i, O|\theta_{old}) \times G_{i-j}(t, t+d, 0|\theta_{old}) \times \beta^1_{t+d}(j, O|\theta_{old}) \times \prod_{w=t+1}^{t+d-1} b_i(O_w) \times b_j(O_{t+d}),$$
$$1 \leq t \leq l, l \leq d \leq (l-t). \tag{4.25}$$

Algorithm 3 : Backward Procedure to Find $\Pr(O,\theta)$

I) Initialization step:

$$\begin{cases} \beta_l^{l+1}(1,O|\theta)=1 \\ \beta_l^d(i,O|\theta)=1, \quad 1\le i\le N, 1\le d\le l \end{cases}. \tag{4.21}$$

II) Induction step:

$$\beta_t^d(i,O|\theta)=G_{i-i}(t,t+1,d-1|\theta)\times\beta_{t+1}^{d+1}(i,O|\theta)\times b_i(O_{t+1})+$$
$$\sum_j G_{i-j}(t,t+1,d-1|\theta)\times\beta_{t+1}^1(j,O|\theta)\times b_j(O_{t+1})$$
$$1\le i\le N, 1\le t\le l-1, 1\le d\le t. \tag{4.22}$$

After the backward variables $\beta_1^1(1,O|\theta),...,\beta_1^1(N,O|\theta)$ are calculated from Eq. (4.22), we can use the following Termination step to estimate the probability of observation sequence O. It should be noted that both Eq. (4.19) and Eq. (4.23) can be used to estimate $\Pr(O|\theta)$.

III) Termination step:

$$\Pr(O|\theta)=G_{1-1}(0,1,0|\theta)\times\beta_1^2(1,O|\theta)\times b_1(O_1)$$
$$+\sum_{j\neq 1}^{N} G_{1-j}(0,1,0|\theta)\times\beta_1^1(j,O|\theta)\times b_j(O_1). \tag{4.23}$$

4.2.4 Steps to find $\gamma_t(i,O|\theta)$

The term $\gamma_t(i,O|\theta)$ is the joint probability of being at state i at time t and observing the full sequence of observations $O_1,O_2,...,O_l$ as:

$$\gamma_t(i,O|\theta)=\Pr(Q_t=i,O|\theta)=\sum_{v=0}^{t}u_v(i,O|\theta)\times G_{i-i}(v,t,0|\theta)\times\beta_t^{t-v+1}(i,O|\theta)\times$$
$$\prod_{w=v+1}^{t} b_i(O_w), i\in E, 1\le t\le l, \tag{4.26}$$

where $\sum_{i=1}^{N}\gamma_t(i,O|\theta)=\Pr(O|\theta)$. Now that all elements of the optimization problem are defined, the described iterative algorithm (see Section 4.2.1) can be used to estimate the unknown characteristic parameters of the model.

4.3 Model Selection

So far, the main assumption made in this thesis is that the structure (topology) of the multistate model is known. In this section, a simple hierarchical and enumerative approach is presented, which can be used to find a reasonable structure for the degradation and observation processes associated with the device under study. As discussed earlier, the model selection involves two steps: (1) configuration selection and (2) parameter estimation. The objective of the configuration selection process is to find a reasonable structure for the degradation and the observation processes associated with a device under study so that the actual degradation process and its associated condition-monitored observation process are reasonably represented by the selected multistate topology. As discussed earlier, most studies in the domain of multistate degradation modeling have assumed that the configuration of the multistate model is known and only unknown characteristic parameters of the selected model (Γ, B) need to be estimated.

To my knowledge, there are three reasons that explain why configuration selection is rarely considered in the literature: (1) the number of possible topology alternatives for the multistate structure can theoretically be very large (all possible combinations of $\zeta = \{N, \Omega, \xi, \lambda, I, V\}$), which makes the model selection process possibly very time consuming , (2) efficient training of such type of topology with 6 aforementioned elements requires large training, validation, and testing data, which are usually not available in real-world problems, and (3) available knowledge including historical information, engineering and field experiences, and available information from similar devices may facilitate determining some of these elements without performing model selection. In this section, a simple hierarchical framework is presented, which can be used to determine a structure for the multistate degradation process with unobservable states when multiple alternatives exist. It should be pointed out here that this framework is only useful when there are limited topology candidates to be compared and there is no guarantee that the result of this framework provides an optimal topology. Investigating such an optimal topology is out of the scope of this thesis.

The main purpose of model selection is to find a model among several alternatives that can better represent actual degradation and observation processes associated with a device under study. Different types of model selection criteria are used in the literature for model selection problems when more than one alternative exist [138]. Among those methods, the maximum likelihood criterion has been widely used due to its simplicity and good convergence properties [139]. Maximum likelihood model selection usually gives higher probability values as the number of parameters increases and may be subject to overfitting [140]. Overfitting means the model describes the training set better, but get worse on other instances of the same phenomenon. The main limitation of the maximum likelihood criterion for model selection is that it does not penalize the large number of parameters in the model. A very common approach to deal with this issue is to use Bayesian Information Criterion (BIC) for model selection. In BIC, there is a penalty term for penalizing a large number of parameters and complex models. This type of criteria is desirable as it aims to select the simplest model that best fits data [140]. In the next subsection, the enumerative model selection framework using BIC is presented.

4.3.1 An Enumerative Approach for Multistate Model Selection

In this section, a simple enumerative approach is presented, which can be used to compare several model alternatives. As described earlier, a multistate model (M) is fully defined if its topology (ζ) and the associated parameters of that topology (θ) are known. Let $M_1, M_2, ..., M_h$ be h known model candidates for a multistate structure. Also, let ζ_i, and θ_i be the topology and the set of parameters associated with the ith model candidate, respectively. Let O denote the set of condition monitoring data $(O^{(1)}, O^{(2)}, ..., O^{(k)})$ available for comparing these model alternatives and $q_{d_k}^{(i)}$ denote the degradation state observed at the last observation point of $\mathrm{O}^{(i)}$. The Bayesian model selection aims to select the model among possible alternatives $(M_1, M_2, ...)$ that has the highest value for the evidence probability $(\Pr(X|M))$, where X is the information available from K samples $(X = O^{(1)}, q_{d_1}^{(1)}, O^{(2)}, q_{d_2}^{(2)}, ..., O^{(K)}, q_{d_k}^{(K)})$. The Bayesian information criterion for the ith model can be defined as [139]:

$$BIC(M_i) = \log \Pr(X|\zeta_i, \hat{\theta}_i) - \alpha \frac{H_i}{2} \log(K), \tag{4.27}$$

where $\hat{\theta}_i$ is the maximum likelihood estimate of θ_i , H_i is the number of parameters in the ith model, K is the number of training samples, and α is the regularization parameter, which takes into account that estimates are not accurate (it is the weight of the penalty term). BIC has been widely and successfully applied for model selection. Now, among several model structures, the one with the highest BIC can be regarded as the best model. Here, the evidence likelihood of the ith model can be calculated as:

$$\Pr(X|\zeta_i,\hat{\theta}_i)=\prod_{k=1}^{K}\Pr(O^{(k)},Q^{(k)}|\zeta_i,\hat{\theta}_i)=\prod_{k=1}^{K}\Pr(O_1^{(k)},O_2^{(k)},...,O_{d_k}^{(k)},q_{d_k}^{(k)}|\zeta_i,\hat{\theta}_i)=$$
$$\prod_{k=1}^{K}\alpha_{d_k}(i,O|\zeta_i,\hat{\theta}_i). \quad (4.28)$$

As discussed in Section 3.3, the configuration of the ith model (ζ_i) involves 6 elements, which are the number of states, transition diagram, transition types, transition rate distributions, CM feature, and number of clusters used for feature representation. Considering limited number of possible options for each of the above elements, one can construct the likelihood function and estimate the unknown parameters based on all available combinations of the above elements. Then, Eqs. (4.27) and (4.28) can be used to find the best model among all possible alternatives.

As in practical cases, a large number of options are possible for each of the above elements, the total number of model alternatives may become extremely large, making the above model selection approach computationally expensive or sometimes infeasible. For example, having only 5 options for each element is equivalent to $15,625(5^6)$ different model alternatives, for which the parameter estimation procedure needs to be applied. Therefore, applying the above model selection criterion to compare model alternatives in real-world problems may be computationally very expensive.

An alternative approach can be used in the sense that the model selection or structural determination is divided into three phases. After data collection and condition monitoring feature extraction, using engineering judgments and other historical knowledge on the degradation and observation processes associated with the device under study, some elements of the configuration are pre-determined (phase I in Figure 4.1). For example, historical information or engineering judgement may verify that the overall degradation of the device can be represented by 4 states and the transition rate distribution follows the Weibull distribution. In such a circumstance, all model alternatives will have 4 levels of health states (N=4) with Weibull-based transition rates. This step can significantly lower the number of model alternatives. The remaining elements (those that are not pre-defined in phase I) now construct the set of possible topology alternatives, denoted by topology 1, ..., topology h. Each topology represents one unique multistate configuration. In phase II, the parameter estimation method can be applied for each possible model alternative to estimate the characteristic parameters (the parameter estimation procedure is repeated h times). The output of Phase II is h trained models with known configuration elements and estimated characteristic parameters (trained model 1, ..., trained model h). Then, in phase III, using a model selection criterion, the trained models are compared and the one, which best satisfies the selection criterion is selected as the best possible model alternative. Now, the multistate model for reliability analysis of the device is fully defined. With such a strategy, the number of model alternatives decreases, making the configuration selection process less time-consuming and more feasible to implement. The summary of the described approach for model selection is shown in Figure 4.1.

As shown in Figure 4.1, after data collection and feature calculation, depending on the information available on the elements of the model structure, some elements of the model structure are determined in phase I. Depending on the information available for the elements of the multistate structure, the number of alternatives varies. The outputs of Phase I are h configuration (topology) alternatives denoted by $\zeta_1, \zeta_2, ..., \zeta_h$. A common example of a topology alternative is a multistate structure with unknown number of states. For example, topology alternatives can be 2-states, 3-states,..., and $h+1$-states topologies. After the topology candidates are determined, the characteristic parameters of the alternatives are estimated in Phase II (model training) by employing the approach given in Section 4.2. The output of phase II is h trained models $(M_1, M_2, ..., M_h)$ with known structure $(\zeta_1, \zeta_2, ..., \zeta_h)$ and estimated characteristic parameters $(\hat{\theta_1}, \hat{\theta_2}, ..., \hat{\theta_h})$. In phase III, using a model selection criterion, all model alternatives are compared and the best model structure is selected. It should be pointed out that for the final model comparison (phase III), an independent validation data set may be used to minimize overfitting. In Chapter 7, the above-discussed framework is employed on a case study for model selection when multiple alternatives exist.

4.4 Numerical Example

In this section, simulation-based numerical experiments are used to illustrate how the proposed parameter estimation procedure presented in Section 4.2 can be employed to estimate the characteristic parameters of a degradation process from historical data. Also, discussions on its correctness and accuracy are provided.

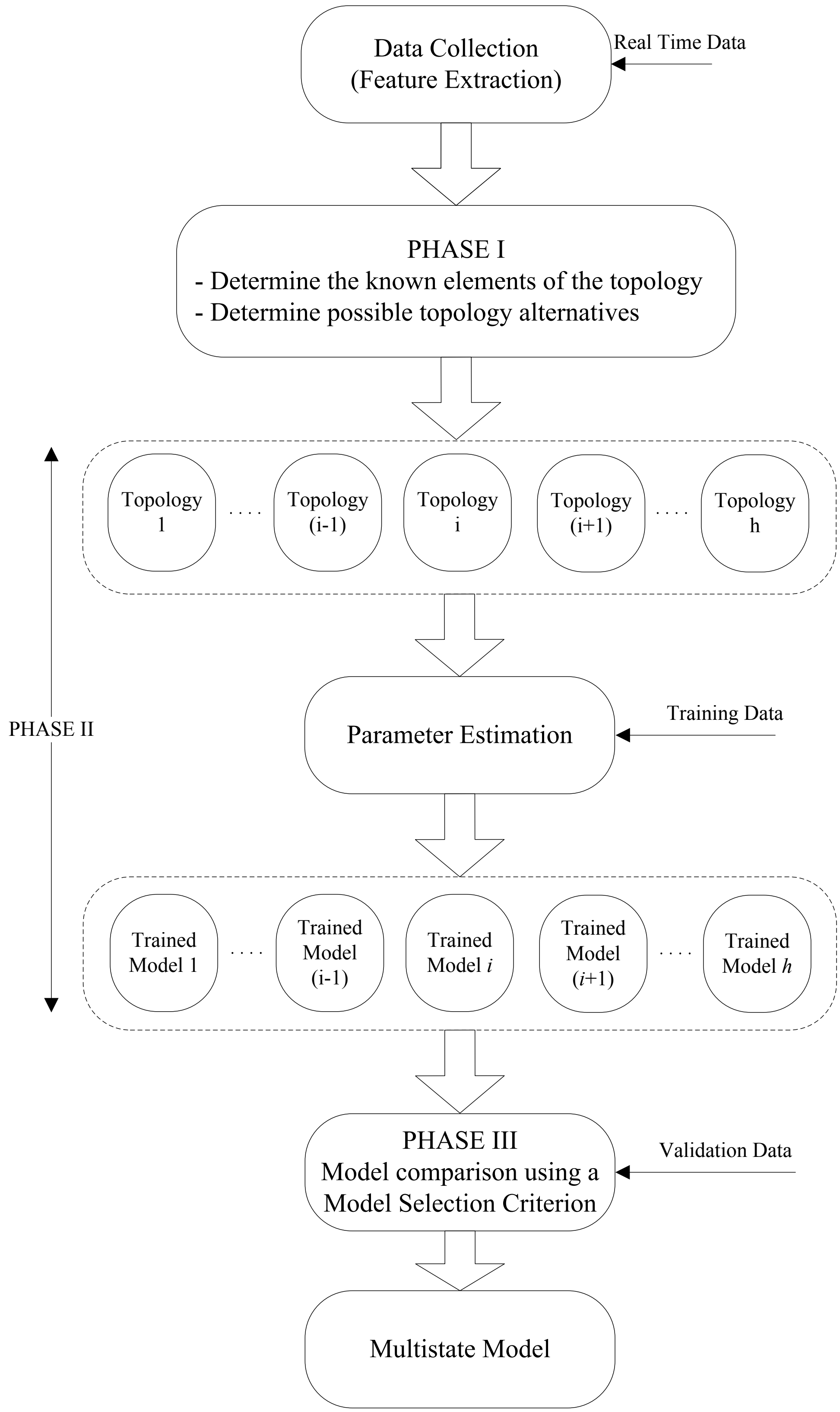

Figure 4.1: Model selection framework for multistate degradation modeling

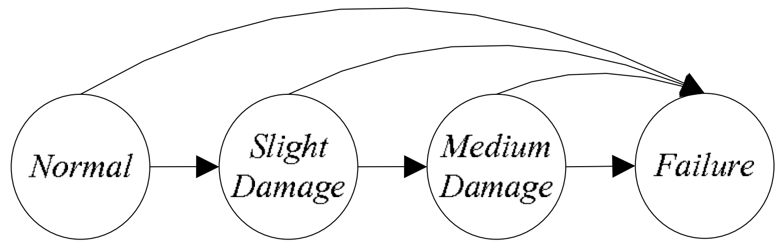

Figure 4.2: Transition diagram of a device with 4 levels of health states

4.4.1 Example Description

A piece of mechanical equipment with a known multistate topology is considered. The elements of the associated multistate topology are described below. It is assumed that the equipment is operating under four levels of health conditions ($N = 4$), namely, normal condition, slight damage, medium damage, and failure. States are ordered from 1 to 4, where state 1 refers to the normal condition and state 4 refers to the failure state. The equipment may pass through all degradation states before failure. However, due to the existence of random shocks, it can also fail directly from any state. Therefore, at each state, 2 types of transitions are possible: (I) transition to the one-step degraded state (immediate neighbor state), and (II) transition to the failure state. The transition diagram (Ω) of this equipment, which presents the connectivity between states, is shown in Figure 4.2.

It is assumed that transitions from state i ($1 \leq i \leq 3$) to the immediate neighbor state depend on the two states involved in the transition and the sojourn time spent at state i. It is also assumed that the transition from state i ($1 \leq i \leq 2$) to state 4 depends on on the two states involved in the transition and the total age of the device. Therefore, the multistate structure (ξ) follows a nonhomogeneous continuous-time semi-Markov process (see Section 3.4). For transition distributions (λ), the Weibull distribution is used, which is the most commonly used distribution to represent degradation [141]. The final distributions of transition rate functions are as follows:

$$\lambda_{i,j}(s,t) = \begin{cases} (\beta_{i,j}/\alpha_{i,j}) \times (t/\alpha_{i,j})^{(\beta_{i,j}-1)}, & (i,j) \in [(1,2),(2,3),(3,4)], \\ (\beta_{i,j}/\alpha_{i,j}) \times ((s+t)/\alpha_{i,j})^{(\beta_{i,j}-1)}, & (i,j) \in [(1,4),(2,4)] \end{cases} . \tag{4.29}$$

The parameters of the transition rate distribution are as follows:
$\alpha = \begin{bmatrix} 0 & 15 & 0 & 21 \\ 0 & 0 & 12 & 28 \\ 0 & 0 & 0 & 7 \end{bmatrix}$, and $\beta = \begin{bmatrix} 0 & 8 & 0 & 3 \\ 0 & 0 & 6 & 12 \\ 0 & 0 & 0 & 4 \end{bmatrix}$, where the elements in the ith row

and jth column of the scale (α) and shape (β) matrices are the characteristic parameters of the Weibull distribution associated with the degradation transition from state i to state j. Now that the structure of the degradation process is fully defined, the structure of the observation process should be determined. All states are only indirectly observable through condition monitoring except the failure state (which is self-announcing). A single condition monitoring indicator referred to as I is monitored periodically (Δ=1), while the equipment is operating. In other words, every one unit of time, the value of the observation process is recorded for further analysis. This temporal sequence of indicator values is referred to as the observation process. It is assumed that the CM indicator has 8 possible discrete outcomes ($M = 8$), which are denoted by $V = \{1, 2, ..., 8\}$. The stochastic relationship between the degradation process and the observation process is represented by a nonparametric discrete distribution referred to as the observation probability matrix (B). Therefore, the observation probability matrix B has 4 rows ($N = 4$) and 8 columns ($M = 8$), where the element in the ith row and jth column of this matrix represents the probability that the output of the observation process is j when the device is in state i. The elements of this matrix are shown below:

$$\text{B}= \begin{bmatrix} 0.45 & 0.35 & 0.15 & 0.05 & 0 & 0 & 0 & 0 \\ 0.05 & 0.10 & 0.30 & 0.35 & 0.10 & 0.05 & 0.05 & 0 \\ 0 & 0 & 0.05 & 0.05 & 0.20 & 0.25 & 0.45 & 0 \end{bmatrix}.$$

Considering the above information on the elements of the multistate topology ($\zeta = \{N, \Omega, \xi, \lambda, V, I\}$) and its associated characteristic parameters ($\theta = \{\Gamma, B\}$), the structure of the multistate model ($M = \{\zeta, \theta\}$) is now fully defined.

4.4.2 Random Sequence Generation

The stochastic behavior of the multistate equipment with degradation states under the non-homogeneous continuous-time hidden semi-Markov structure is simulated by the Monte Carlo simulation method and corresponding observation process and degradation process are generated. Then, simulated observation sequences are used as the input for the estimation procedure. The results of the estimation procedure will then be compared with the actual parameters, which were originally used to generate the observation sequences. In this subsection, the procedure under which degradation and observation processes are simulated, is described.

The simulation process explained here generates multiple sequences of degradation states and their corresponding observation values based on the multistate structure defined in Section 4.4.1. For each simulated sequence of data, the state of the device and the condition monitoring indicator values are extracted over time. The output of the simulation consists of multiple temporal sequences of observations with known states, which are generated according to the successive visited states. It should be noted here that only observation process is used as the input for parameter estimation. Since throughout this thesis, transition rate function is considered as the fundamental describer of the degradation process, a simple technique to generate random sequences based on a NHCTHSMP in terms of transition rate functions is introduced.

A random temporal sequence involves both the degradation process (successive sequence of visited states and the time of each transition) and the observation process (successive observed condition monitoring indicators) from time zero up to the failure point. In order to do this, the couple (X_n, T_n) is first generated for $1 \leq n \leq n_f$, where n_f is the transition number at which the failure occurs ($n_f = \{n | X_n = N\}$). Then, based on the generated sequence of states (X_n), the corresponding observation values $U_m (1 \leq m \leq d_k)$, where d_k is the number of observation points, are generated. The simulation process, which is based on the inverse transform technique, is described below. Let a be a uniform random variable in the range $[0, 1]$. If $b = F^{-1}(a)$, then b is a random variable with CDF, F. It is possible to define the cumulative hazard function between state i and j based on the corresponding transition rate function as:

$$\Lambda_{i,j}(s,t) = \int_0^t \lambda_{i,j}(s,u)du. \tag{4.30}$$

Now, based on the inverse transform technique, we can generate random number $T^s_{i,j}$ for the time to transition from state i to state j, given that state i is reached at time s. Let a be a uniform random number from $U[0, 1]$, then a random number from F can be generated as:

$$\begin{aligned} &1 - F_{i,j}(s,t) = \\ &exp(-\Lambda_{i,j}(s,t)) \rightarrow \Lambda_{i,j}(s,t) = -\log(1-a) \rightarrow T^s_{i,j} = \Lambda^{-1}_{i,j}(s, -\log(1-a)). \end{aligned} \tag{4.31}$$

The summary of all steps used to generate a random sequence (X_n, T_n) is shown in Algorithm 4. A sample sequence of the degradation process is shown in Figure 4.3. In this sample, which corresponds to the life of a single piece of equipment, there are three transitions as $X_1 = 2$, $X_2 = 3$, and $X_3 = 4$ with corresponding

Algorithm 4 : Random Generation of Degradation Process

Step 1. Let $X_0 = 1, T_0 = 0$, and $c = 0$ and move on to step 2.

Step 2: Let $i = X_c$. For each $j \in FS_i$, generate a separate random number $T^s_{i,j}$ denoted by from Eq. (4.31), given that $s = T_c$.

Step 3: Let $c = c + 1$ and find the state and the time of the next transition as $X_c = \arg \min_{j \in FS_i} \{T^s_{i,j}\}$, where X_c takes the j value that has the smallest $T^s_{i,j}$. This step is based on the fact that the next transition is realized according to the event that occurs first in a competition among all possible transitions out of state i.

Step 4: If $X_c = N$, terminate the algorithm and output (X, T), otherwise move back to Step 2.

---

transition times $T_1 = 15$, $T_2 = 28$, and $T_3 = 38$. Algorithm 5 shows how to generate $U_j \quad (1 \leq j \leq d)$, which is the corresponding values of the observation process with length d associated with (X_n, T_n). Now, based on the predetermined observation

Algorithm 5 : Random Generation of Observation Process

Step 1: Let j=0 and go to step 2.

Step 2: Set $j = j + 1$. Sample a random number a from $U[0, 1]$. The process U_j equals v_f, if $\sum_{z=1}^{f-1} b_{X^0_j}(z) < a \leq \sum_{z=1}^{f} b_{X^0_j}(z), 1 \leq f \leq m$, where X^0_j is the state of the device at the jth observation point.

Step 3: If $X^0_j = N$, terminate the algorithm and output (U), otherwise move back to Step 2

---

interval Δ, the value of condition monitoring indicator at each observation point can be simulated. A sample sequence of the observation process is shown in Figure 4.3. In this sample, which corresponds to the life of a single piece of equipment, the outputs of the observation process for a single device, which takes values from 1-8, are shown. It can be seen from this figure that as the device ages over time, its level of degradation increases. Also, the CM output varies with time and indirectly reflects the deterioration process. The output of the above-described simulation process is K independent sequences of observation process denoted by $O = \{O^{(1)}, O^{(2)}, ..., O^{(K)}\}$, which can be used as the input for parameter estimation.

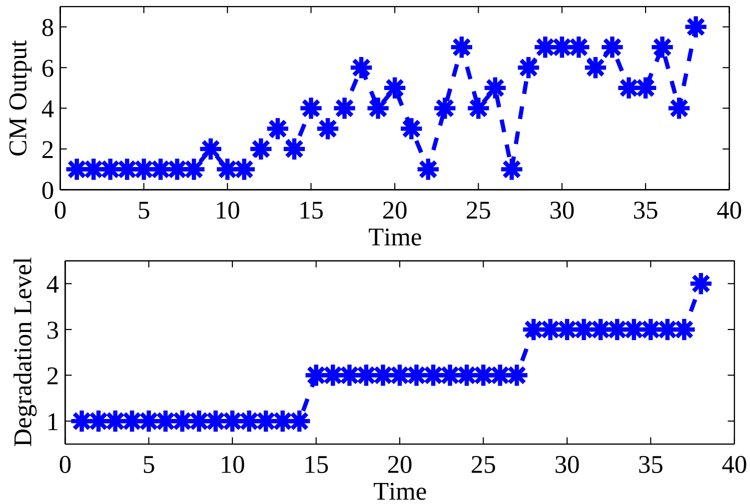

Figure 4.3: Sample realization of the degradation and observation processes

4.4.3 Parameter Estimation

After the procedure explained in Section 4.4.2 is employed to generate several independent sequences of degradation and observation processes, the simulated sequences of the observation process can be employed as the input for parameter estimation. In other words, it can assumed that the characteristic parameters of the multistate structure (θ) are not known and only several sequences of observation process are available to be used for parameter estimation. The results of the estimation procedure will then be compared with the true parameter values, which were originally used to simulate the observation sequences. To evaluate the effect of available training data on the results, five different values for the number of historical observations are considered ($K = 20, K = 40, K = 80, K = 160$, and $K = 320$). In addition, to reach more consistent results, for each case of K, the parameter estimation method is applied 50 times, that is 50 different sets of data are simulated for each case of K. We used Matlab 2011 on a stand-alone PC with CPU 2.3 GHz and 16 GB of RAM for all numerical experiments. The *fminsearch* function in Matlab is used for optimization purposes in the estimation procedure.

As discussed in Section 4.2, the proposed parameter estimation process required initial estimates for all unknown parameters. For the transition distributions, the initial value of 40 is assumed for all scale parameters and the initial value of 1 is assumed for all shape parameters. Also for the observation transition matrix, the initial value of $(1/(M-1))$ is assumed for all entries except those in the last row and the last column, which are related to the observable failure state. Therefore, entry $(4, 8)$ of matrix B is 1. The threshold of 0.005 is considered as the stopping criteria for the average log-likelihood function improvement. To be able to evaluate the correctness of the estimates for each case of K and also to confirm the improvement in estimation error after each step of the proposed estimation procedure, a well-known measure called mean squared error (MSE) is used, which measures the difference between the estimated values and the actual values as:

$$MSE(\theta, \hat{\theta}) = \sum_{i=1}^{n}((\theta_i - \hat{\theta}_i)^2/n), \tag{4.32}$$

where θ_i and $\hat{\theta}_i$ are, respectively the actual and the estimated values of the ith parameters and n is the number of unknown parameters. As there are 5 possible transitions for the device, there are 5 unknown shape parameters and 5 unknown scale parameters associated with the degradation process. Also, as there are 4 states and 8 possible outcomes for the condition monitoring indicator, there are 21 $((4 - 1) \times (8 - 1))$ unknown parameters associated with the observation process. Because the ranges of the scale parameters, shape parameters, and entries of the observation probability matrix are different with each other, we separately calculate MSE for each of these parameters set. The results of the numerical experiments are shown in Section 4.4.4.

4.4.4 Results

In this section, the results of the numerical experiments are described from three aspects: (I) Log-likelihood improvement after each iteration or the ability of the estimation procedure to improve likelihood value over iterations, (II) Estimation results, which show the accuracy of the estimation results (mean, standard deviation, and mean squared error (MSE)), and (III) CPU time results, which show how long the estimation procedure may take to provide estimation results.

I) Log-likelihood Improvement

As discussed earlier, it is expected that the proposed iterative estimation method improves log-likelihood function after each iteration of the algorithm. The result of implementing the parameter estimation method using simulated data verifies the ability of the proposed estimation method to increase the log-likelihood function after each iteration. In Figure 4.4, the average log-likelihood value from iteration to iteration of the estimation procedure calculated using 80 simulated sequences of the observation process is presented. It can be seen in this figure that, the likelihood function improves at a much faster rate in the first few iterations, and then its improvement gets slower. Similar observations have been seen for $K = 20, 40, 80, 160,$ and 320.

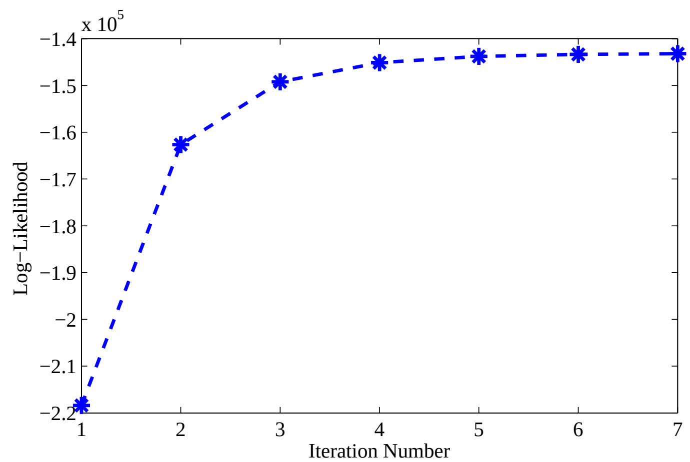

Figure 4.4: Log-likelihood improvement versus iteration number for $K = 80$

II) Estimation Results

Table 4.1 presents the true values of each characteristic parameter used for simulation, the average, the standard deviation (STD), and the mean squared error (MSE) for the estimated parameters for all cases of K based on 50 independent sets of simulated data. It can be verified from Table 4.1 that the averages of the estimated values are reasonably close to the true values, particularly when a larger number of life histories (K) are employed for estimation. In other words, as expected, the mean estimated values approach the true values as K increases. Moreover, that the MSE values for the estimated parameters converge to zero as K increases is also verified. As the estimation results for the elements of the observation probability matrix are very close to one another, the detailed results are not shown here. The average MSE values of all three sets of parameters (α, β, and B) are shown in Figures 4.5 and 4.6 for each case of K. The consistent trends in Figures 4.5 and 4.6 verify that MSE converges to zero as the number of available life histories for estimation (K) increases. In summary, it can be concluded that the estimated values with an acceptable level of accuracy are found, in particular for a sufficient training sample. It is worth mentioning that by tightening the stopping criteria, we may get better estimates in a longer CPU time.

Table 4.1: Estimation results for 50 estimation runs

Parameters		$\alpha_{1,2}$	$\alpha_{1,4}$	$\alpha_{2,3}$	$\alpha_{2,4}$	$\alpha_{3,4}$	$\beta_{1,2}$	$\beta_{1,4}$	$\beta_{2,3}$	$\beta_{2,4}$	$\beta_{3,4}$
True Values		15	21	12	28	7	8	3	6	12	4
Mean	K=20	15.16	22.72	11.69	28.24	7.26	8.72	3.35	6.68	16.08	4.35
	K=40	15.05	21.60	11.92	28.09	7.05	8.02	3.28	6.20	14.44	4.20
	K=80	14.98	21.08	11.94	28.01	7.08	8.15	3.16	5.72	12.88	4.05
	K=160	15.04	21.09	11.88	27.96	7.14	8.00	3.03	5.81	12.30	4.00
	K=320	15.08	21.19	11.90	27.91	7.15	7.99	2.99	5.83	12.23	4.00
STD	K=20	0.70	5.51	0.98	4.44	1.15	2.71	1.17	2.60	7.32	1.30
	K=40	0.42	3.95	0.52	1.34	0.52	1.26	0.94	1.64	4.73	0.95
	K=80	0.24	1.99	0.45	0.73	0.38	0.95	0.61	1.07	2.13	0.62
	K=160	0.23	1.46	0.27	0.52	0.28	0.75	0.43	0.86	1.73	0.47
	K=320	0.15	0.93	0.22	0.31	0.21	0.44	0.33	0.60	0.93	0.38
MSE	K=20	0.50	32.67	1.04	19.41	1.36	7.70	1.47	7.08	69.20	1.77
	K=40	0.17	15.65	0.27	1.76	0.26	1.56	0.93	2.68	27.84	0.92
	K=80	0.06	3.89	0.20	0.52	0.14	0.90	0.39	1.19	5.20	0.38
	K=160	0.05	2.09	0.09	0.26	0.10	0.55	0.19	0.76	3.03	0.22
	K=320	0.03	0.88	0.06	0.10	0.06	0.19	0.11	0.39	0.90	0.14

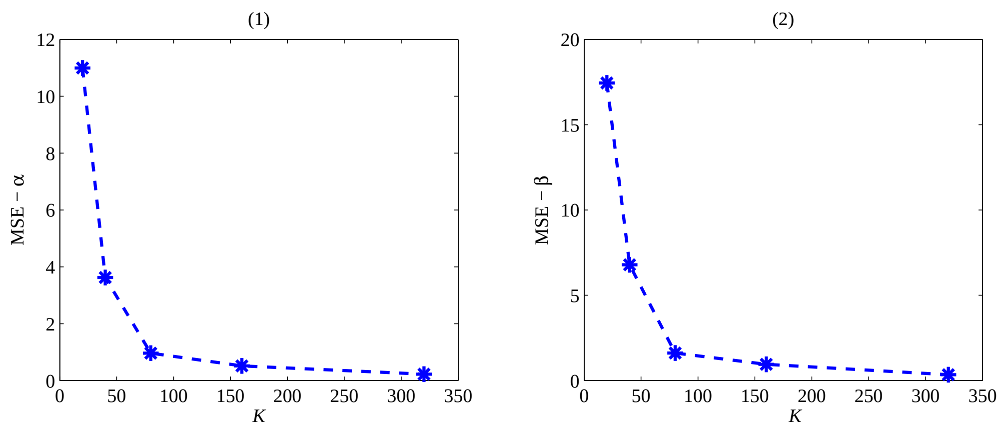

Figure 4.5: Average MSE (based on 50 runs) for α and β

III) CPU time

As expected, the estimation procedure took relatively longer when a larger number of observations (K) were employed. Figure 4.7 shows the average CPU time for different levels of K considering 50 independent estimation runs. Each point in this figure shows how long it took on average to use K independent sequences of observation process for parameter estimation. In all cases, the final estimates are achieved in less than 5 hours of CPU time. As in practical application of this unsupervised estimation method, the estimation procedure is an offline course of action, the proposed estimation model is reasonable in terms of CPU time. It should also note that CPU time is very sensitive to the structure of the multistate model and maximum life of the device, as both can highly affect the size of the likelihood function and therefore increase the CPU time needed for parameter estimation.

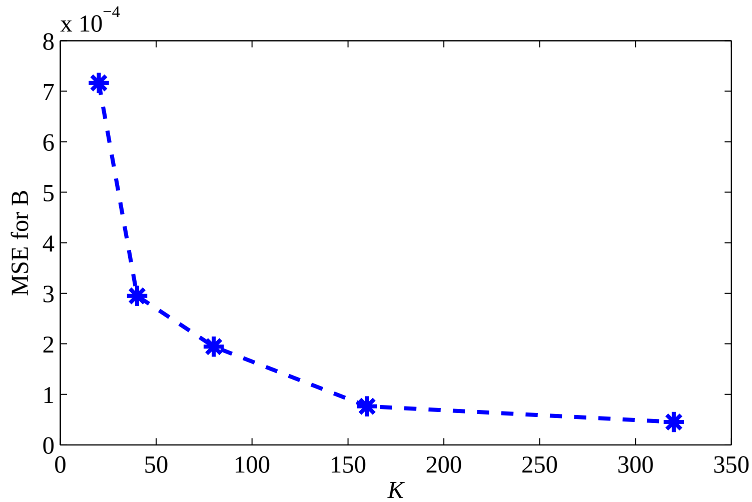

Figure 4.6: Average MSE (based on 50 runs) for B

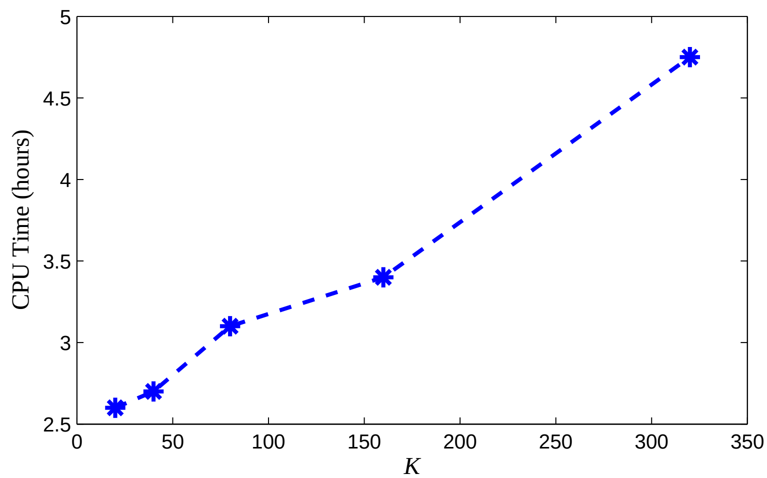

Figure 4.7: CPU time evaluation

4.5 Summary

The primary step to use a stochastic process for modeling the degradation and observation processes associated with a device under study is to find a reasonable structure that can represent the stochastic behaviour of these processes. The result of this chapter can be used to estimate the characteristic parameters of a known multistate model by employing historical condition monitoring data. After the multistate model is trained, it can be used for diagnostics and prognostics (see Chapters 6 and 7). The method proposed in this chapter for parameter estimation can utilize available condition monitoring data to train a nonhomogeneous continuous-time hidden semi-Markov process that can be used to represent the degradation and observation processes associated with a device under condition monitoring. In Chapter

7, this parameter estimation method is used on turbofan engine data sets to estimate the characteristic parameters of associated multistate structures. Although a simple enumerative framework is presented to find a reasonable multistate structure among available model structures, developing a more efficient approach for model selection is an important direction for future research. In the remainder of this thesis, how to use a trained multistate process for diagnostics and prognostics and maintenance decision making is discussed.

Chapter 5

Key Diagnostics and Prognostics Measures

5.1 Introduction

One of the main objectives of condition monitoring is to provide useful information on the current and future health status of a device under operation. Generally, this type of information is reported by some important characteristic and performance measures, which are easy to understand and directly or indirectly represent some important aspects of the degradation and observation processes associated with the device under study. These important measures should be defined in a way that they can finally be used for maintenance decision making.

Performance measures can be categorized into different classes from several viewpoints. From the calculation point of view and the data required for calculating a measure, these measures can be static or dynamic. Static means that the measures are computed without using any information on the actual operation of the device. Static measures are usually calculated based on some historical information or prior knowledge on the degradation process. Therefore, static measures are constant for different devices of the same type, that is, measures are independent of the actual operating condition, which may result in several degradation and observation patterns. A well-known example of static measure is the mean time to failure (MTTF) that is calculated independent of the actual operation of the device.

As in reality, stochastic properties of the degradation and observation processes result in different degradation and observation patterns, it is more reasonable to employ information available on the actual operation of the device for performance measure calculation. These types of measures are referred to as dynamic measures, that is, they change over the age of the device and also reflect the dynamic behavior of the degradation and observation processes. It should be pointed out that under condition monitoring, these measures can be conditional on the condition monitoring data, that is, all historical condition monitoring data (up to the current time point) extracted at previous observation points are used for calculation. Examples are conditional mean remaining useful life, which is calculated from the profile of the observation process extracted at previous condition monitoring points (condition monitoring data).

From the application point of view, performance measures can be divided into diagnostic measures, which deal with the current health status of the device, and prognostic measures, which deal with the future health status of the device. In this chapter, dynamic diagnostic and prognostic measures are introduced for a multistate structure evolving according to a NHCTHSMP. This chapter is organized as follows: Sections 5.2 and 5.3 respectively present the definition and mathematical expressions of diagnostic and prognostic measures. In Section 5.4, numerical examples are used to demonstrate how these measures can be used for online diagnosis and prognosis. Finally in Section 5.5, the conclusions obtained from the results of this chapter are discussed. The results of this chapter are reported in [27].

5.2 Diagnostic Measures

As discussed earlier, the objective of developing diagnostic measures is to provide useful information on the current health status of the device. In this section, important dynamic diagnostic measures are presented, which employ condition monitoring data to reflect certain aspects of the current degradation level of a device under study. For each measure, the definition, application, and its mathematical expression are illustrated in details. While the definitions of some of these measures are derived from the literature, this is the first time that such measures are discussed for a multistate degradation process with hidden states, which evolves according to a NHCTHSMP. It should be pointed out here that the common assumption throughout this chapter is that the structure of the multistate model (M) is already known.

5.2.1 Conditional State Probability

This measure provides the probability of being in a certain state at the current point of time, given the profile of condition monitoring data. We use $\omega_t(i|O, M)$ to denote the probability that the device is at the degradation state i at time t, given that the multistate model (M) is known and the sequence of $O_1, O_2, ..., O_t$ is observed. This measure can be calculated as follows:

$$\omega_t(i|O, M) = \Pr(Q_t = i|O_1, O_2, ..., O_t, M) = \frac{\Pr(Q_t = i, O_1, O_2, ..., O_t|M)}{\Pr(O_1, O_2, ..., O_t|M)} = \frac{\alpha_t(i, O|M)}{\sum_{i=1}^{N} \alpha_t(i, O|M)}, \quad (5.1)$$

where Q_t is the state of the device at the tth observation point, $\alpha_t(i, O|M)$ is the forward variable, denoting the joint probability of being in state i at the tth observation point and observing sequence $O_1, O_2, ..., O_t$. Details on how to calculate $\alpha_t(i, O|M)$ are given in Section 4.2.2. The conditional state probability presents the probability of being at each degradation level, given the condition monitoring data. This measure can be further employed to find the average degradation level and its confidence interval. It is clear that $\sum_{i=1}^{N} \omega_t(i|O, M) = 1, \forall t > 0$.

5.2.2 Average Degradation Level (ADL)

Finding a monotonic degradation index, which reflects the damage development, has always been a challenging topic in real-world condition monitoring frameworks [142]. There are two main reasons for the importance of such measures [143]. First, an oscillating measure cannot reflect the damage development trends perfectly. Second, the damage usually shows accelerated growth with the running time, that is, the degradation speed increases with time. Therefore, a feature with monotonic trend can better represent the degradation process. There are reported studies on finding a feature having a monotonic relationship with the actual health status of the device, such as [113, 144]. However, feature selection and feature fusion techniques are not always successful at finding such features that could also have a reasonable physical meaning.

As in this thesis, the device is assumed to be nonrepairable, its health status never improves over time, that is, the degradation level either stays at the same level or degrades with time. Therefore, a reasonable measure to monitor the overall health status of the device is the estimated average of the degradation process, which is equivalent to the average degradation states, given the condition monitoring data. At any time point t, given the condition monitoring data, the average degradation level (ADL) can be calculated as follows:

$$ADL(t|O,M) = \sum_{i=1}^{N} i \times \omega_t(i|O,t). \tag{5.2}$$

Similar to the average degradation level, the most likely state (MLS) at time t can also be estimated as $MLS(t|O,M) = \text{arg}\max_i \omega_t(i|O,M)$.

5.2.3 Conditional Hazard Rate

Hazard rate is one of the most important and commonly used performance measures used in reliability analysis of mechanical systems. Generally, hazard rate at time t is referred to as the rate of failure during the next instant of time, given no failure before time t. Hazard rate also equals to the negative of the derivative of $ln(R(t))$. It should be pointed out here that for the multistate degradation structure under study in this thesis, the hazard rate function is not explicitly defined and therefore it needs to be calculated from the known transition rates between states. When condition monitoring data are available, the instantaneous conditional hazard rate at time t calculated at time t_p (time of the pth condition monitoring point) can be defined as follows:

$$h(t|O_1, O_2, ..., O_p, L > t_p, M) = \lim_{\Delta t \to 0} \frac{R(t|O_1, O_2, ..., O_p, L > t_p, M) - R(t + \Delta t|O_1, O_2, ..., O_p, L > t_p, M)}{\Delta t \times R(t|O_1, O_2, ..., O_p, L > t_p, M)}, \tag{5.3}$$

where $R(t|O_1, O_2, ..., O_p, L > t_p, M)$ is the conditional reliability of the device at time t, given that the device has not failed up to time t_p and the sequence $O_1, O_2, ..., O_p$ is observed. Details on how to calculate this measure can be found in Section 5.3.2. It can be seen from the formula given in Eq. (5.3) that the hazard rate is the instantaneous rate of failure for the survivor at time t during the next instant of time.

A discrete form of hazard rate can be calculated from the probability of failure within the next observation interval. Given that the device has not failed at time t_p and the sequence $O_1, O_2, ..., O_p$ is observed, the conditional hazard rate for the next observation cycle can be calculated as:

$$h'(t_p|O_1, O_2, ..., O_p, L > t_p, M) = \frac{1 - R(t_{p+1}|O_1, O_2, ..., O_p, L > t_p, M)}{(t_{p+1} - t_p)}, \quad (5.4)$$

where $1 - R(t_{p+1}|O_1, O_2, ..., O_p, L > t_p, M)$ is the probability that the device fails before the next condition monitoring point. It should be noted that Eq. (5.4) becomes equivalent to the instantaneous conditional hazard rate at time t_p (Eq. (5.3)) if the observation interval is small enough. In this thesis, Eq. (5.4) is referred to as the conditional hazard rare, but readers should note that it is not the instantaneous hazard rate. The detailed definition and mathematical expression for reliability function are given in Section 5.3.2. The next possible derivation of hazard rate is the average conditional hazard rate over the next observation interval, which is defined as:

$$AFR(t_p|O_1, O_2, ..., O_p, L > t_p, M) = \frac{\int_{t_p}^{t_{p+1}} h(t|O_1, O_2, ..., O_p, L > t_p, M)}{(t_{p+1} - t_p)} = \frac{\ln R(t_p|O_1, O_2, ..., O_p, L > t_p, M) - \ln R(t_{p+1}|O_1, O_2, ..., O_p, L > t_p, M)}{(t_{p+1} - t_p)} = \frac{-\ln R(t_{p+1}|O_1, O_2, ..., O_p, L > t_p, M)}{(t_{p+1} - t_p)}. \quad (5.5)$$

As the overall health status of the device under study cannot improve over time, it is expected that the above measures, which reflect the dynamic characteristic of the degradation process, has a non-decreasing trend with the age of the device. However, as it is conditional on the history of condition monitoring data, noisy measurements can result in a non-monotonic trend. It should be pointed out here that hazard rate can also be considered as a prognostic measure, as it provides information on the future instances of time. In this thesis, hazard rate is treated as a dynamic diagnostic measure.

5.3 Prognostic Measures

Although important diagnostic measures can provide useful information on the current health status of the device, prognostic measures are more attractive to maintenance decision makers as they provide information on the future health status of a device, which can be used for maintenance decision making. Recent research trend also verifies the relative importance of prognostic measures compared to diagnostic measures. This subsection is devoted to important prognostic measures, which (1) are calculated from available condition monitoring data and (2) provide information on the future health status (degradation level) of the device.

5.3.1 Conditional Future State Probability

This measure is very similar to the one given in Section 5.2.1 for the conditional state probability except that the probability of the state of interest at a future time point is investigated. In other words, conditional on the profile of condition monitoring data, it tries to predict the probability of being in states 1,...,N at a time point in future. This measure can be defined as the probability that the device is at state i at a future time point t, given that it has not failed up to time t_p ($t > t_p$) and the condition monitoring sequence $O_1, O_2, ..., O_p$ is observed. With this predictive measure, we can also estimate the most likely degradation pattern from the current time point until a time point in future. This measure can mathematically be defined as:

$$\upsilon_t(i|O_1, O_2, ..., O_p, L > t_p, M) = \Pr(Z_t = i|O_1, O_2, ..., O_p, L > t_p, M) = \frac{\Pr(Z_t = i, O_1, O_2, ..., O_p, L > t_p|M)}{\Pr(O_1, O_2, ..., O_p, L > t_p|M)} = \frac{\Pr(Z_t = i, O_1, O_2, ..., O_p, L > t_p|M)}{\sum_{j=1, j\neq N}^{N} \alpha_{t_p}(j, O|M)}. \quad (5.6)$$

The term $\Pr(Z_t = i, O_1, O_2, ..., O_p, L > t_p|M)$ is equivalent to $\sum_{j=1}^{N-1} \Pr(Z_t = i, O_1, O_2, ..., O_p, Q_p = j, L > t_p|M)$, where:

$$\Pr(Z_t = i, O_1, O_2, ..., O_p, Q_p = j, L > t_p|M) = \sum_{z=0}^{p} u_z(j, O|M) \times \prod_{w=z+1}^{p} b_j(O_w) \times \Pr(Z_t = i, Q_p = j|Q_{z-1} \neq j, Q_z = j, M), \quad (5.7)$$

where $u_z(j, O|M)$ is the joint probability of observing the sequence of $O_1, O_2, .., O_z$ and being at state j at the zth observation point for the first time (see Section 4.4.3) and $b_i(j)$ is the probability that the jth observation value is observed while the state of the device is i (an element of the observation probability matrix B). The final and the most difficult step is to find $\Pr(Z_t = i, Q_p = j|Q_{z-1} \neq j, Q_z = j, M)$, which is the probability of being in states i and j at times t and t_p, respectively, given that state j was observed for the first time at the zth observation point ($Q_{z-1} \neq j, Q_z = j$).

This measure can be computed based on the following steps:

$$\Pr(Z_t = i, Q_p = j | Q_{z-1} \neq j, Q_z = j, M) \simeq \sum_k \int_{t_p}^{t} \dot{Q}_{j,k}(t_z, \tau | M) \times \Pr(Z_t = i | Z_\tau = k, Z_{\tau^-} \neq k, M) d\tau, \quad (5.8)$$

where $\dot{Q}_{j,k}(t_z, \tau|M)$ is the probability of transition from state j to state k at time τ given that state j is reached at time t_z. It should be pointed out here that it is assumed that the observation intervals are small enough, so that transition times can be approximated by the endpoints of observation intervals.

The term $\Pr(Z_t = i|Z_\tau = k, Z_{\tau^-} \neq k, M)$ is the probability of being in state i at time t, given that state k is reached at time τ. This term is the solution of the following systems of equations:

$$\phi_{i,j}(s, t|M) = Pr(Z_t = j|Z_s = i, Z_{s^-} \neq i, M) = (1 - H_i(s, t - s|M))\delta_{i,j} + \sum_l \int_s^t \dot{Q}_{i,l}(s, \tau|M) \times \phi_{l,j}(\tau, t|M) d\tau, \quad (5.9)$$

where $\delta_{i,j}$ is the Kronecker delta. Directly solving the above equation is very complicated and instead approximation methods such as discretization can be employed. More details on how to approximately solve Eq. (5.9) by discretization can be found in [135, 145]. After the solution of Eq. (5.9) is found, Eqs. (5.6)-(5.8) can be used to obtain $\upsilon_t(i|O_1, O_2, ..., O_p, L > t_p, M)$.

5.3.2 Reliability Function

One of the most important prognostic measures is the conditional reliability function, which represents the probability that the device continues operation beyond a time point t ($L > t$) that is greater than the current time point t_p, given that the device has not failed yet ($L > t_p$) and the sequence $O_1, O_2, ..., O_p$ is observed. This important measure is the fundamental element of the dynamic replacement model given in Chapter 6. Most prognostic measures are calculated based on this important measure. The conditional reliability function can be defined as:

$$R(t|O_1, O_2, ..., O_p, L > t_p, M) = \Pr(L > t|O_1, O_2, ..., O_p, L > t_p, M) = \frac{\Pr(L > t, O_1, O_2, ..., O_p, L > t_p|M)}{\Pr(O_1, O_2, ..., O_p, L > t_p|M)} = \frac{\sum_{j=1, j\neq N}^{N} \Pr(L > t, O_1, O_2, ..., O_p, Q_p = j|M)}{\sum_{j=1, j\neq N}^{N} \Pr(O_1, O_2, ..., O_p, Q_p = j|M)}, \quad (5.10)$$

where L is the total age of the device, $\sum_{j=1,j\neq N}^{N} \Pr(O_1, O_2, ..., O_p, Q_p = j|M) = \sum_{j=1,j\neq N}^{N} \alpha_{t_P}(j, O|M)$ and

$$\Pr(L > t, O_1, O_2, ..., O_p, Q_p = j|M) = \sum_{z=0}^{p} u_z(j, O|M) \times \prod_{w=z+1}^{p} b_j(O_w) \times \Pr(L > t, Q_p = j|Q_{z-1} \neq j, Q_z = j, M). \quad (5.11)$$

The remaining step is to find $\Pr(L > t, Q_p = j|Q_{z-1} \neq j, Q_z = j, M)$, which is the joint probability of being in state j at at time t_p and the reliability of the device at time t, given that state j is reached at the zth CM point (unconditional to the condition monitoring information). This measure can be calculated as:

$$\Pr(L > t, Q_p = j|Q_{z-1} \neq j, Q_z = j, M) \approx (1 - H_j(t_z, t - t_z|M)) + \sum_{k\neq N} \int_{t_p}^{t} \dot{Q}_{j,k}(t_z, \tau|M) \times R(t|k, \tau, M)d\tau, \quad (5.12)$$

where $R(t|i, s, M) = \Pr(L > t|X_n = i, T_n = s, M)$ is the solution of the following system of equations:

$$R(t|i, s, M) = (1 - H_i(s, t - s|M)) + \sum_{j\neq N} \int_{s}^{t} \dot{Q}_{i,j}(s, \tau|M) \times R(t|j, \tau, M)d\tau. \quad (5.13)$$

Eq. (5.13) can be solved by either directly solving the discretized system of evolution equation or by backward recursive method as reported in [146]. In this backward approach, $R(t|i, s, M)$ is first calculated at all possible points for state $N-1$ through $R(t|i, s, M) = 1 - H_{N-1}(s, t - s)$ (see Eq. (3.10)). Then, it is recursively calculated for states $N - 2, N - 3, ...,$ and 1, respectively. Another equivalent approach to calculate the conditional reliability function is from Eq. (5.6) as:

$$R(t|O_1, O_2, ..., O_p, L > t_p, M) = \sum_{i\neq N} \upsilon_t(i|O_1, O_2, ..., O_p, L > t_p, M). \quad (5.14)$$

The above relationship simply means that the conditional reliability at time t is the conditional probability of not being in the failure state N at time t. In this thesis, backward recursive is used to estimate the reliability function (with discretization step of 1). Therefore, all results calculated based on the conditional reliability function are subject to discretization error.

5.3.3 Remaining Useful Life Distribution

Remaining useful life (RUL), also called remaining service life, residual life or remnant life, refers to the time left before observing a failure given the current machine age and condition, and the operation profile [17]. It is defined as the conditional random variable $RUL_t = (L - t|L > t, \vartheta(t))$, where L denotes the random variable of failure time (age of the device), t is the current age, and $\vartheta(t)$ is the past condition profile up to the current time t. Remaining useful life (RUL) is one of the most important measures used in online health monitoring of devices under gradual degradation. RUL provides maintenance decision makers a useful tool by quantifying how much time is left until the failure point. When condition monitoring data are used to calculate RUL, it is called the conditional (or alternatively called dynamic) remaining useful life. The distribution of remaining useful life can be found from the conditional reliability function given in Eq. (5.10). The cumulative distribution function (CDF) of remaining useful life at time t, given that the device has not failed yet and condition monitoring data up to time t_p, can be calculated as:

$$\Pr(RUL_{t_p} \leq t|O_1, O_2, ..., O_p, M) = 1 - R(t + t_p|O_1, O_2, ..., O_p, L > t_p, M). \quad (5.15)$$

Based on the above cumulative distribution function (CDF) of the remaining useful life, the probability density function of the remaining life can be calculated. In real-world prognostics, the average of the remaining useful life, referred to as the mean remaining life or mean residual life (MRL), is used to represent the estimated time to failure at each point of time. The MRL at time t_p (time of the pth observation point) can be directly calculated from Eq. (5.10) as:

$$MRL(t_p|O_1, O_2, ..., O_p, L > t_p, M) = \int_0^\infty R(t_p + x|O_1, O_2, ..., O_p, L > t_p, M)dx. \quad (5.16)$$

Using Eq. (5.15), the corresponding conditional percentile confidence interval (or alternatively called prediction interval of percentile interval) for the remaining life can be calculated at any time point. The lower bound (LB) and the upper bound (UB) for the $(1 - \alpha)\%$ percentile interval are respectively the $\alpha/2$ and $1 - \alpha/2$ percentiles of the remaining useful life distribution. The percentile confidence interval for the remaining useful life provides a better picture of the actual remaining useful life as it provides the possible range for the remaining life at a certain level of confidence. The variance of the remaining useful life at time t_p denoted by

$(Var(t_p|O_1, O_2, ..., O_p, M))$ is another important property of the remaining useful life. Thus, variance can be calculated as follows:

$$\begin{aligned}
&Var(t_p|O_1, O_2, ..., O_p, L > t_p, M) = \\
&E((L - t_p)^2|O_1, O_2, ..., O_p, L > t_p, M) - (MRL(t_p|O_1, O_2, ..., O_p, L > t_p, M))^2 = \\
&2\int_0^{\infty} xR(t_p + x|O_1, O_2, ..., O_p, L > t_p, M)dx - MRL(t_p|O_1, O_2, ..., O_p, L > t_p, M)^2.
\end{aligned} \tag{5.17}$$

5.3.4 Warning Levels

Implementing maintenance actions usually requires maintenance setup activities. As maintenance setup times may be costly and time-consuming, it is crucial for maintenance decision makers to consider a maintenance lead time for maintenance decision making. For example, if the maintenance lead time for replacing a gearbox is 30 days, the maintenance decision maker may initiate maintenance setup at least 30 days before the expected time of failure to avoid downtime. Therefore, maintenance decision making for systems with maintenance setup times involves two important decisions: (1) when to initiate maintenance setup and (2) when to stop the operation of a degraded device for replacement. The objective is to make sure that maintenance setup is completed when the device is ready to be replaced.

Most previous measures introduced in this thesis provide information on the time of failure, which may only be useful to determine when to replace a degraded device. To consider maintenance setup time, warning levels can be defined in a way that the importance of performing necessary maintenance actions is categorized into several risk levels. Defining several warning levels helps maintenance decision makers to understand how their production is likely to stop according to either an unexpected failure or maintenance lead time. Let us assume that there are W warning levels for the device under study. For each warning level i, the probability range of failure within the next d units are denoted as (p_1^i, p_2^i). The warning levels are defined as follows:

Definition: The device is considered to be at the ith warning level at time t, if the probability that the device fails in the next d units of time is greater than p_1^i and less than p_2^i, where p_1^i and p_2^i are the ith warning level thresholds defined by maintenance decision makers, and d is the maintenance setup time. Now, the warning level of the device at time t_p (time of the pth observation point) is as:

$$WL(t_p|O_1, O_2, ..., O_p, M) = \sum_{i=1}^{W} i \times I_{p_1^i \leq \Pr(L<t_p+d|O_1,O_2,...,O_p,L>t_p,M)<p_2^i}, \qquad (5.18)$$

where I_A is the indicator function, which is equal to 1 when A is true.

Depending on the value of $\Pr(L < t_p + d|O_1, O_2, ..., O_p, L > t_p, M)$, the warning level can get any integer value between 1 and W. For example, a maintenance team may define three levels of warning as normal, risky, and very risky with the associated probability threshold of [0,50%), [50%,75%) and [75%,100%]. Now, if at time t, the conditional probability of failure within the next d time units is between [75%,100%], then the device is considered to be in the 3rd warning level at time t. This means that in terms of the ability of the maintenance team to finish maintenance setup before the device fails, the current warning level of the device is very risky. A decision making tool that can employ condition monitoring data to determine when to initiate maintenance setup needs to be developed in future work.

5.4 Numerical Example

In this section, simulation-based numerical experiments are used to evaluate the correctness of the main diagnostic and prognostic measures given in this chapter and demonstrate their applications in online monitoring of deteriorating systems. The results are reported separately for diagnostic and prognostic measures.

5.4.1 Example Description

The results presented in this chapter is based on 100 samples simulated for a fleet of devices of the same type. Each device starts with a different degree of initial unknown wear and manufacturing variation, which is considered to be normal, i.e., it is not considered a fault condition. The device is operating normally at the start of each time series, and develops a fault at a point during the series. The fault grows in magnitude until the system failure. It is assumed that the device can have 5 different levels of health states, referred to as: (1) normal condition, (2) slightly damaged, (3) medium damaged, (4) severely damaged, and (5) failure. Each record in a run-to-failure trajectory, which corresponds to a given operation cycle, has the output of the observation process. Here, operation cycle refers to one unit of operation time. All failures occur due to either a gradual degradation or a random failure directly from an intermediate state.

A single indicator is monitored periodically at the end of each cycle (Δ=1), while the equipment is operating. This means that the time between two observation points is equivalent to one operation cycle. The temporal sequence of indicator values is referred to as the observation process. The stochastic behavior of the multistate equipment with degradation states under the non-homogeneous continuous-time hidden semi-Markov assumptions is simulated by the Monte Carlo simulation method and corresponding observation process and degradation process are generated. It is assumed that the CM indicator has 10 possible discrete outcomes ($M = 10$). Therefore, the observation probability matrix B has 5 rows ($N = 5$) and 10 columns ($m = 10$). For transition distributions, the Weibull distribution, which is the most commonly used distribution to represent degradation [141], is used. It is assumed that the transition to the neighbor degraded state depends on the level of degradation and the time spent at each state, while the transition to the failure state depends on the level of degradation and the total age of the device. The transition from state 4 to state 5 also depends on the level of degradation, the time spent at state 4, and the total age of the device. The mathematical formula for transition rate between states is as follows:

$$\lambda_{i,j}(s,t) = \begin{cases} (\beta_{i,j}/\alpha_{i,j}) \times (t/\alpha_{i,j})^{(\beta_{i,j}-1)}, & (i,j) \in [(1,2),(2,3),(3,4)], \\ (\beta_{i,j}/\alpha_{i,j}) \times ((s+t)/\alpha_{i,j})^{(\beta_{i,j}-1)}, & (i,j) \in [(1,5),(2,5),(3,5)], \\ (\beta_{4,5}/\alpha_{4,5}) \times (t/\alpha_{4,5})^{(\beta_{4,5}-1)} + & \\ (\beta^2_{4,5}/\alpha^2_{4,5}) \times ((s+t)/\alpha^2_{4,5})^{(\beta^2_{4,5}-1)}, & (i,j) = (4,5) \end{cases} . \tag{5.19}$$

The parameters used in Eq. (5.19) and the elements of the observation probability matrix (B) are as follows:

$$\alpha = \begin{bmatrix} 0 & 15 & 0 & 0 & 75 \\ 0 & 0 & 20 & 0 & 70 \\ 0 & 0 & 0 & 15 & 70 \\ 0 & 0 & 0 & 0 & 15 \end{bmatrix}, \beta = \begin{bmatrix} 0 & 3 & 0 & 0 & 3 \\ 0 & 0 & 4 & 0 & 6 \\ 0 & 0 & 0 & 5 & 10 \\ 0 & 0 & 0 & 0 & 8 \end{bmatrix}, \alpha^2_{4,5} = 65,\ \beta^2_{4,5} = 14,$$

$$B = \begin{bmatrix} 0.55 & 0.30 & 0.10 & 0.05 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.05 & 0.15 & 0.35 & 0.25 & 0.10 & 0.05 & 0.05 & 0 & 0 & 0 \\ 0 & 0 & 0.05 & 0.05 & 0.40 & 0.35 & 0.15 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.05 & 0.15 & 0.45 & 0.25 & 0.05 & 0.05 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1 & 0.3 & 0.6 \end{bmatrix}.$$

In Figure 5.1, a sample realization of the observation process and its corresponding degradation process obtained from simulation are shown. This figure represents the stochastic behavior of one of the 100 simulated samples. It can be seen from this figure that as the device ages over time, it′s level of degradation increases. Also, the CM output varies with time and indirectly reflects the deterioration process.

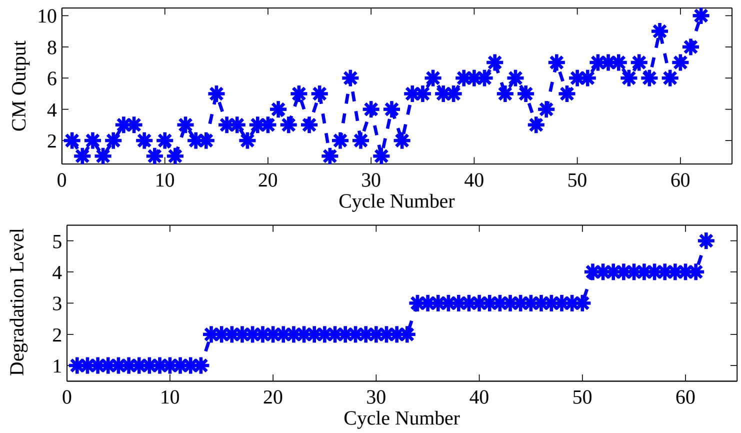

Figure 5.1: Sample realization of the degradation and observation processes

5.4.2 Diagnostic Results

In this subsection, the correctness of the main diagnostic measures provided for the degradation level are evaluated through the 100 simulated samples. First, the most likely state (MLS) at each observation point estimated from condition monitoring data is treated as the diagnosed (estimated) level of degradation. The result is then compared with the actual level of degradation. The classification accuracy or the degradation detection accuracy (%) is defined as the percentage of time-points at which the diagnosed level of degradation using MLS equals the actual level of degradation. Let d_k be the number of observation points for the kth simulated sample and $t_i^{(k)}$ is the corresponding time point of the ith observation point. The degradation detection accuracy for sample k can now be calculated as:

$$\text{Degradation Detection Accuracy } (k) = 100 \times \frac{\sum_{i=1}^{d_k} I_{MLS(t_i^{(k)}|O^{(k)},M)=Q_i^{(k)}}}{d_k}. \qquad (5.20)$$

The classification accuracy for the 100 samples shown in Figure 5.2, verifies that overall, the most likely state calculated using the most updated condition monitoring data (see Eq. (5.1)) can provide reasonable estimates for the actual level of degradation. In other words, this measure can be used for online degradation diagnosis of a deteriorating device using condition monitoring data.

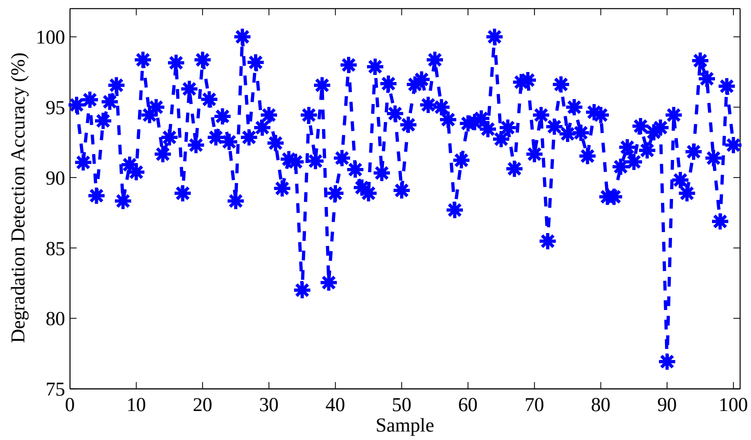

Figure 5.2: Degradation detection accuracy (%) using MLS

In addition to the most likely state, the estimated average degradation level (ADL) as shown in Eq. (5.2) can also be used to represent the actual degradation process. The main difference between the MLS and ADL is that the former gives an integer value corresponding to the most likely state while the latter gives the average state level, which can get any value between 1 and N. In Figure 5.3, the mean squared error of estimating the actual degradation level with the average degradation level for each sample is presented. The MSE of the degradation detection accuracy for the k sample can be fined as:

$$\text{MSE of ADL}(k) = \frac{\sum_{i=1}^{d_k} (ADL(t_i^{(k)}|O^{(k)}, M) - Q_i^{(k)})^2}{d_k}. \tag{5.21}$$

It can be observed from Figure 5.3 that estimation results based on average degradation level is also very close to the actual degradation level for almost all samples and can be used as an effective tool for online degradation diagnosis of deteriorating devices.

As described earlier, the introduced dynamic diagnostic measures are expected to have a non-decreasing trend with the age of the device. To visually demonstrate the monotonicity of the proposed dynamic measures, the results for the conditional hazard rate (Eq. (5.4)) and average degradation level are shown for samples 1-4. For the average degradation level (Figure 5.5), the actual degradation level is also shown. It can be seen from Figures 5.4-5.5 that overall, these two measures have non-decreasing trends over the life of the device, that is, their absolute values do not decrease (they either increase or stay the same) with the age of the device.

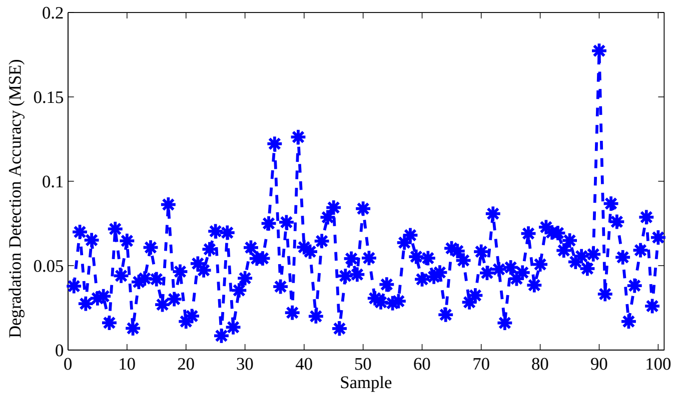

Figure 5.3: Degradation detection accuracy (MSE) using ADL

Such measures with a rough monotonic trend are of high interests for maintenance decision makers and can be used as an online tool to identify the current health status of mechanical devices under degradation and condition monitoring.

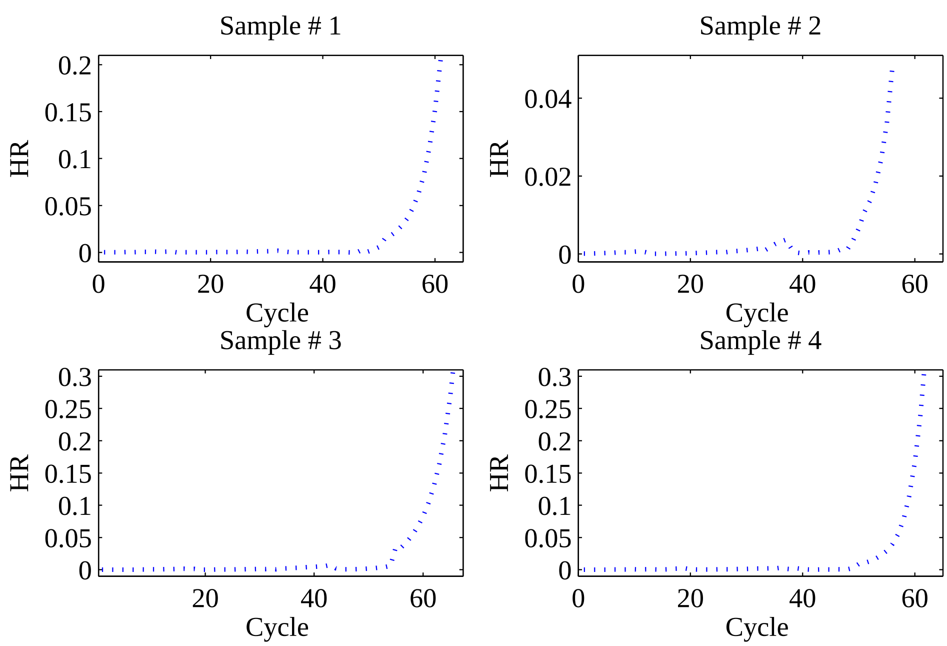

Figure 5.4: Conditional hazard rate for samples 1-4

5.4.3 Prognostic Results

As discussed earlier in this chapter, the objective of prognostic measures is to provide some information on the future levels of degradation associated with the device under study. The main measures being investigated in this section are the residual life and warning level. In this subsection, the effectiveness of using

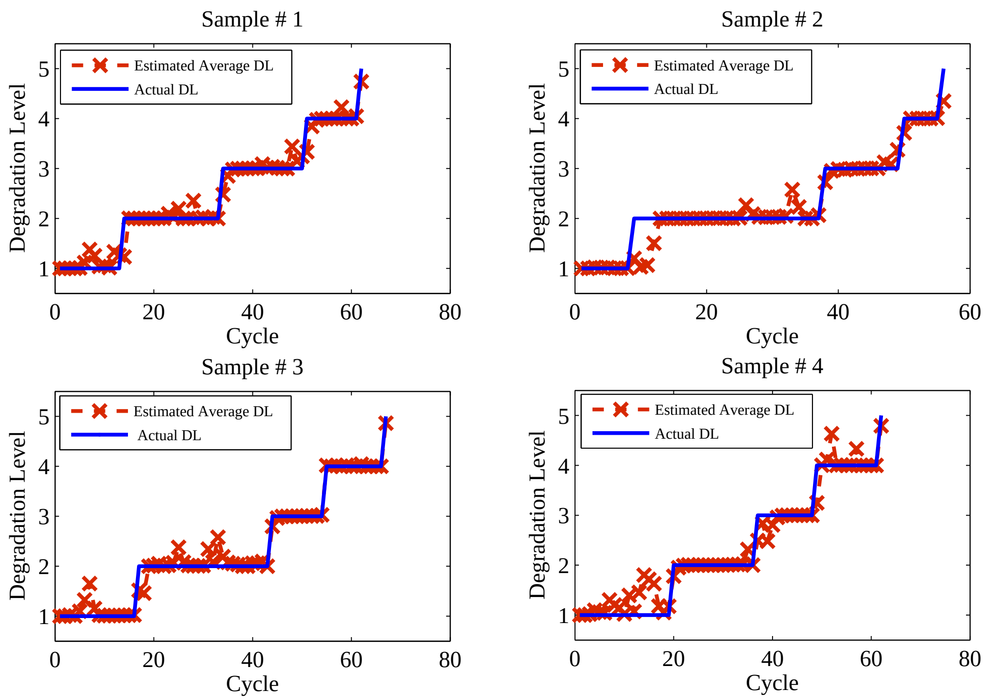

Figure 5.5: Average degradation level-actual degradation level for samples 1-4

mean remaining life as a measure to estimate the actual remaining life is presented. To enable the comparison of the prognostics results for all 100 samples, multiple prediction points were considered at different stages (in terms of % of lifetime). The mean remaining life was estimated at 10 different points equivalent to $50\%, 55\%, 60\%, 65\%, 70\%, 75\%, 80\%, 85\%, 90\%$, and 95% of the real lifetime of each sample. Therefore, for each sample, mean remaining life is calculated 10 times at different stages of life. The estimated MRL values are then compared to the actual remaining life. For each sample, the relative error is used to measure the error of estimation as Relative Error (RE) $= \frac{|\text{Estimated MRL - Actual Remaining Life}|}{\text{Total Life}}$. Table 5.1 presents the average and the standard deviation of the relative error for 100 samples.

It can be seen from Table 5.1 that overall, the estimated mean residual lives are reasonably close to the actual remaining life and therefore can be used as an online measure to estimate the remaining life of the device. It can also be verified from Table 5.1 that the estimation results get closer the true values as the device ages, i.e., MRL is a better estimate for remaining useful life when the device is older.

In Figure 5.6, the estimated mean total life (current time + estimated mean

Table 5.1: Prognostic results at different points (in terms of % of lifetime)

% of Lifetime	50	55	60	65	70	75	80	85	90	95
Mean of RE (%)	6.80	6.36	5.94	5.55	5.12	5.20	5.08	4.83	4.12	3.40
STD of RE (%)	5.64	5.48	5.21	4.25	4.12	4.10	3.92	3.74	3.60	3.47

residual life), its 95% percentile confidence interval, and the actual total life for the 100 samples made at time points equivalent to 80% lifetimes are presented. The results are shown in an increasing order of total lifetime. It can be seen from this figure that in most cases the estimated total life (80% lifetime + estimated mean residual life) is relatively close to the actual total life. Also in almost all cases, the actual life falls within the provided confidence intervals. The coverage rate (percentage of covering the true values of the parameter in repeated sampling) calculated for the 100 samples is 95%. Similar results were observed for mean residual estimation and its associated prediction interval calculated at other time points (% of lifetime).

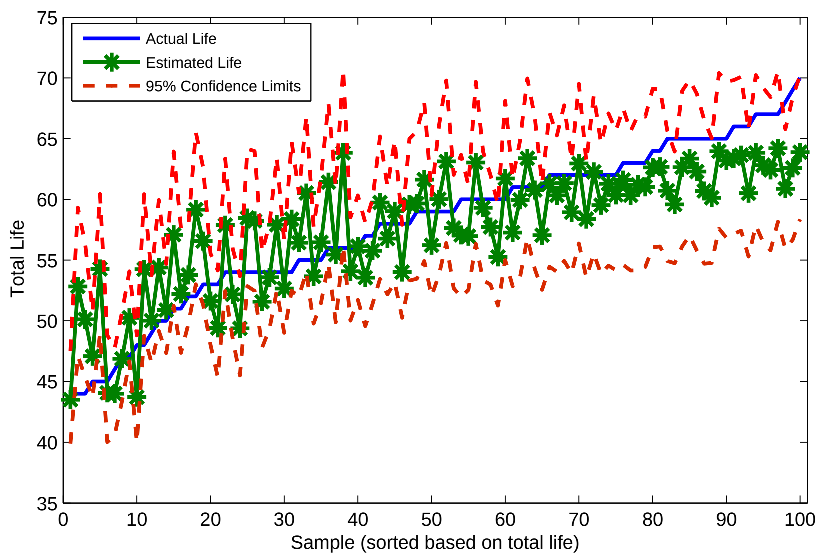

Figure 5.6: Estimated total life and its confidence interval for 100 samples (calculated at 80% total life of each sample)

To demonstrate how conditional mean residual life can be used as an online tool for remaining useful life estimation, in Figure 5.7, the results for 4 samples are reported. As shown in this figure, (1) the mean residual life is a reasonable estimate for the remaining useful life, (2) mean residual life gets closer to the actual value as the device ages, (3) the true values lies within the 95% percentile interval (in almost all cases), and (4) percentile interval gets narrower as the device ages.

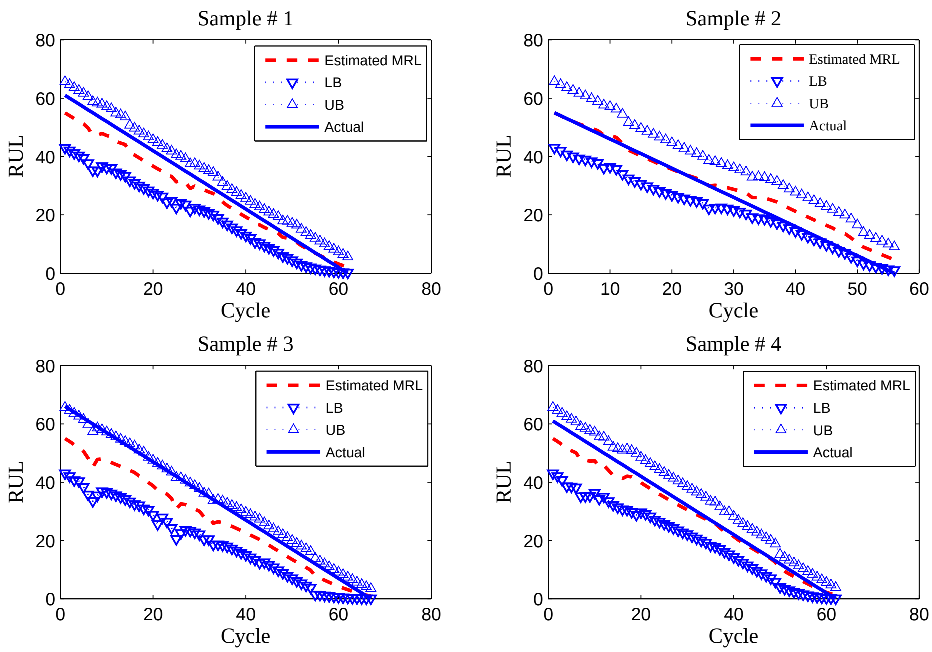

Figure 5.7: RUL analysis for samples 1-4

The final prognostic measure studied here is the warning level. Warning levels correspond to the condition of the device in terms of being able to survive until the maintenance setup is completed. Assume that the maintenance initiation setup time is 10 cycles, and the probability intervals associated with the four levels of warning are defined as $[0\%, 50\%)$, $[50\%, 75\%)$, $[75\%, 90\%)$, and $[90\%, 100\%]$. For example, the device is considered to be in the warning level 1, if the probability of failure within the next 10 cycles is less than 50%. In Figure 5.8, the probability of failure within the next 10 cycles and the associated probabilities for the warning levels (dash lines) are shown for sample No. 1. Four different areas (warning levels) are observable in this figure, where each area represents one warning level. Such results can be generated and employed as an online tool to determine the levels of risk associated with being able to survive until the maintenance initiation setup is completed.

5.5 Summary

One of the most commonly used tools for degradation diagnostics and prognostics for mechanical systems is to track their health status using some important per-

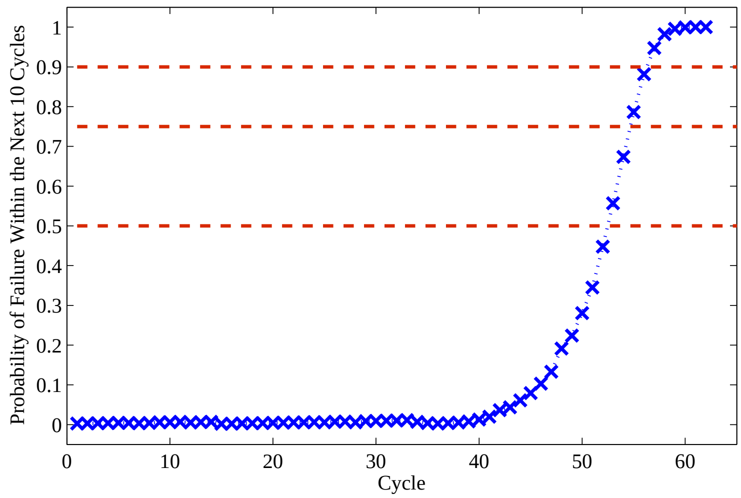

Figure 5.8: Failure probability within the next 10 cycles - sample No.1

formance measures. A performance measure for a condition-monitored device is considered dynamic if available condition monitoring data is employed for calculating that measure. Two types of dynamic performance measures can be employed for degradation analysis of deteriorating systems referred to as (1) diagnostic measures and (2) prognostic measures. In this chapter, the formulas for some important dynamic measures are introduced. Each measure is defined to reflect a certain aspect of the degradation process so that it can be employed as an online tool for diagnostics and prognostics. Finally, the application of some of these measures was shown through numerical examples. In the next chapter, some of these measures are used to determine the optimal strategy for the replacement, that is, when to replace a degraded device.

Chapter 6

Condition-Based Replacement Model

6.1 Introduction

In previous chapters, a model for a multistate degradation process (Chapter 3), an unsupervised estimation method to characterize the parameters of the degradation and observation processes (Chapter 4), and important diagnostic and prognostic measures for online health monitoring (Chapter 5) are introduced. In this chapter, using the results given in previous chapters, a maintenance model is developed, which can employ dynamic information on the current health status of the device for decision making. Maintenance activities are the set of actions during the life cycle of a system (or a device), which intend to keep it in working condition as much as possible. Condition-based maintenance is a type of preventive maintenance, which is based on monitoring a parameter of the condition of a device through methods such as direct measurement, inspection, and condition monitoring. The ultimate objective of condition monitoring is to be able to provide some useful information that can be used for maintenance decision making in order to avoid unnecessary maintenance actions and prevent catastrophic and costly failures. Maintenance decision making in the presence of condition monitoring data involves finding the answer for the key question, that is, when (under what condition) to replace a degraded device. When the device is under condition monitoring, decision makers can utilize the most updated information on the degradation and observation processes obtained through condition monitoring indicators for decision making.

A replacement policy that utilizes some information on the actual degradation and observation processes is a dynamic condition-based replacement policy, that is, the dynamic behavior of the degradation process can affect maintenance decisions. In this thesis, it is assumed that the replacement policy is a block replacement policy where the maintenance decision points are exactly the condition monitoring points. In other words, at each condition monitoring point and after observing the most updated condition monitoring data, the decision on whether to replace a degraded device or wait until the next observation point is made. As a result of a block replacement policy, the device is finally replaced at failure (failure replacement) or at a certain monitoring point (preventive replacement), whichever occurs first. Therefore, this type of replacement policy is a condition-based replacement policy, in which the history of the observation process until the decision-making points are used for decision making and the maintenance actions depend on the condition of the device. It should be pointed out here that replacement can be a physical replacement or a full repair (overhaul), such that the device becomes as good as new after the repair. As very similar kinds of condition-based replacement policies have been discussed in the literature as in [41, 114, 147], this chapter aims to modify available results so that a condition-based replacement model for a device under multistate degradation and condition monitoring can be developed. Some of the notation used in this chapter is adopted from [41, 147].

This chapter presents the mathematical framework for a dynamic replacement policy for a device with multistate health levels and unobservable states, where the degradation process follows a NHCTHSMP. The concepts used in solving a semi-Markov decision process (SMDP) are employed to solve this problem. The multistate model discussed in Chapter 3 with a known structure and characteristic parameters as described in Chapter 4 are used as the inputs of this chapter. Also, some important prognostic measures such as conditional reliability function are used to develop the replacement model.

This chapter is organized as follows. In Section 6.2, the assumptions and elements of the cost-effective replacement policy are introduced. The mathematical details for the proposed replacement policy in a control-limit form are described in Section 6.3. A simulation-based numerical example is used in Section 6.4 to show how such dynamic replacement policy can be used for maintenance decision making. Also, the effectiveness of the proposed structure is demonstrated. Finally, Section 6.5 summarizes the results of this chapter.

6.2 Assumptions of the Maintenance Model

The objective of the proposed replacement policy is to minimize the expected long-run average unit cost of the device, which depends on the long-run average cost and the long-run average duration of a replacement cycle. Therefore, it is important to identify the elements that affect the cost and duration of a replacement cycle. Here, replacement cycle refers to the interval between two replacement points. It should be pointed out here that the assumption made for the degradation and observation processes associated with the device are the same as those described in Section 3.2. The detailed assumptions for the cost elements of the replacement model are described below.

- The device operates until it fails or replaced, whichever occurs first.

- Maintenance decisions are made only at certain discrete points, which are equivalent to the condition monitoring observation points. These points are referred to as maintenance decision making points hereafter. The results presented in this chapter are based on the assumption that the intervals between two decision points are constant (Δ), however, results can be converted to the case with nonidentical maintenance decision intervals. Also, it is assumed that the time between two observation points is short enough, so that at most one degradation transition can occur.

- There are two possible maintenance decisions (actions) at each decision point: (1) replace the device immediately, and (2) do nothing, that is, wait until the next decision point. The set of possible maintenance actions is denoted by $A = \{1, 0\}$, where 1 means preventive replacement (or replace immediately) and 0 means do nothing, that is, wait until the next decision epoch.

- The replacement of the device (whether it is a failure replacement or a preventive replacement) costs c_r and takes t_r units of time. In other words, c_r and t_r are the cost and time of a replacement, respectively.

- There is additional cost and time for a failure replacement denoted by c_f and t_f, respectively. As a result, the total cost and time of a failure replacement are $c_r + c_f$ and $t_r + t_f$, respectively.

- The system downtime is also subject to unit downtime cost of c_d. Therefore, the downtime cost of a preventive replacement is $c_d \times t_r$ and the downtime cost of a failure replacement is $c_d \times (t_r + t_f)$. Therefore, the total cost per unit-time of preventive replacement and failure replacement are respectively $(c_r + c_d \times t_r)/t_r$, and $(c_r + c_f + c_d \times (t_r + t_f))/(t_r + t_f)$.

- It is assumed that the cost of extracting condition monitoring data over time is zero.

- The condition of the device at time t_p (pth decision point), which is denoted by ϑ^{t_p}, includes two types of known information, (1) the sequence of observed condition monitoring data until time point t_p ($Y^{t_p} = O_1, O_2, ..., O_p$) and the overall health status of the device ($\Psi^{t_p} = 1$, that is the device is still operating at time t_p and $\Psi^{t_p} = 0$, that is the device is failed at time t_p). Therefore $\vartheta^{t_p} = \{Y^{t_p}, \Psi^{t_p}\}$. The set of all possible conditions for the device is denoted by S, for which we have $\vartheta^t \in S, \forall t > 0$. As it is assumed that the failure state is self-announcing, that is, the overall health status of the device (Ψ) is known over time. This means that the condition of the device is known over time.

6.3 Condition-Based Replacement Model

This section presents the structure of the condition-based replacement policy for a device with multistate health conditions, which is under condition monitoring.

6.3.1 Elements of the Condition-Based Replacement Model

The condition-based maintenance model considered in this chapter determines the best maintenance actions at certain discrete points, referred to as maintenance decision points, which are equivalent to condition monitoring observation points. The objective of this condition-based maintenance policy is to minimize the long-run average unit cost of the device considering replacement cost, failure cost, and downtime cost. The condition-based replacement policy problem is formulated through a semi-Markov decision process (SMDP) [148], which has been used in the literature to find the solution of dynamic maintenance problems. Some applications of SMDP in maintenance modeling can be found in [149, 150]. Markov and semi-Markov decision models are powerful tools for analyzing stochastic sequential decision processes with infinite horizons [148]. The main elements of the proposed condition-based replacement policy are as follows:

I. $A = \{1, 0\}$: The set of maintenance decision actions, where 1 means preventive replacement (PR) and 0 means do nothing (DN).

II. $\vartheta^t = \{Y^t, \Psi^t\}$: The condition of the device at time t, which includes Y as the set of observed condition monitoring data from time zero to time t and Ψ as the overall health level of the device. The condition of the device is known over time.

III. S: The set of possible conditions for the device ($\Psi^t \in S, \forall t$).

IV. γ: A condition-based replacement policy, which determines which maintenance action to choose at each decision point depending on the condition of the device. In other words, a replacement policy ($\gamma : S \to A$) is a decision policy as a rule for replacement, which depends only on the condition of the device (ϑ). For any given condition, the replacement policy should choose a single action from A.

V. $D(\vartheta, \gamma)$: The output (maintenance action) of replacement policy γ when the condition of the device is ϑ. As there are only two possible maintenance actions, then (1) $D(\vartheta, \gamma) = 0$, refers to do nothing and (2) $D(\vartheta, \gamma) = 1$, refers to immediate replacement.

In the decision process with the above elements, the cost incurred until the next interval and the time until the next decision point depend on the condition of the device and the maintenance action chosen in that condition. Different replacement policies yield different maintenance actions and costs. The objective here is to find the condition-based replacement policy γ^* that minimizes the long-run average cost per unit time of the device. The long-run average cost per unit time is the total long-run cost of a replacement policy (cost of replacement plus cost of failure plus cost of downtime) divided by the total long-run cycle time of a replacement policy (time required for replacement plus additional time of failure plus operation time). It will be shown later that the approach used in this thesis is similar to [41, 114, 147], in the sense that the replacement policy γ belongs to a class of control-limit policy in which the device is replaced when the failure risk (as a control index) exceeds a threshold value defined by policy γ. This control-limit policy will be illustrated in the remainder of this section.

Under policy γ, the expected time and cost from time t to the completion of the replacement cycle (beginning of the next operation cycle), given the condition ϑ^t, can be computed from the following equations in terms of each possible maintenance actions:

$$T_t(\vartheta^t,\gamma) = \begin{cases} t_r, & \text{if } D(\vartheta^t,\gamma)=1 \\ \displaystyle\int_0^{\Delta} R(t+\tau|\vartheta^t)d\tau + (t_r+t_f)\times\big(1-R(t+\Delta|\vartheta^t)\big) + & \\ \displaystyle\sum_{\vartheta^{t+\Delta}\neq F} \Pr(\vartheta^{t+\Delta}|\vartheta^t)\times T_{t+\Delta}(\vartheta^{t+\Delta},\gamma), & \text{if } D(\vartheta^t,\gamma)=0 \end{cases}, \quad (6.1)$$

$$C_t(\vartheta^t,\gamma) = \begin{cases} c_r + c_d\times t_r, & \text{if } D(\vartheta^t,\gamma)=1 \\ (c_r+c_f+c_d\times(t_r+t_f))\times\big(1-R(t+\Delta|\vartheta^t)\big) + & \\ \displaystyle\sum_{\vartheta^{T+\Delta}\neq F} \Pr(\vartheta^{t+\Delta}|\vartheta^t)\times C_{t+\Delta}(\vartheta^{t+\Delta},\gamma), & \text{if } D(\vartheta^t,\gamma)=0 \end{cases}, \quad (6.2)$$

where Δ is the time between two maintenance decision points (observation interval), $R(x|\vartheta^t)$ is the conditional reliability of the device at time x , given the condition of the device at time t as ϑ^t, $(1-R(t+\Delta|\vartheta^t))$ is the conditional probability of failure within the time interval $(t, t+\Delta)$, given the condition ϑ^t, $\int_0^{\Delta} R(t+\tau|\vartheta^t)d\tau$ is the mean operation time within the next maintenance decision interval $(t, t+\Delta)$, $\Pr(\vartheta^{t+\Delta}|\vartheta^t)$ is the probability of observing condition $\vartheta^{t+\Delta}$ at time $t+\Delta$, given condition ϑ^t at time t, and F is any condition at which the device is failed.

Eq. (6.1) simply states that at time t, if the maintenance action based on policy γ is to replace a degraded device, then the remaining time to the start of the next replacement cycle is just t_r, which is the time required for replacement. However, if the maintenance action at time t based on policy γ is to wait until the next decision point (do nothing), then the remaining time to the start of the next replacement cycle is the sum of the expected time to replacement if the device fails within the next decision making interval and the expected time to replacement if the device does not fail within the next decision making interval. Similarly, the cost before the start of the next replacement cycle can be defined based on the two possible maintenance options as shown in Eq. (6.2). As the only option at the point of an unexpected failure is immediate replacement, then we have: $D(\vartheta^t,\gamma)=1$, if $Z_t = N$ (or $\Psi^t = 0$).

The above equations will be used in the remainder of this section to find the optimal replacement policy. In the following sections, the structure of the cost function, replacement policy, and the solution procedure to find the optimal replacement policy are illustrated.

6.3.2 Cost Function

As stated earlier, the final objective of the replacement policy is to minimize the expected long-run average cost per unit time of the device. According to the renewal theory, the cost function can be represented as:

$$g(\gamma) = \frac{E(C_0(\vartheta^0, \gamma))}{E(T_0(\vartheta^0, \gamma))}, \tag{6.3}$$

where $g(\gamma)$ is the average cost per unit time for the policy γ and ϑ^0 is the condition of the device at time 0, that is, the device is at its perfect functioning state and no condition monitoring data have been recorded yet. Here, $E(T_0(\vartheta^0, \gamma))$ and $E(C_0(\vartheta^0, \gamma))$ are the expected time and cost of a replacement cycle, respectively. Now let γ^* be the optimal condition-based replacement policy and g^* be the corresponding optimal long-run average cost per unit time of the device. Then we have:

$$g^* = \min_{\gamma} \frac{E(C_0(\vartheta^0, \gamma))}{E(T_0(\vartheta^t, \gamma))} = \frac{E(C_0(\vartheta^0, \gamma^*))}{E(T_0(\vartheta^t, \gamma^*))}. \tag{6.4}$$

As stated earlier, the cost of a preventive replacement is $c_r + c_d \times t_r$ and the additional cost of a failure replacement is $c_f + c_d \times t_f$. Therefore, the expected cost of a replacement cycle is $c_r + c_d \times t_r + (c_f + c_d \times t_f) \times E(\Pr(L \leq T_\gamma))$, where L is the total life of the device, T_γ is the replacement time determined by policy γ, and $E(\Pr(L \leq T_\gamma))$ is the expected probability of a failure replacement in a replacement cycle. Also, as the time required for a preventive replacement is t_r and the time required for a failure replacement is $t_r + t_f$, the expected sum of replacement time and failure time in a replacement cycle is $t_r + t_f \times E(\Pr(L \leq \gamma))$. It should be noted that the expected length of the replacement cycle is the sum of replacement time, failure time, and operation time. Therefore, the expected length of a replacement interval is $t_r + t_f \times E(\Pr(L \leq T_\gamma)) + E(\min(L, T_\gamma))$, where T_γ is the replacement time based on policy γ and $E(\min(L, T_\gamma))$ is the expected operation time of an operation cycle based on policy γ. Based on the above discussion and according to the theory of renewal reward process [151], the long-run average cost per unit time for policy γ is as follow:

$$g(\gamma) = \frac{c_r + c_d \times t_r + (c_f + c_d \times t_f) \times E(\Pr(L \leq T_\gamma))}{E(\min(L, T_\gamma)) + t_r + t_f \times E(Pr(L \leq T_\gamma))}. \tag{6.5}$$

In the next subsection, the structure of the replacement policy as a control-limit policy and steps to be implemented to estimate the cost function and its elements under a known policy γ are illustrated.

6.3.3 Structure of the Replacement Policy

In this subsection, the structure of the dynamic replacement policy for a condition-monitored device is reviewed. It will be shown that this replacement model is from the family of the infinite horizon decision process with average cost function, which can be solved through a well-known method used in dynamic programming, referred to as policy iteration method [149, 150]. As noted earlier, a replacement policy is a mapping function from $\vartheta \in S$ (condition of the device) to $a \in A$ (maintenance actions). Let $V(\vartheta, \gamma)$ be the relative average cost associated with policy γ in infinite horizon, given that the condition of the device is ϑ. The corresponding policy iteration algorithm can be implemented in a general form as follows [150]:

Algorithm 6 : Policy Iteration Algorithm For a Replacement Policy

Step 1. Initialization: start with a stationary policy (γ_0) and set $\gamma = \gamma_0$.

Step 2. Value Determination: consider $V(\vartheta, \gamma)$ and $g(\gamma)$ as unknowns. Find the solution of these unknowns using the following system of linear equations for each $\vartheta \in S$.

$$V(\vartheta, \gamma) = \\ r(\vartheta, D(\vartheta, \gamma)) - g(\gamma) \times y(\vartheta, D(\vartheta, \gamma)) + \sum_{\vartheta' \in S} \Pr(\vartheta, \vartheta', D(\vartheta, \gamma)) \times V(\vartheta', \gamma), \quad \forall \vartheta \in S, \tag{6.6}$$

where $V(\vartheta, \gamma)$ is the relative expected cost-to-go of replacement policy (γ) in infinite horizon when the initial condition is ϑ, $r(\vartheta, D(\vartheta, \gamma))$ is the immediate cost of replacement policy γ when the condition of the device is ϑ, $y(\vartheta, D(\vartheta, \gamma))$ is the expected transition time of the replacement policy (expected time to the next decision point) γ when the condition of the device is ϑ, and $\Pr(\vartheta, \vartheta', D(\vartheta, \gamma))$ is the probability that the replacement policy γ transfers the condition of the device from ϑ to ϑ'.

Step 3. Policy Improvement: For each $\vartheta \in S$, determine the action a' that yields:

$$\arg\min_{a'} \{ r(\vartheta, a') - g(\gamma) \times y(\vartheta, a') + \sum_{\vartheta' \in S} \Pr(\vartheta, \vartheta', a') \times V(\vartheta', \gamma) \}, \quad \forall \vartheta \in S \tag{6.7}$$

Step 4. Iteration: the resulting optimum actions for each condition $\vartheta \in S$ obtained from Eq. (6.7), yields the new stationary policy γ'. If γ and γ' are identical (or if $g(\gamma) = g(\gamma')$), then γ is optimal. Otherwise, set $\gamma = \gamma'$ and move back to step 2.

It is expected that the policy iteration algorithm converges after a finite number of iterations [148]. Directly applying the above policy iteration for solving the condition-based replacement problem is subject to a challenge. The challenge is that the number of possible conditions for the device ($\vartheta \in S$) can be extremely large. This makes the total number of possible policies beyond any practical bound. If the condition monitoring data can take any of the m possible outputs and the maximum possible number of condition monitoring points is L_{max} (depending on the maximum total life of the device), then the maximum number of possible conditions is $\sum_{i=1}^{L_{max}} m^i$. For example, if the maximum possible failure time is 100, the maintenance decision interval is 1, and there are 8 possible condition monitoring outputs, the number of possible conditions for a device over time is $\sum_{i=1}^{100} 8^i = 2.33 \times 10^{90}$. This makes directly applying the above policy iteration algorithm infeasible in real-world applications. Therefore, it is desirable to find the optimal policy without having to analyze every possible combinations of conditions and maintenance actions [149].

To be able to apply the policy iteration technique, we can modify its steps in the sense that the condition-based replacement policy becomes a condition-based control-limit replacement policy. As a result, a control index dependent on the condition of the device can be used to determine whether or not to replace a degraded device. Such control index is used in many papers as in [41, 147]. Such replacement policy is described in details in the remainder of this chapter. As only two maintenance decision actions are possible at each decision point (do nothing or immediate replacement), the policy improvement step can be redefined in the sense that by comparing the expected cost of each of these two possible actions, the action with lower cost can be determined. Now, let us define the policy improvement system of equations at time t for condition ϑ^t considering the two possible actions. The policy improvement system of equations can be constructed based on the following optimality equation:

$$V(\vartheta^t, \gamma) = \min \begin{cases} V^1(\vartheta^t, \gamma), & D(\vartheta^t, \gamma) = 1 \\ V^2(\vartheta^t, \gamma), & D(\vartheta^t, \gamma) = 0 \end{cases}, \tag{6.8}$$

where $V^1(\vartheta^t, \gamma)$ is the relative average cost of replacement policy γ if the maintenance decision is immediate replacement and $V^2(\vartheta^t, \gamma)$ is the relative average cost of replacement policy γ if the maintenance decision is do nothing (or wait until the next inspection point). Using Eq. (6.7), $V^1(\vartheta^t, \gamma)$ and $V^2(\vartheta^t, \gamma)$ can be calculated as:

$$V^1(\vartheta^t,\gamma) = c_r + c_d \times t_r - g(\gamma) \times t_r + V(\vartheta^0,\gamma), \tag{6.9}$$

and

$$\begin{aligned} V^2(\vartheta^t,\gamma) = &\left[c_r + c_f + c_d(t_r+t_f) + V(\vartheta^0,\gamma)\right]\left[1 - R(t+\Delta|\vartheta^t)\right] - \\ &g(\gamma)\left(\tau_t(t+\Delta|\vartheta^t) + (t_r+t_f)(1-R(t+\Delta|\vartheta^t))\right) + \\ &\left[\sum_{\vartheta^{t+\Delta}\neq F} \Pr(\vartheta^t,\vartheta^{t+\Delta},0) \times V(\vartheta^{t+\Delta},\gamma)\right] \end{aligned}. \tag{6.10}$$

The notation $\tau_t(t+\Delta|\vartheta^t)$ in Eq. (6.10) is the mean operation time in the interval $(t, t+\Delta)$ given ϑ^t, which can be expressed as follows:

$$\tau_t(t+\Delta|\vartheta^t) = \int_0^\Delta R(t+\tau|\vartheta^t)d\tau. \tag{6.11}$$

Now, for any policy γ, it is possible to perform the policy improvement step by comparing the cost of two possible options ($V^1(\vartheta^t,\gamma)$ and $V^2(\vartheta^t,\gamma)$), given the condition of the system (ϑ^t) and the replacement policy (γ). Then, the best decision between the two possible options can be determined for any condition ϑ and any policy γ. According to Eq. (6.9) and Eq. (6.10), if $V^2(\vartheta^t,\gamma) < c_r + c_d \times t_r + V(\vartheta^0,\gamma) - g(\gamma) \times t_r$, then $V(\vartheta^t,\gamma) = V^2(\vartheta^t,\gamma)$, that is, the optimal maintenance action at time t under policy γ is do-nothing ($D(\vartheta^t,\gamma) = 0$) and if $V^2(\vartheta^t,g) > c_r + c_d \times t_r + V(\vartheta^0,g) - g(\gamma) \times t_r \geq 0$, then $V(\vartheta^t,\gamma) = V^1(\vartheta^t,\gamma)$, that is the optimal action is immediate replacement ($D(\vartheta^t,\gamma) = 1$). As shown in Appendix 1, the condition-based replacement policy can be presented as a control-limit replacement policy, which is summarized as follows:

$$\begin{cases} CI(\vartheta^t) \geq g(\gamma) & \rightarrow D(\vartheta^t,\gamma) = 1 \\ \Psi^t = 0 & \rightarrow D(\vartheta^t,\gamma) = 1 \\ CI(\vartheta^t) < g(\gamma) & \rightarrow D(\vartheta^t,\gamma) = 0 \end{cases}, \tag{6.12}$$

where:

$$CI(\vartheta^t) = \left\{ \frac{\left[c_f + c_d \times (t_f)\right]\left[1 - R(t+\Delta|\vartheta^t)\right]}{\tau(t+\Delta|\vartheta^t) + t_f\left[1 - R(t+\Delta|\vartheta^t)\right]} + \frac{\varpi(\vartheta^t,\gamma)}{\tau(t+\Delta|\vartheta^t) + t_f\left[1 - R(t+\Delta|\vartheta^t)\right]} \right.. \tag{6.13}$$

The above replacement policy is a stationary policy (with respect to the condition of the device), which assigns to each condition a fixed maintenance action. The term $CI(\vartheta^t)$ is a control index that can be calculated at each maintenance decision point t, given the condition of the device (ϑ^t). Now, for any given policy γ and its corresponding cost $g(\gamma)$, the decision on replacing a degraded device can be determined using Eq. (6.12). This equation also verifies that the replacement policy can be defined according to its cost. The final replacement time based on policy γ can now be expressed as:

$$T_\gamma = \min\left\{L, t = j\Delta \geq 0; CI(\vartheta^t) \geq g(\gamma)\right\}. \tag{6.14}$$

Now that the structure of the replacement policy in a control-limit form is determined for each policy γ, the next step is to perform the value determination step. It can be seen that the only required unknown in Eq. (6.6) that needs to be found at each step of the policy iteration algorithm is $g(\gamma)$, which is the average cost of a new policy. Therefore, the next steps are to find (1) the expected long-run average cost per unit time of the device in a replacement cycle for policy γ ($g(\gamma)$), and (2) the elements used in Eqs. (6.12)-(6.13) to determine when to replace a degraded device. Following the steps given in the policy iteration algorithm, an initial policy can be used as a start point to iteratively find the optimal replacement policy. The challenging step is the value determination step, which deals with evaluating a policy and estimating the two elements $E(\Pr(L \leq T_\gamma))$ and $E(\min(L, T_\gamma))$, which are respectively the expected probability of failure replacement in a replacement cycle and the expected operation time of a replacement cycle (expected operation time) under policy γ (Eq. (6.5)). In this thesis, a Monte-Carlo simulation method and an analytical approach (Appendix A) are given to estimate these two elements. The mathematical details on how to obtain the replacement policy shown in Eq. (6.12) are given in Appendix A.

6.3.4 Summary of the Replacement Policy

The replacement policy presented in the previous section has two phases, which are training phase and implementation phase. In the training phase, the structure of the optimal replacement policy and its associated cost are determined. In the implementation phase, using condition monitoring data, the trained control-limit-based replacement policy is used for final maintenance decision making. The summary of steps for each phase is illustrated as follows:

I. Training Steps to Develop a Replacement Policy: The steps needed to train the proposed replacement policy are given in Algorithm 7.

The result of the above training steps is a series of replacement policies and their associated costs, which should converge to the optimal replacement policy. Now

Algorithm 7 : Training Steps to Develop a Replacement Policy

Step 1. Set $k = 0$; select an arbitrary replacement policy (γ_0) and its associated cost ($g(\gamma_0)$). Remember, a replacement policy is defined with its associated cost (see Eq. (6.12)). One possible initial policy is to consider Do Nothing for all conditions. The cost of this policy is $\dfrac{c_r + c_d \times t_r + (c_f + c_d \times t_f)}{E(L) + t_r + t_f}$.

Step 2. Find $E(\Pr(L \leq T_\gamma))$ and $E(\min(L, T_\gamma))$ from the Monte-Carlo simulation method or from the approach given in Appendix A using $E(\Pr(L < T_\gamma | L > 0, \vartheta^0))$, $E(\min(L, T_\gamma) | L > 0, \vartheta^0)$. In the associated Monte-Carlo simulation, random temporal degradation processes and their associated observation processes are generated. Then, control-limit replacement policy (Eq. (6.12)) is applied on each simulated run-to-failure data to determine when to replace the degraded device. The term $E(\Pr(L \leq T_\gamma))$ is the percentage of times that the device failed before the replacement policy could determine the replacement time (percentage of failure replacement) and $E(\min(L, T_\gamma))$ is the average of the operation cycle considering the replacement time of each sample based on the current policy. It is clear that a larger number of simulation runs and simulated samples can improve the correctness of estimation results. The elements used in Eq. (6.13) are calculated according to the results shown in Appendix A.

Step 3. Set $k = k + 1$ and find $g(\gamma_k)$ using Eq. (6.5). Here, $g(\gamma_k)$ is the cost associated after applying the kth policy γ.

Step 4. If $g(\gamma_k) = g(\gamma_{k-1})$, then γ_k is the optimal replacement policy ($\gamma^* = \gamma_k$). Otherwise, move back to step 2.

that the optimal condition-based replacement policy (γ^*) and its associated cost ($g(\gamma^*)$) are known, Eq. (6.12) can be used as an online tool to determine whether or not to replace a degraded device at a certain condition. In the implementation phase, the result of replacement policy on each sample is determined. The summary of the implementation phase is as follows:

II. Implementation Steps: The steps needed to implement the proposed replacement policy as an online tool for maintenance decision making is given in Algorithm 8. It should be pointed out here that the device is always replaced in the case of an unexpected failure. Then, steps 1-3 need to be repeated to find the replacement time of the new device.

Algorithm 8 : Implementation Steps for Replacement Policy

Step 1: Set $p = 0$. The maintenance decision at this point (time zero) is do nothing.

Step 2: Set $p = p + 1$. At the pth condition monitoring observation point (pth maintenance decision making point), collect the condition monitoring indicator value and determine the condition of the device (ϑ^{t_p}).

Step 3: Calculate $CI(\vartheta^{t_p})$ from Eq. (6.13). If $CI \geq g(\gamma^*)$, then replace the device immediately, otherwise do nothing, that is, wait until the next condition monitoring point and move back to step 2. Also, replace the device when it fails.

6.3.5 Alternative Cost Function

In the replacement policy discussed in Section 6.3.4, the objective was to minimize the expected long-run average cost per unit of time of a replacement cycle. Sometimes, the maintenance decision makers are interested in minimizing the long-run average cost of an operation cycle. In this case, the cost function based on policy γ is as follows:

$$g(\gamma) = \frac{c_r + c_f \times E(\Pr(L \leq T_\gamma))}{E(\min(L, T_\gamma))}. \tag{6.15}$$

Here, the cost of failure includes additional failure cost and downtime cost. The new replacement policy can be obtained by (1) modifying the new cost of failure replacement so that it also includes downtime cost and (2) set t_r, t_f, and c_d, to zero. This type of cost structure has been used in some articles, such as [41, 147].

6.3.6 Inspection Data with Missing Points

As mentioned earlier, the results of this chapter are also applicable for devices under non-periodic observation intervals, where the intervals between observation points are not equal. Therefore, condition monitoring data with missing points can be considered a special case of a non-periodic maintenance decision making.

6.3.7 Reducing the Effect of Outliers

Condition monitoring data in real-world systems are usually subject to noise due to errors of measurement or the nature of the condition monitoring data. Such noisy data may result in wrong maintenance decisions. For example, a single noisy observation at a condition monitoring point may reflect that the device is very likely to be in a highly-damaged health state, while it is actually still in a normal state. This kind of outlier may result in a non-monotonic hazard function over time that can affect the calculation of the control-limit policy. For example, a noisy data may result in a high value of $CI(\vartheta^t)$ and therefore unnecessary maintenance action. One possible approach to deal with noisy data is to check the control-limit more than once. For example, the replacement policy can be changed to a n-points decision making rule as follows:

$$\begin{cases} CI(\vartheta^{t-i\times\Delta}) \geq g(\gamma), 0 \leq i \leq (n-1) & \rightarrow D(\vartheta^t,\gamma) = 1 \\ \Psi^t = 0 & \rightarrow D(\vartheta^t,\gamma) = 1 \\ Otherwise, & \rightarrow D(\vartheta^t,\gamma) = 0 \end{cases} . \tag{6.16}$$

In the above equation, the control-limit is checked at n consecutive points and the device is replaced only if the control index of all n points is higher than the control-limit. It is obvious that applying the above policy is subject to the risk of missing a failure when the degradation process is very close to the end of the life of the device and is evolving very fast. In Appendix A, an alternative approach to deal with non-monotonic hazard function is discussed.

6.3.8 Time-Based Preventive Replacement Policy

One of the most commonly used replacement policies in the literature and real-world problems is the time-based (age-based) preventive maintenance (TBM) policy under which the device is replaced at a priori-known time T_ϱ or at a failure, whichever occurs first. This constant replacement T_ϱ is determined regardless of the actual degradation process associated with a device. The optimal replacement time is the time for which the following long-run average cost per unit time is minimized:

$$T_\varrho^* = \min \left\{ L, \arg\min_{T_\varrho} \frac{c_r + c_d \times t_r + (c_f + c_d \times t_f) \times (1 - R(T_\varrho|\vartheta^0))}{\int_0^{T_\varrho} R(x|\vartheta^0)dx + t_r + t_f \times (1 - R(T_\varrho|\vartheta^0))} \right\}, \tag{6.17}$$

where $R(x|\vartheta^0)$ is the unconditional reliability function at time x (dependent only on the initial condition of the device, which is being in the working perfectly state). The solution for the optimal replacement time can be found by direct optimization or search methods. This type of replacement policy is useful to be compared with condition-based replacement policy in order to find out whether or not the condition-based replacement policy is worth implementing.

6.4 Numerical Example

In this section, simulation-based numerical experiments are used to demonstrate (1) how the proposed condition-based replacement model can be used for maintenance decision making and (2) how effective the proposed condition-based replacement model is with respect to the option of Do Nothing. The structure of the numerical example is the same as the one used in Chapter 5. A device with 5 levels of health states with a degradation process evolving according to a NHCTHSMP is considered. The cost of replacement (c_r) is assumed to be \$100,000 and the time required for replacement (t_r) is assumed to be 1 cycle. There is also an additional cost of failure (c_f), which is equal to \$300,000. Also, there is a downtime cost for each unit of downtime (c_d), which is \$30,000. The additional time required for failure replacement (t_f) is assumed to be 2 cycles. Later, in this section, by varying c_f/c_r, the effect of the cost of failure on the condition-based replacement model is demonstrated.

In order to be able to apply the condition-based replacement policy, the structure of the control-limit index and the optimal cost of replacement policy need to be determined (Training Step). First, the steps listed in Section 6.3.4 are implemented to find γ^* and its associated cost $g(\gamma^*)$. Then, the results of applying optimal replacement policy on 100 random trajectories of independent run-to-failure data are presented. Each device is replaced at failure or a preventive replacement time determined by the replacement policy, whichever occurs first. To find out how effective the CBM policy is, its associated cost is compared with the case that no replacement policy exists, that is the device is always replaced at failure. It should be noted that in theory, the unit cost of the system considering no maintenance should be $\dfrac{c_r + c_f + c_d \times (t_r + t_f)}{E(L) + t_r + t_f}$, where $E(L)$ is the expected operation time to failure for the device.

Table 6.1 presents the results for the training phase of the condition-based maintenance policy, at which the structure of the condition-based replacement policy is determined. As shown in Table 6.1, the investigation for the optimal replacement policy is started from an arbitrary policy (γ_0), then steps summarized in Section 6.3.4 are iteratively applied until the replacement policy converges to the optimal replacement policy. As noted earlier, a good choice for initial replacement policy is to consider the do nothing (DN) policy, which has a unit cost of 8213.6. Results show that the optimal replacement policy has an average unit cost of (\$2921.6). In addition, applying the proposed condition-based replacement policy will result in an average replacement cycle of 48.7.

Table 6.1: Search for the optimal replacement policy

Policy No. (k)	g(γ_k)	$E(\min(L < T_{\gamma_k}))$	$E(\Pr(L < T_{\gamma_k}))$	$g(\gamma_{k+1})$
0	8213.6	51.06	8.1%	3047.7
1	3047.7	48.77	4.5%	2932
2	2932	48.72	4.3%	2920.9
3	2920.9	48.7	4.3%	2921.6
4	2921.6	48.7	4.3%	2921.6

Therefore, the following control-limit based replacement policy can be used for replacement decision making at any time point t.

$$D(\vartheta^t, \gamma^*) = \begin{cases} 1, & \text{if } CI(\vartheta^t) \geq 2921.6 \\ 0, & \text{if } CI(\vartheta^t) < 2921.6 \\ 1, & \text{if } \Psi^t = 0 \end{cases} . \tag{6.18}$$

Now, that the structure of the replacement policy is determined, it can be used for maintenance decision making. The result of applying the proposed replacement policy on 100 random trajectories of independent run-to-failure data is shown in Figure 6.1. For each sample, the actual failure time and the proposed replacement time based on the optimal replacement policy are shown. It can be observed from Figure 6.1 that for 97 samples, the device is replaced at a replacement time suggested by the optimal replacement policy (replacement time $<$ failure time) and for 3 samples the device is replaced at an unexpected failure. Finally, in Table 6.2, the result of applying the condition-based replacement model (CBM) is compared with the case with no maintenance (DN). It can be observed from this table that applying the proposed condition-based replacement model results in cost reduction of 65.64%.

Table 6.2: Effectiveness of CBM

Type	Average Unit Cost	Average Replacement Time
DN	8168	56.99
CBM	2806.3	49.11
Cost Reduction (%)	65.64%	

To demonstrate how the control-limit policy can be actually used for maintenance decision making, two samples are selected in the sense that for the first one (Figure 6.2.a), the replacement policy suggested to replace a degraded device before

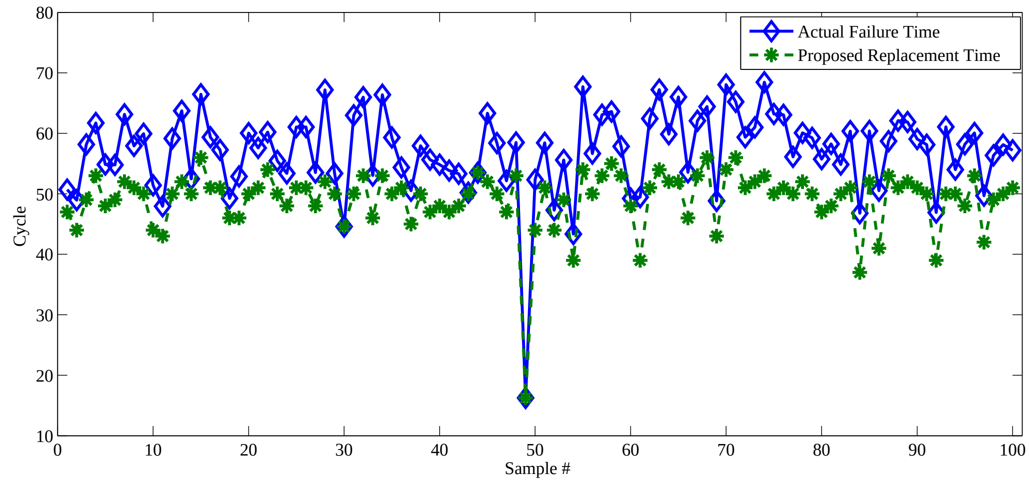

Figure 6.1: CBM replacement time - actual failure time

the actual failure time and for the second one (Figure 6.2.b) the actual failure occurred before the replacement policy could prevent it. Also, it can be observed that for the second case, the control index was very close to the threshold at the failure point. This observation (inability of the CBM model to predict failure) is very common as the optimal replacement policy cannot always predict failures before they occur. The key point here is that the maintenance policy can minimize the average unit cost over a long period of time.

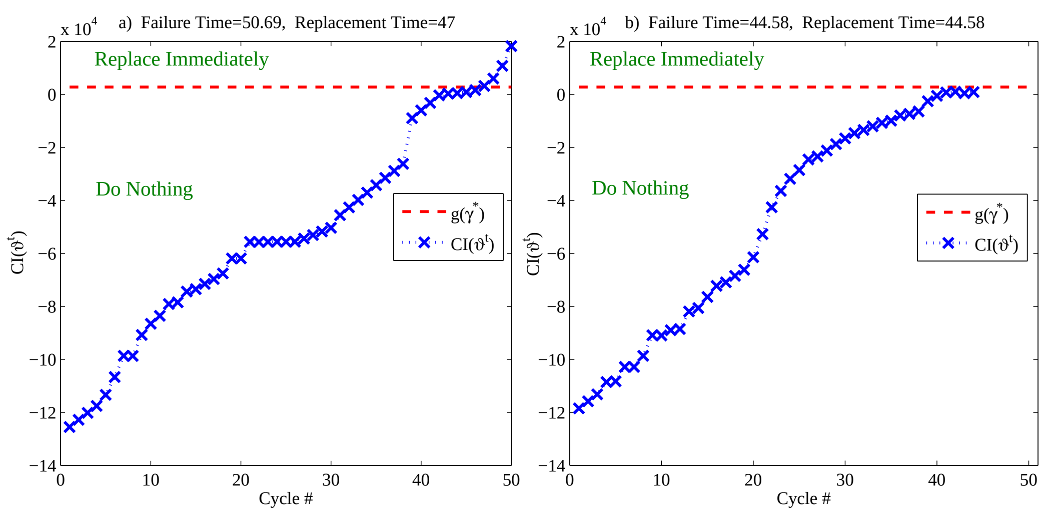

Figure 6.2: Implementation of replacement policy on two samples

To analyze the effect of cost of failure on the results of the maintenance policy, the replacement policy was applied for several combinations of c_f/c_r (other elements are zero). Results are shown in Table. 6.3. For each case of c_f/c_r, the actual costs of the condition-based maintenance policy and no maintenance policy (DN) for 100 samples are shown. It can be verified from the results that CBM results in lower maintenance cost than the no maintenance policy (replace at failure) for all cases. In addition, it can be observed that the condition-based replacement policy is more effective when the cost of failure gets larger, that is, the cost reduction is larger. The final measure investigated here is the average replacement time (ART). It can be observed from this table that the average replacement time of the replacement policy decreases when the failure cost is larger.

Table 6.3: Results of CBM for different combinations of c_f/c_r

c_f/c_r	ART	Cost of CBM	Cost of DN	Cost Reduction (%)
0	57.0	1755	1755	0.0%
0.25	54.2	1927	2193	12.2%
0.5	52.7	1993	2632	24.3%
1	51.1	2074	3509	40.9%
2	49.8	2131	5264	59.5%
3	49.0	2226	7019	68.3%
4	48.4	2316	8774	73.6%
5	48.0	2293	10528	78.2%
6	47.7	2348	12283	80.9%
7	47.3	2262	14038	83.9%
8	47.1	2291	15792	85.5%
9	46.9	2324	17547	86.8%
10	46.7	2356	19302	87.8%

6.5 Summary

In this chapter, a condition-based replacement model is proposed, which can be used as an online tool to determine when to replace a degraded device. The input for this CBM framework is just the condition monitoring data. It turns out that the condition-based maintenance model is from the control-limit family in the sense that a function of the conditional reliability is calculated over time as the control index and then it is compared to a pre-defined warning level (threshold) to check whether or not the device needs to be replaced. The effectiveness of the CBM framework for replacement decision making was demonstrated through simulation-based numerical examples. In Chapter 7, the application of this condition-based replacement policy on maintenance decision making for turbofan engines is demonstrated.

Chapter 7

Case Study

In Chapter 3, the structure of a flexible stochastic process and its application for multistate degradation process are discussed. A parameter estimation method to estimate the characteristic parameters of a known multistate structure is developed in Chapter 4. In Chapter 5, important diagnostic and prognostic measures for a multistate structure under the NHCTHSMP are discussed. Finally in Chapter 6, a condition-based replacement model is introduced, which can employ condition monitoring data for maintenance decision making. Simulation-based numerical experiments are used for illustration purposes at the end of each chapter.

To demonstrate the application of all of the results of this thesis on diagnostic and prognostic health monitoring of mechanical systems, a single case study on turbofan engines extracted from NASA Prognostic Data Repository [90] is considered. This well-known collection of publicly available data sets focuses on prognostic data sets, which can be used for the development of prognostic algorithms. The data sets in this repository have sufficient number of samples, which can be used as training, validation, and testing sets. Most data sets in this repository include temporal sequences of condition monitoring data, which are recorded over time from a normal state until a so-called failure point.

In Section 7.1, the structure of the turbofan degradation data set used in this thesis is described. In Section 7.2, the structure of the multistate structure to be used to model the degradation process and its associated observation process is determined using the results obtained in Chapter 4. In Sections 7.3 and 7.4, important diagnostic and prognostic measures introduced in Chapter 5 are calculated for several engines and their effectiveness is discussed. Finally in Section 7.5, a condition-based replacement policy obtained from the results given in Chapter 6 is developed for replacement decision making of turbofan engines and its effectiveness is evaluated. The advantage of the condition-based maintenance model over the no maintenance case and the conventional age-based replacement model is discussed in Section 7.5.2. Also the advantage of using multistate modeling over binary modeling is discussed in Section 7.5.3. Section 7.5.4 discusses the relative effectiveness of the replacement policy with respect to an ideal replacement policy. The summary of this chapter is given in Section 7.6.

7.1 Data Description

The data set given in [90] consists of multivariate time series signals that are collected from turbofan engine dynamic simulation process. The engine run-to-failure simulation was carried out using C-MAPSS (Commercial Modular Aero-Propulsion System Simulation), a well-known simulation program for transient operation of modern commercial turbofan engines, which allows input variations of health related parameters. This type of engine represents a modern dual-spool, high-bypass ratio turbofan engine that has been the focus of many controls and diagnostics/health management studies over the past few years [152]. A comprehensive logic structure is developed in a manner similar to that used in real engine controllers such that integrator-windup problems are avoided and the controller and regulators perform as intended over the full range of flight conditions and power levels.

Hundred engines' run-to-failure time series trajectories are considered in this thesis (data set FD001), which can be considered to be from a fleet of engines of the same type. In this chapter, this data set is divided to three data sets as follows: (1) 60 engines are considered for training, (2) 20 engines are considered for validation, and (3) 20 engines are considered for testing. Each engine starts with different degrees of initial unknown wear and manufacturing variation, which is considered to be normal, i.e., it is not considered a fault condition. The engine is operating normally at the start of each time series, and develops a fault at some point during the series. The fault grows in magnitude until the system failure. Each record in a run-to-failure trajectories, which corresponds to a given operation cycle, is a 24-element vector, consists of three values for the operational settings and 21 values for engine performance measurements, which are contaminated with noise. All failures are caused by HPC (High-Pressure Compressor) degradation. For more details on the engine run-to-failure simulation and the structure of the data used in this thesis, interested readers may refer to [90]. In Figure 7.1, 4 different sensor measurements for Engine No.1 are presented.

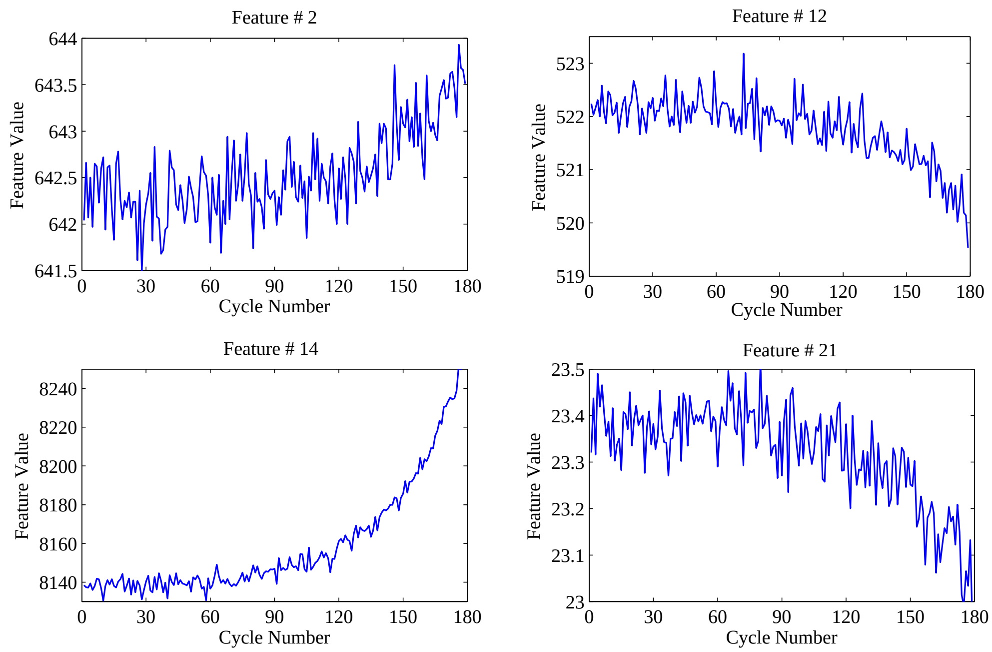

Figure 7.1: Sample condition monitoring feature values for Engine No.1

It can be seen from Figure 7.1 that these features have non-identical but sensitive trends over the age of the device. From the data set, it can also be observed that there are features that vary very little, or fluctuate very much over time, that is, no observable trend over time can be found. In Figure 7.2, 4 different performance measurements of such type for Engine No. 8 are presented. More detailed investigation on the data set also verifies that different degradation patterns result in vastly different lifetimes. In Figure 7.3, the total lifetime for each engine is presented. It can be seen from this figure that lifetime varies from 128 cycles to 362 cycles (mean=206.3 cycles, standard deviation=46.34 cycles, median = 199 cycles). It should be pointed out that in the calculation of likelihood function, important measures, and replacement policy, each three cycles are considered as one time unit for computational convenience. The results are then superimposed to original cycle values. Therefore, the final results presented in this chapter are reported according to the original cycle numbers, but are still subject to the discretization error.

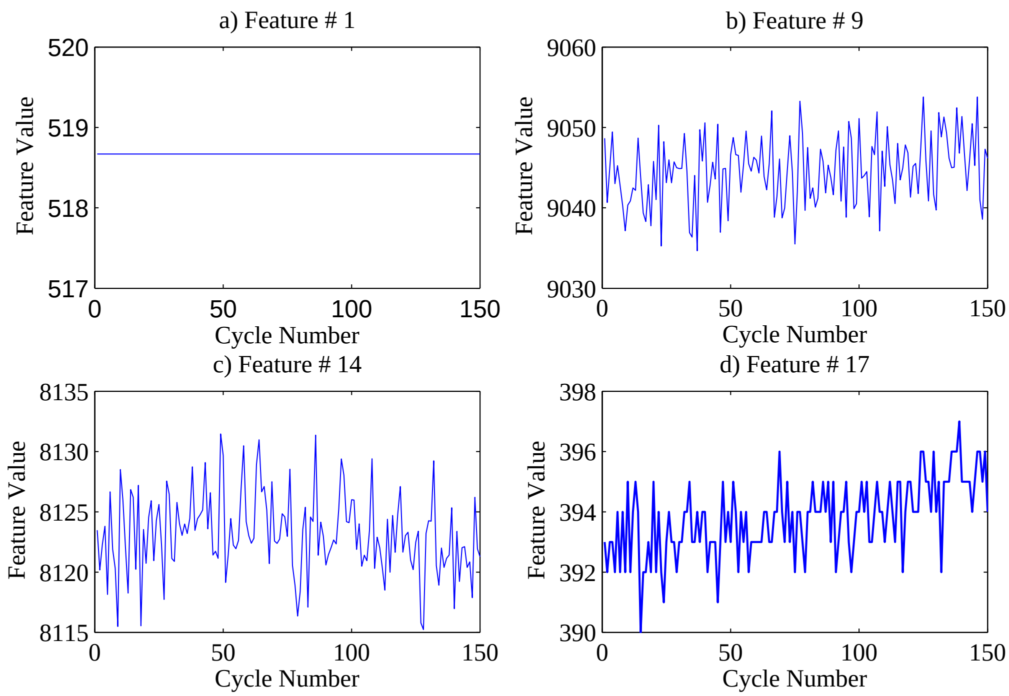

Figure 7.2: Sample condition monitoring feature values for Engine No. 8

7.2 Multistate Structure

As discussed in Chapter 4, the first step of using a multistate degradation process for diagnostics and prognostics is to determine a reasonable structure for the multistate model in the sense that the stochastic behavior of the degradation and observation processes can be represented. Structure determination or alternatively called model selection for any type of device involves two steps, which are (1) configuration (topology) selection and (2) parameter estimation. The purpose of the configuration selection step is to determine a reasonable topology for the associated multistate model with indirectly observable states. The main elements that determine a multistate topology with indirectly observable states are the number of states (N), transition diagram (Ω), transition types (ξ), transition rate's statistical structure (λ), condition monitoring feature (I), and the number of clusters to be used for final feature representation (V). Therefore, the multistate topology can be denoted by $\zeta = \{N, \Omega, \xi, \lambda, I, V\}$. The detailed description of these elements is illustrated in Section 3.3. With regards to the parameter estimation step, parameters to be estimated (θ) are the ones that characterize the degradation process (Γ) and the observation process (B).

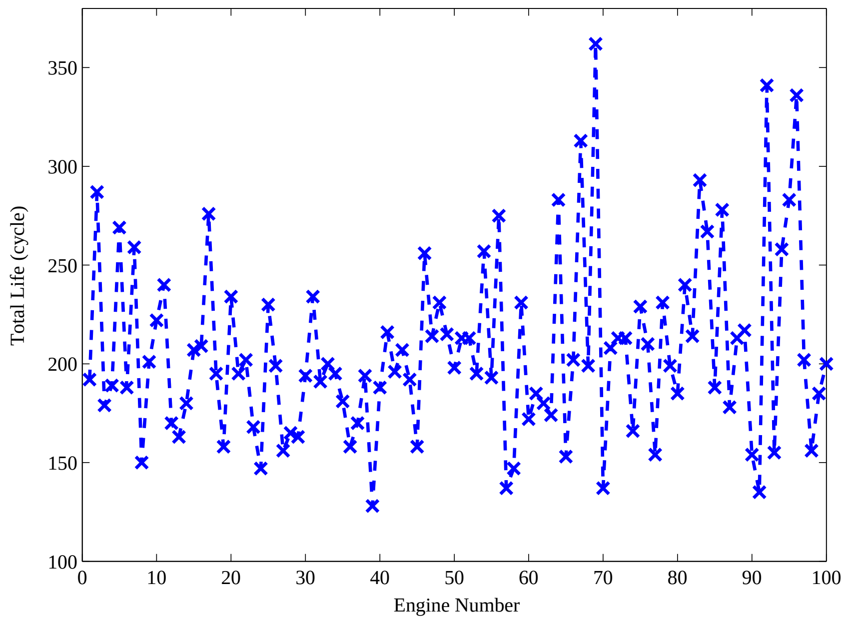

Figure 7.3: Actual life times of 100 engines

In this chapter, the steps defined in Section 4.1 are employed to find the structure of the multistate model. First, in phase I, the elements of $\{\Omega, \xi, \lambda, \mathrm{I}, V\}$ are predetermined. Here, the number of degradation states (N) is considered as the only topology elements that is unknown and needs to be found. Therefore, multistate model alternatives have similar multistate structure elements ($\{\Omega, \xi, \lambda, \mathrm{I}, V\}$) with different number of states. For each model alternative, parameter estimation using training data set is performed to determine the characteristic parameters of the associated structure. Then the model selection criterion (Eq. (4.27)) is used to determine the best structure among possible candidates. The effectiveness of the selected structure is further investigated through diagnostic and prognostic measures and the condition-based replacement model. It should be pointed out that the final multistate structure used in this chapter is not necessarily the best possible multistate structure for the case study. Developing a more effective model selection is an interesting direction for future work. In the following, the steps used in this thesis to determine $\{\Omega, \xi, \lambda, \mathrm{I}, V\}$ are described.

I: Transition Diagram or Connectivity Between States (Ω): This element defines the relationship (connectivity) between degradation states. To cover soft failures (such as a wear process) and hard failures (such as shock process), a progressive one-step left-to-right multistate structure is considered where transitions are allowed either to the neighbor state or to the failure state. It is interesting to note that the results of the parameter estimation show that it is very unlike that the device has a hard failure. Figure 7.4 shows N-states left-to-right progressive multistate structure as an example of a possible structure.

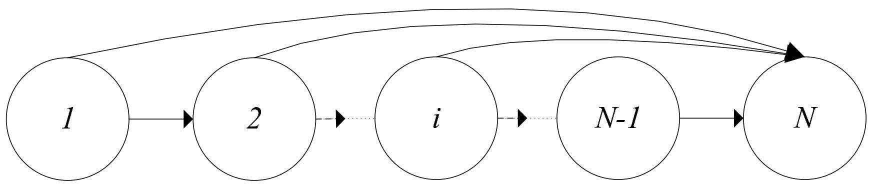

Figure 7.4: N-states left-to-right progressive multistate structure

II. Transition Types (ξ): As stated in Section 1, the nonhomogeneous semi-Markov structure is very flexible to be used for degradation transition modeling as it covers many other structures, such as Markov, homogeneous semi-Markov, and explicit-duration Markov processes. It is assumed that the transitions to the neighbor state depend on the level of degradation and the time spent at each state, while the transitions to the failure state depend on the level of degradation and the total age of the device. The structures of these transition rates are given in Eq. (7.1).

III. Transition Rate Distribution (λ): For the distribution of transition rates, the power law process in the form of Weibull-based hazard function is selected, which is the most widely used distributions for degradation modeling [141]. The final statistical form of transition rate function is as follows:

$$\lambda_{i,j}(s,t) = \begin{cases} \frac{\beta_{i,j}}{\alpha_{i,j}}(\frac{t}{\alpha_{i,j}})^{\beta_{i,j}-1}, & 1 \leq i < N, j = j+1 \\ \frac{\beta_{i,j}}{\alpha_{i,j}}(\frac{s+t}{\alpha_{i,j}})^{\beta_{i,j}-1}, & 1 \leq i < N-1, j = N \end{cases} . \tag{7.1}$$

IV: Condition Monitoring Feature (*I*): In the multistate structure considered in this thesis, the final output of the observation process is assumed to be a single indicator having indirect information regarding the actual health status of the device. Here, the notation I is used to denote the selected feature for health monitoring. For this case study, there are more than one condition monitoring feature available to be used as the representative of the observation process. One possible approach to combine all these features is to use a feature fusion (combination) technique. Among all feature fusion techniques, Principal Component Analysis (PCA) is one of the most widely feature reduction method used in condition monitoring [82, 153, 154, 155]. PCA is known to be efficient in compressing information and eliminating correlations between variables. PCA transforms linearly a set of observations to another set of linearly uncorrelated observations known as the principal component. The first principal component accounts for the largest variability in the data and therefore includes more useful information than other principal components. In this chapter, the first principal component (FPC) of the set of candidate features is used as the representative of the observation process. After observing the trend of the FPC for all engines, it is found that it roughly has a non-decreasing trend over time. Figure 7.5 presents the first principal component values for Engine No. 1 over time.

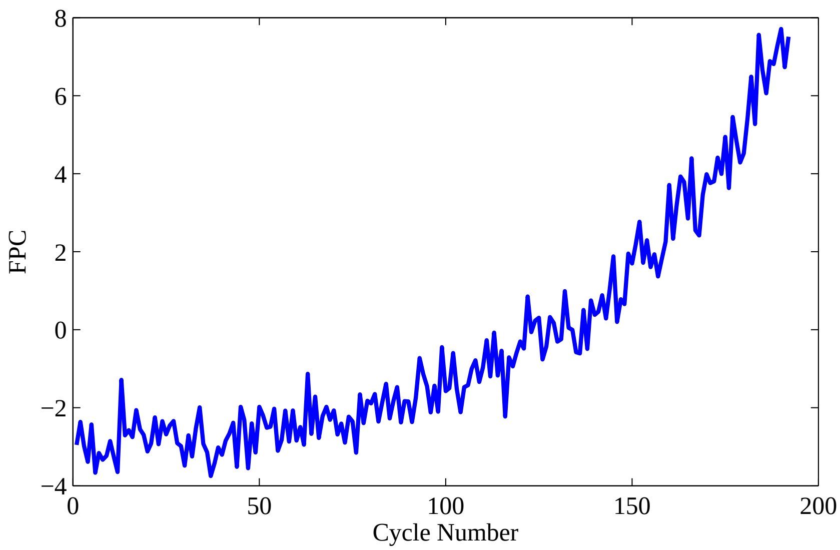

Figure 7.5: FPC for Engine No. 1

V:Feature Dimension (V): As discussed earlier in this thesis, the final CM feature should be represented in a nonparametric discrete form. In other words, the set of original observation values needs to be converted to several discrete levels. Among numerous techniques used in the literature to convert the value of a feature in a continuous domain to a discrete domain without having information about its relationship with the actual state levels, unsupervised k-means clustering has been successfully applied in the domain of condition monitoring [83, 156]. The key challenge of using k-means clustering for dimension reduction is to find the optimal values for k, which represents the number of discrete levels for the CM indicators (m), in the sense that the original stochastic behaviour of the observation process keeps as unchanged as possible.

Different types of criteria are used in the literature for this purpose. The discussion on the best method to be used for clustering is beyond the scope of this thesis. The objective here is to convert original feature values to discrete levels in the sense that the distribution of the feature (stochastic behaviour) does not change very much. Generally, if a very small number of clusters are selected, the feature cannot be a reasonable representative of the degradation process, i.e. it does not vary much with the age of the device. Therefore, it is important for the number of clusters to be reasonably large to better reflect the degradation process. However, considering a very large number of clusters not only may result in misrepresentation of the original data, but also is computationally expensive. As in this case study, the selected feature represents a degradation process, it is reasonable to have a feature with a monotonic trend over the age of the device. To evaluate the variability of a monotonic trend, the well-known nonparametric score called Mann-Kendall [157] can be modified in the sense that weights are considered according to the distance between two points. Now, for the sequence of condition monitoring indicators extracted from time zero to the failure time, this index can be calculated as:

$$M - K_m = \sum_{k=1}^{K} \sum_{i=1}^{d_k} \sum_{j=1, j>i}^{d_k} (t_j^{(k)} - t_i^{(k)}) \times sgn(O_j^{(k)} - O_i^{(k)}), \tag{7.2}$$

where $sgn(x) = \begin{cases} -1 & \text{if}\, x < 0 \\ 0 & \text{if}\, x = 0 \\ 1 & \text{if}\, x > 0 \end{cases}$, m is the number of clusters, K is the number of samples, $t_i^{(k)}$ is the time of the ith observation point for the kth sample and $O_i^{(k)}$ is the output of the observation process at time $t_i^{(k)}$ for the kth sample. Now, the best value for m is selected in the sense that increasing it does not change very much the monotonicity index given in Eq. (7.2).

Figure 7.6 presents the change in the monotonicity index by increasing the number of clusters. It can be seen that monotonicity stays almost unchanged for $m \geq 22$. Therefore, the number of possible CM outcomes is selected as 22 (m=22). Figure 7.7 presents the final CM outputs for Engine No. 1 when m=22. These superimposed discretized values are considered as the input for parameter estimation. It should be pointed out here there is no guarantee that this index can find the optimal number of possible CM outputs.

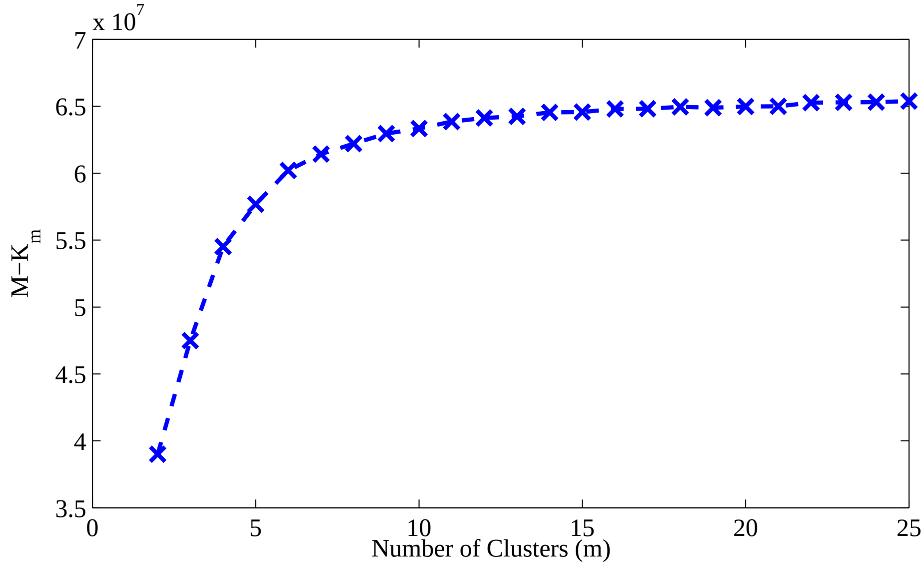

Figure 7.6: Change in the monotonicity index versus the number of clusters

7.2.1 Final Multistate Model Using BIC

Now that the elements of $\{\Omega, \xi, \lambda, \mathrm{I}, V\}$ are defined, possible model alternatives can be considered as structures with exactly same elements of $\{\Omega, \xi, \lambda, \mathrm{I}, V\}$ and different values of N (number of states). To determine the value of N (number of states), as shown in Figure 4.1 for model selection, first the parameter estimation method (see Chapter 4) is used to train all model alternatives considering different values for N. Then, the BIC measure introduced in Eq. (4.27) is calculated for each structure candidate to determine the reasonable values of N. Figure 7.8 presents the BIC for different model alternatives with different values of N. It should be pointed out that to avoid over-fitting, an independent validation data set composed of 20 engines (engines No. 61-80) is used. It can be observed from this figure that N=7 has the highest value among other alternatives. It should be noted that as the decreasing trend for BIC is observed for $N > 7$, the training procedure is performed only up to N=10. It is important to note that other methods can also be used to compare different model alternatives. Finding the best possible structure among several candidates is out of the scope of this work. Later in this chapter, the advantage of this structure over other candidate structures are also discussed through the cost of the corresponding maintenance models. Now that the

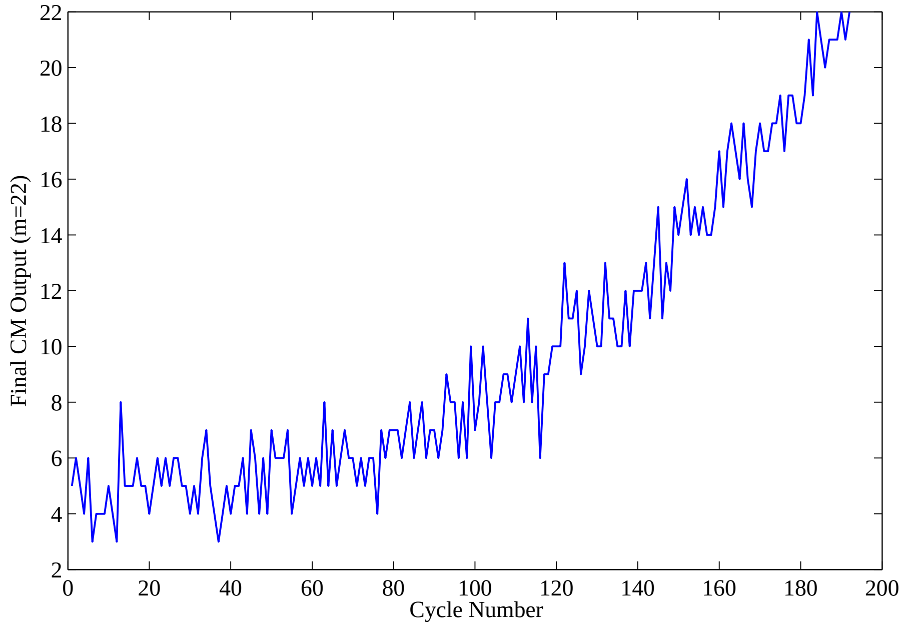

Figure 7.7: First principal component for Engine No. 1 and the discretized values

structure of the multistate model is known, it can be used as a tool for diagnostic and prognostic health monitoring as well as maintenance decision making.

7.3 Health Monitoring Using Diagnostic Measures

In this section, the average degradation level (ADL) is selected as a diagnostic measure for health monitoring of the 20 Engines (Engines No. 91-100) in the testing data set. Results shown in Figure 7.9 verify that, overall, this measure roughly has a non-decreasing trend over time. This condition monitoring measure, which is sensitive to the level of damage over time, can directly reflect the level of degradation and therefore can be used as an online tool for degradation monitoring. The results shown in this figure for all 20 engines indirectly supports that the selected multistate structure is a reasonable representative of the degradation process. In addition, as shown in Figure 7.9, this degradation measures can represent different degradation patterns, from rapid degradation as occurred for Engine No. 93 as well as slow degradation as occurred for Engine No. 92.

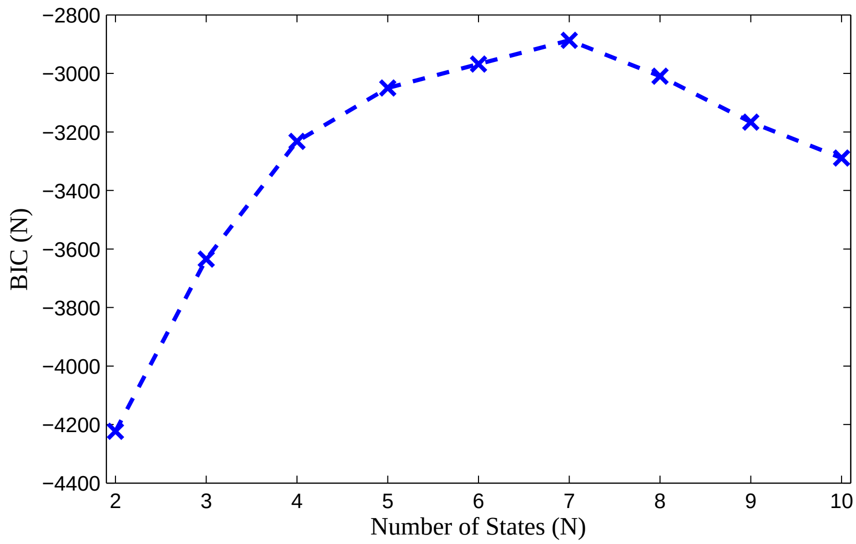

Figure 7.8: Bayesian Information Criterion (BIC) - the number of states (N)

7.4 Prognostic Measures

The measure of interest calculated in this section for prognostic engine monitoring is the mean of the remaining life and its associated percentile confidence interval (95%). For Engines No. 97-100 in the testing data set, the MRL with the actual remaining life and the 95% percentile confidence interval are shown in Figure 7.10 (estimation starts at cycle 1). It can be seen from this figure that overall the estimated mean residual life is far from the actual residual life, but it gets closer to the actual remaining life particularly for larger ages of the device (close to the failure point). This makes sense as over time, more condition monitoring data are employed for estimation. Also, as the devices ages, the percentile interval for the remaining life becomes narrower, that is, the prediction uncertainty decreases. Such figure can be used as an online tool for estimating the remaining useful life.

In Tables 7.1 and 7.2, the estimated mean failure time (MFT) and the 95% percentile confidence interval for the total life ([LB,UB]) are calculated for Engines No. 81-100 at different points in terms of the percentage of the lifetime. As can be seen in these tables, estimation points are at 70%, 75%, 80%, 85%, 90%, and 95% of the lifetime of each engine. It should be pointed out that the estimated total life is the estimated residual life plus the time point at which the estimation is made. Also, for each engine the actual failure time (AFT) is reported. The following can be observed from Tables 7.1 and 7.2.

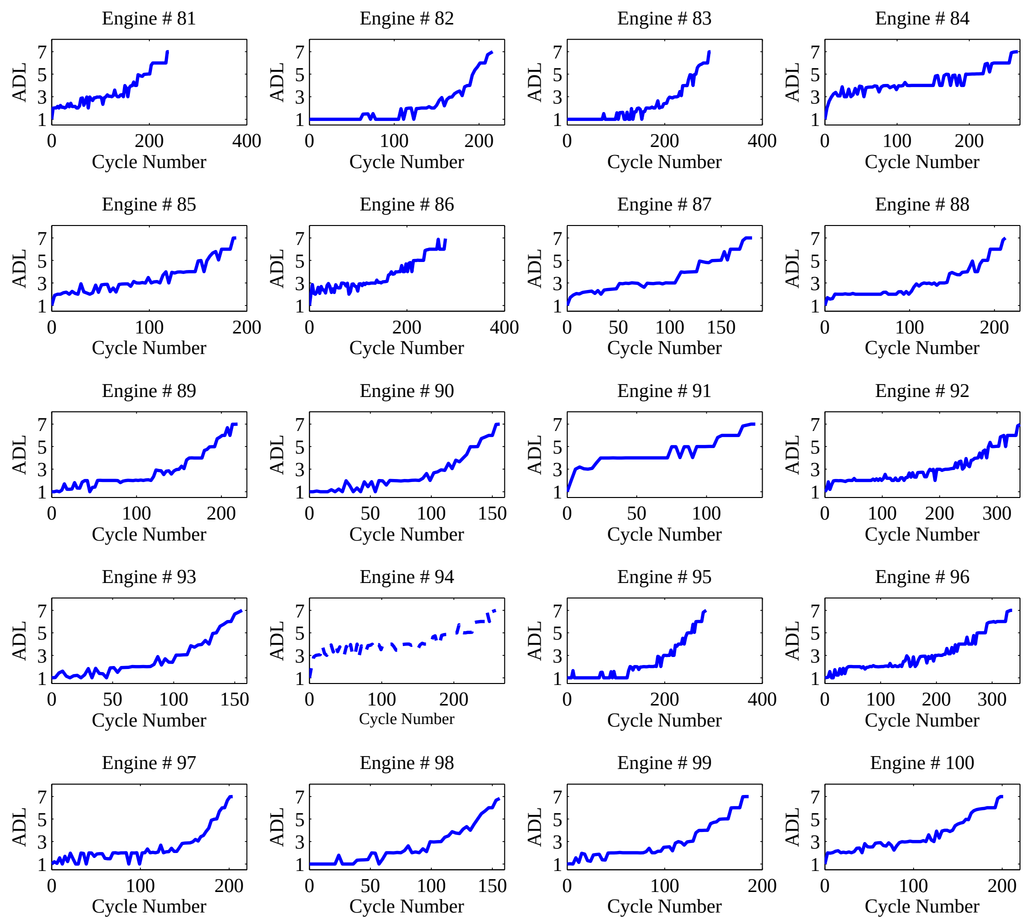

Figure 7.9: Average degradation level (ADL) for condition monitoring

- Overall, the mean failure time provides reasonable estimates for the actual failure time, particularly for larger ages of the engines.
- Estimated mean failure times get closer to the actual failure time as the device ages.
- The provided percentile intervals cover the actual failure times in almost all cases, particularly at larger ages.
- The percentile intervals are getting narrower as the device ages.
- In more cases, the estimated failure time is lower than the actual failure time. This is generally better than the case in which the estimated failure time is higher than the actual failure time. The reason is that overestimating the failure time may result in unexpected (or catastrophic) failures.

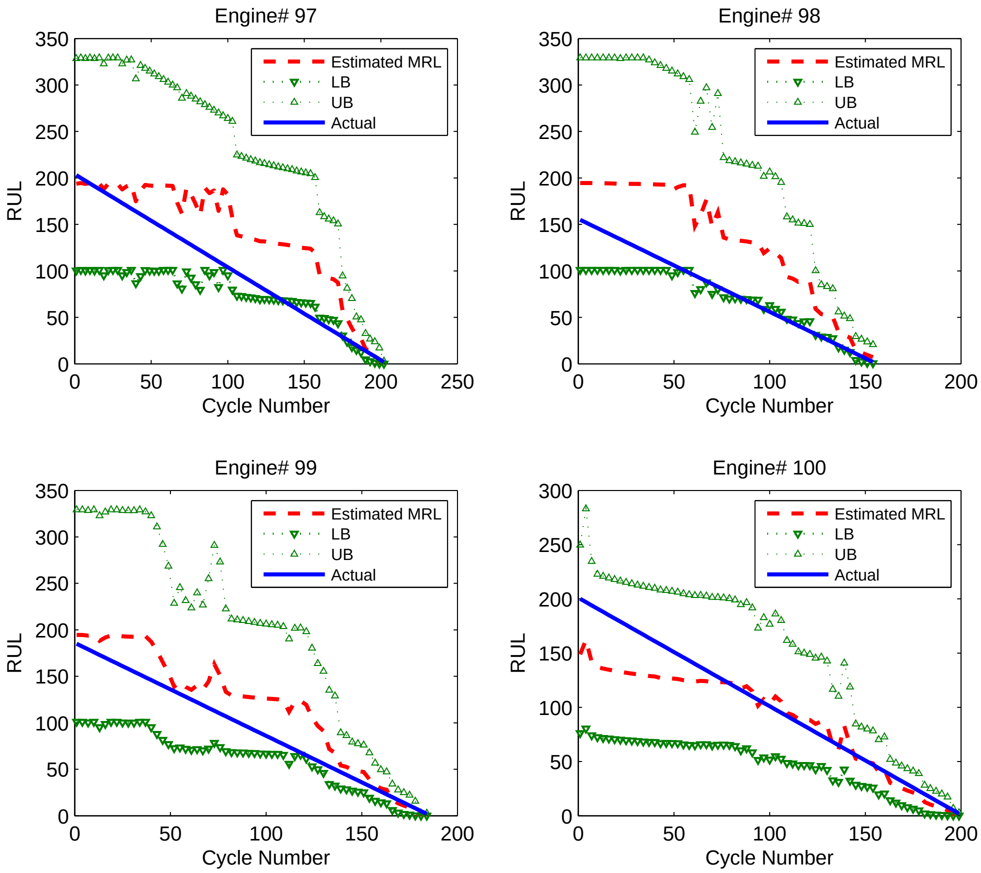

Figure 7.10: Remaining useful life estimation for Engines No. 97-100

7.5 Condition-based Maintenance

In this section, the condition-based replacement policy introduced in Chapter 6 is applied on the turbofan engine testing data set to find out how effective it is to find out when to replace a degraded engine. This section is composed of four subsections. First, the structure of the control-limit replacement policy is determined. Then, how the proposed condition-based replacement model can be used for replacement decision making for Engines No. 81-100 is shown. The effectiveness of the maintenance model is evaluated through comparison with the Do-Nothing policy (DN) and aged-based replacement policy (TBM) under different levels of c_f/c_r. The effectiveness of the multistate structure compared to the binary case is also demonstrated. Finally, comments on the possible improvement of the condition-based model and its effectiveness compared to an ideal replacement policy are provided.

7.5.1 Development of the Replacement Policy

In order to be able to apply the condition-based replacement policy, the structure of the control-limit index and the optimal cost of replacement policy need to be determined (training step). First, the steps listed in Section 6.3.4 are implemented to find γ^* and its associated cost $g(\gamma^*)$. Then, the results of applying the replacement policy on 20 engines in the testing data set are presented. Each device is replaced at failure or at a preventive replacement time determined by the replacement policy, whichever occurs first. To find out how effective the CBM policy is, its associated cost is compared with the case that no replacement policy exists, that is the device is always replaced at failure. The cost of replacement (c_r) is assumed to be \$100,000 and the additional cost of failure replacement is assumed to be \$300,000. All other elements are considered to be zero.

Table 7.3 presents the result for the training phase of the condition-based maintenance policy, at which the structure of the condition-based replacement policy is determined. As shown in Table 7.3, the investigation for the optimal replacement policy is started from an arbitrary policy (γ_0), then steps summarized in Section 6.3.4 are iteratively applied until the replacement policy converges to the optimal replacement policy. As noted earlier, a good choice for initial replacement policy is to consider the do nothing policy, which has a cost of $g(\gamma_0) = \$2039$. Also, the optimal replacement policy has an average cost of \$523.2. In other words, in theory applying the proposed condition-based replacement policy will result in a long-run average unit cost of \$523.2. The expected operation time in each cycle based on this policy is 191.13. Results also verify that the probability of failure replacement is theoretically zero.

Now that the structure of the replacement policy is determined, it can be used as an online tool to determine the replacement time of degraded engines. The engine is replaced at the suggested replacement time or the failure, whichever occurs first. Figure 7.11 shows how the developed replacement policy can be used for engines No. 81 and 83. Whenever the condition-based control index exceeds the limit, the device needs to be replaced. It can be seen from this figure that the suggested replacement time is lower than the actual failure time for both engines.

To evaluate the effectiveness of the replacement model, the associated cost of implementing it on the 20 engines in the testing data set is presented. Figure 7.12

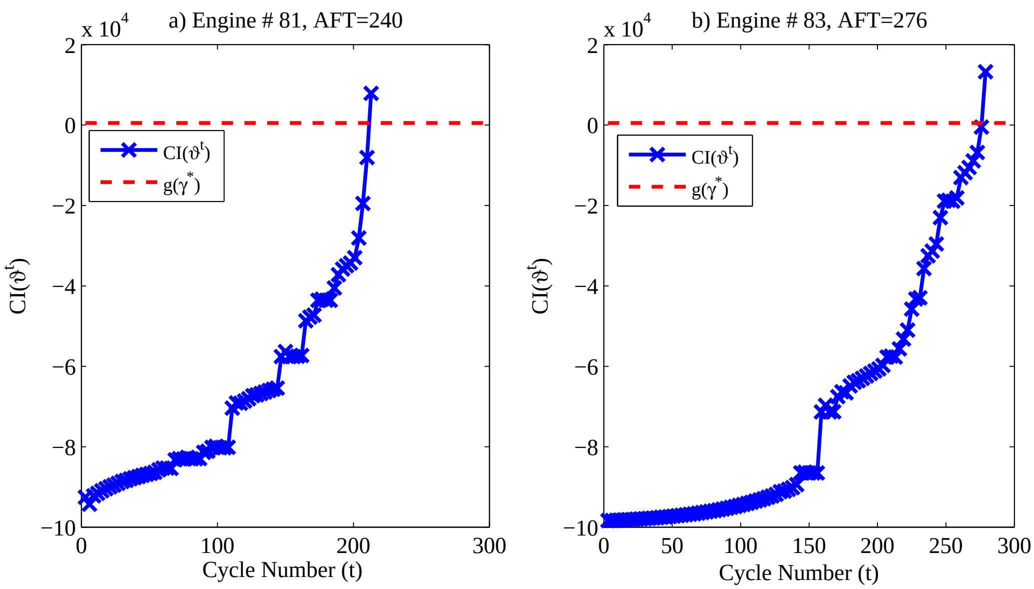

Figure 7.11: Two examples of the control-limit policy

shows the effective replacement time versus the actual failure time for engines No. 81-100. Results in this figure verifies that (1) the replacement times are lower than the failure time (that is all replacements are preventive replacement) and (2) the replacement times are reasonably close to the actual failure times.

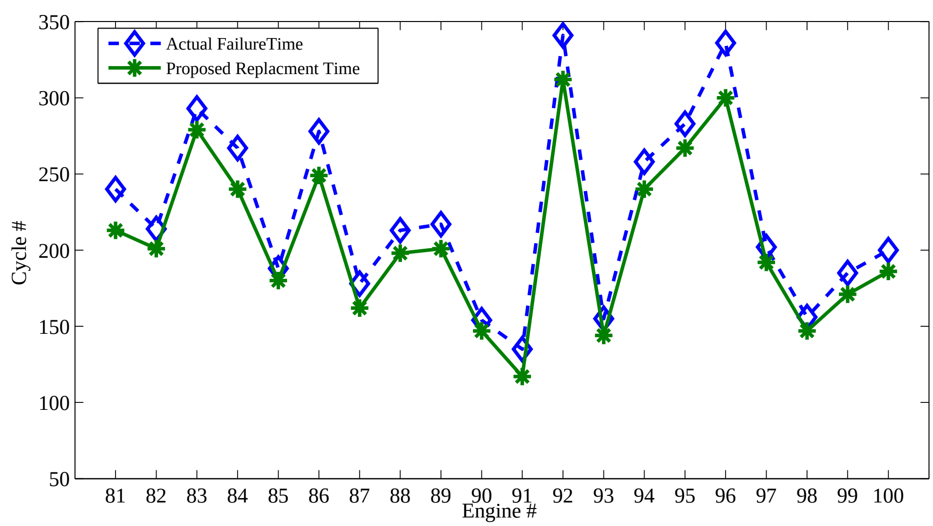

Figure 7.12: Suggested replacement time - actual failure time

To analyze the effect of cost of failure on the results of the maintenance policy, the replacement policy was applied considering several combinations of c_f/c_r. Results are shown in Table. 7.4. For each case of c_f/c_r, the actual costs of the condition-based maintenance policy (CBM) and no maintenance policy (DN) and the average replacement time (ART) for the 20 engines in the testing data set are shown. It can be verified from the results in this table that CBM results in lower maintenance cost than the no maintenance policy (replace at failure) for all cases except for the case where the cost of failure replacement is zero. In addition, it can be observed that the condition-based replacement policy is more effective when the cost of failure gets larger, that is, the cost reduction is larger. Also, on average, when the cost of failure is larger, the device is replaced at an earlier point.

7.5.2 Comparison between Condition-based Replacement Model and Time-Based Replacement Model

As discussed in Section 6.3.8, one of the most commonly used replacement policies in the literature and real-world problems is the time-based preventive maintenance policy under which the device is replaced at a known time T_ϱ or at a failure, whichever occurs first. This constant time T_ϱ is determined regardless of the actual degradation process associated with a device. The advantage of using a time-based replacement policy (TBM) is that it does not need any information on the actual degradation process, therefore, no need for the condition monitoring framework. A reasonable way to demonstrate the effectiveness of a condition monitoring system is to compare its cost saving with the one from an age-based replacement policy. If the cost saving is high enough so that it can cover the cost of implementing a condition-monitoring system, then it can be considered as an effective tool for replacement decision making.

In Table 7.5, the average replacement times as well as the average unit cost for both CBM and TBM considering Engines No. 81-100 are shown. Results in this table verify that (1) the cost of condition-based maintenance is lower than the cost of time-based maintenance, (2) the condition-based maintenance can result in higher cost reduction for larger failure costs, and (3) the time-based replacement times (ART_2) are earlier (in almost all cases) than the condition-based replacement times (ART_1), that is TBM is more conservative than CBM and can result in more unnecessary (early) maintenance actions.

7.5.3 Comparison between Multistate Structures and Binary structures

In this subsection, the effectiveness of the multistate structure compared to the binary structure is evaluated. In Table 7.6, the average replacement time for the CBM and TBM, the unit cost of CBM, TBM, and Do-Nothing, and the cost reduction of using CBM compared to (1) TBM and (2)do-nothing for multistate structure with different levels of N are shown. ART_1 and ART_2 are respectively the average operation times associated with the CBM and TBM. To obtain more consistent results, the experiment is repeated for 10 different combinations of c_f/c_r $(1 \leq c_f/c_r \leq 10)$. Results in Table 7.6 are the average of these 10 cases. From the result in this table, it can be verified that (1) CBM and TBM with multistate structures have lower cost than the binary structure, (2) the cost of the CBM model with seven states (N=7) is the lowest among all other multistate structures, (3) condition-based maintenance performs better than the aged-based maintenance and do-nothing maintenance strategy, (4) the average replacement time of the multistate structure is higher than the one in TBM, and (5) the average replacement times in multistate structures are higher than binary structure. These results can justify the use of multistate structure for degradation modeling.

7.5.4 Comparison between CBM and an Ideal Replacement Policy (IRP)

From the cost point of view, an ideal block replacement policy is a policy under which the device is replaced at the observation point right before the actual failure time. Based on such a policy, not only unnecessary maintenance actions are avoided, but also catastrophic and unexpected failure are prevented. It should be pointed out that such a policy is only ideal when the maintenance setup time is negligible. Although achieving such a replacement policy may be impossible in real-world problems, calculating the cost of such a policy can provide maintenance decision making another tool to evaluate the effectiveness of a condition-based replacement policy. Figure 7.13 shows the relative difference (%) between the cost of the condition-based replacement policy (CBRP) and an ideal replacement policy (IRP) applied on Engines No. 81-100. The results shown is based on the average of 10 levels of c_f/c_r $(1 \leq c_f/c_r \leq 10)$. It can be observed from this figure that (1) overall, multistate structures can result in lower maintenance cost compared to a binary structure and

(2) the difference between the cost of the proposed replacement policy and an ideal replacement policy is reasonably low for $N > 5$ (less than 10%) and approximately 7.7 % for $N = 7$. This 7.7 % is the maximum possible cost improvement from any replacement policy compared to the one proposed in this thesis.

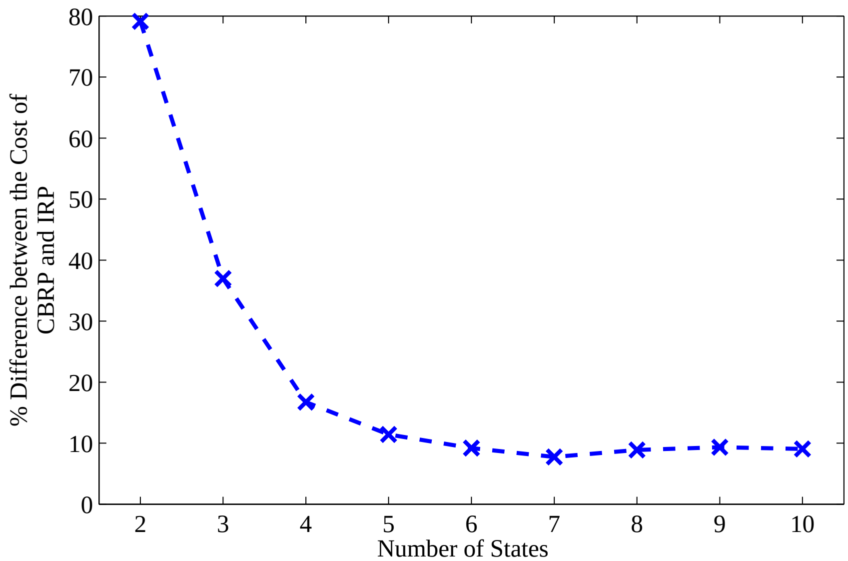

Figure 7.13: Comparison between CBRP and an IRP

7.6 Summary

In this chapter, the application of the proposed multistate modeling on diagnostic and prognostic health monitoring and maintenance decision making is shown through a case study on turbofan engine data set from the NASA prognostic Data Repository. The initial elements of a multistate structure are defined and then different model alternatives are compared with a model selection criterion to find the optimal structure of the multistate process. Diagnostic and prognostic measures are then calculated for online health monitoring of the device. Finally, a condition-based replacement model is developed, which can be used to determine when to replace a degraded engine. Results obtained in this chapter can verify that using a multistate degradation process for modeling the actual degradation process can result in significant cost reduction and can prevent catastrophic and costly failures. Also, the relative effectiveness of the CBM compared to the TBM and the relative effectiveness of using multistate structure compared to the binary structure are shown. The results applied on condition monitoring of turbofan engines can verify that multistate structures are very effective in modeling the degradation process and developing online diagnostic and prognostic health monitoring of systems under gradual degradation and condition monitoring.

Table 7.1: Prognostic results for engines No. 81-90

Engine #	Measure	Estimation Point (in terms of %) Lifetime						AFT
		70%	75%	80%	85%	90%	95%	
81	LB	197.4	202.6	207.8	214.0	217.0	228.4	240
	MFT	222.2	225.5	224.5	227.7	224.3	231.7	
	UB	255.4	256.0	247.7	246.5	237.5	241.5	
82	LB	220.8	230.2	222.3	227.1	211.0	207.1	214
	MFT	282.8	290.3	266.8	270.4	227.3	218.8	
	UB	366.0	366.0	330.8	333.1	248.7	238.1	
83	LB	274.1	279.9	281.9	276.4	280.3	282.0	293
	MFT	327.8	330.7	322.6	299.5	296.2	292.3	
	UB	366.0	366.0	366.0	330.6	317.0	307.5	
84	LB	208.0	215.7	224.6	237.2	241.4	255.4	267
	MFT	226.8	230.6	237.3	249.2	249.7	258.5	
	UB	250.8	250.7	254.9	265.4	263.5	268.0	
85	LB	169.0	170.4	171.6	177.8	183.9	181.8	188
	MFT	208.0	194.7	193.2	193.9	197.7	190.9	
	UB	266.0	228.0	223.6	214.9	216.6	204.9	
86	LB	222.2	235.5	236.0	245.9	253.0	264.3	278
	MFT	243.8	255.7	250.7	258.6	260.4	266.9	
	UB	272.9	282.5	270.8	276.2	274.2	277.1	
87	LB	154.8	158.9	159.8	163.8	165.3	171.6	178
	MFT	180.7	181.4	176.6	178.0	175.2	177.4	
	UB	219.4	211.3	200.0	197.4	189.6	190.3	
88	LB	191.6	204.5	199.3	207.3	206.4	203.0	213
	MFT	235.6	246.6	222.9	229.0	221.1	212.3	
	UB	299.4	307.3	254.5	258.3	241.1	226.7	
89	LB	202.2	195.2	202.5	205.2	207.7	207.9	217
	MFT	248.0	220.8	229.4	225.0	222.4	216.1	
	UB	313.5	257.1	274.3	252.9	242.3	230.3	
90	LB	172.5	166.6	160.8	150.0	153.0	150.7	154
	MFT	234.7	212.4	199.5	166.8	167.6	161.2	
	UB	318.5	278.2	262.8	189.7	187.5	176.1	

Table 7.2: Prognostic results for engines No. 91-100

Engine#	Measure	Estimation Point (in terms of %) Lifetime						AFT
		70%	75%	80%	85%	90%	95%	
91	LB	111.1	115.5	118.1	116.6	123.7	129.3	135
	MFT	127.2	129.8	131.8	126.3	129.7	132.0	
	UB	148.2	149.3	150.5	140.7	142.1	142.7	
92	LB	286.1	301.9	301.3	306.0	315.7	325.8	341
	MFT	324.8	337.1	324.0	321.0	328.7	334.9	
	UB	366.0	366.0	354.7	341.2	346.5	348.9	
93	LB	157.2	163.0	155.3	154.4	152.5	149.4	155
	MFT	202.1	206.5	180.5	177.3	167.8	159.1	
	UB	266.1	268.8	214.8	209.1	188.2	173.6	
94	LB	206.1	213.2	218.8	230.3	236.8	246.5	258
	MFT	226.8	231.5	233.9	243.1	249.1	251.8	
	UB	254.5	256.4	254.6	260.8	266.2	264.9	
95	LB	256.3	260.8	261.9	269.3	269.1	271.8	283
	MFT	310.0	304.0	297.0	291.3	283.4	280.9	
	UB	366.0	366.0	356.7	320.8	302.8	294.9	
96	LB	276.3	280.5	295.7	298.0	305.0	318.7	336
	MFT	315.5	305.7	316.1	312.1	315.0	325.1	
	UB	366.0	344.8	343.2	331.8	330.2	338.4	
97	LB	211.6	218.8	211.7	217.6	201.0	197.7	202
	MFT	271.1	277.4	258.2	261.3	221.5	207.5	
	UB	353.0	358.6	324.7	324.5	253.4	221.8	
98	LB	164.0	160.9	157.0	161.0	156.1	151.1	156
	MFT	222.0	205.1	185.1	184.2	171.8	161.7	
	UB	303.3	268.5	226.3	215.2	192.5	176.5	
99	LB	179.0	170.3	175.9	175.1	174.1	178.2	185
	MFT	225.7	195.2	198.0	192.1	185.4	185.9	
	UB	292.8	230.4	227.6	215.7	202.2	199.3	
100	LB	183.5	177.5	182.5	180.8	184.9	192.8	200
	MFT	223.9	200.8	204.7	196.4	199.6	199.2	
	UB	281.9	232.0	234.7	217.0	219.1	212.2	

Table 7.3: Search for control-limit replacement policy

Policy No. (k)	g(γ_k)	$E(\min(L < T_{\gamma_k}))$	$E(\Pr(L < T_{\gamma_k}))$	$g(\gamma_{k+1})$
0	2039	188.37	0%	530.9
1	530.9	190.95	0%	523.7
2	523.7	191.13	0%	523.2
3	523.2	191.13	0%	523.2

Table 7.4: CBM for different combinations of c_f/c_r for engines No. 81-100

c_f/c_r	ART	Cost of CBM	Cost of DN	Cost Reduction (%)
0	224.7	445.1	445.1	0
0.25	220.5	459.6	556.4	18.7
0.5	216.9	461	667.7	30.9
1	212.9	469.8	890.3	47.2
2	208.7	479.3	1335.4	64.4
3	207.3	482.4	1780.5	72.9
4	206.4	484.5	2225.7	78.2
5	205.7	486.2	2670.8	81.8
6	205.4	487	3116.0	82.4
7	205.4	487	3561.1	86.3
8	205.2	487.3	4006.2	87.8
9	204.9	488	4451.4	89.0
10	204.8	488.4	4896.5	90

Table 7.5: Comparison between CBM and TBM

c_f/c_r	ART_1	ART_2	Cost of CBM	Cost of TBM	Cost Reduction (%)
0	224.7	224.7	445.1	445.1	0
0.25	220.7	224.7	459.8	556.4	18.7
0.5	216.9	199.9	461	650.5	29.1
1	212.9	156.0	469.8	775.2	39.4
2	208.7	126.0	479.3	793.7	39.6
3	207.3	114.0	482.4	877.2	45
4	206.4	105.0	484.5	952.4	49.12
5	205.7	102.0	486.2	980.4	50.4
6	205.4	99.0	487	1010.1	51.8
7	205.4	96.0	487	1041.7	53.2
8	205.2	93.0	487.3	1075.3	54.7
9	204.9	90.0	488	1111.1	56.1
10	204.8	90.0	488.4	1111.1	56

Table 7.6: Comparison between multistate and binary structures

N	ART_1	ART_2	Cost of CBM	Cost of TBM	Cost of DN	Cost Reduction 1(%)	Cost Reduction 2(%)
2	131.4	126.3	804.7	835.0	2893.4	4.1	66.0
3	168.2	102.6	615.3	1082.1	2893.4	42.6	74.5
4	191.3	99.9	524.4	1065.6	2893.4	49.7	77.0
5	199.9	102.6	500.6	1022.8	2893.4	50.2	77.6
6	203.9	103.8	490.5	1010.5	2893.4	50.6	77.9
7	206.6	107.1	484.0	972.8	2893.4	49.5	78.2
8	205.5	128.7	489.2	819.0	2893.4	40.1	78.0
9	203.8	125.7	491.1	828.5	2893.4	40.6	78.1
10	206.2	136.5	490.0	786.1	2893.4	37.4	77.8

Chapter 8

Summary and Future Work

This chapter summarizes the contributions on using a multistate stochastic model for degradation analysis and describes some problems that remain to be further addressed, and suggests directions for future work.

8.1 Summary of Contributions

This thesis aims to provide maintenance decision makers a tool that can utilize condition monitoring data to facilitate online degradation and health monitoring and effective maintenance decision making for devices under continuous degradation. The contribution made in this thesis is summarized in four categories as described in the next four subsections.

8.1.1 Multistate Degradation Modeling Using NHCTHSMP

As most mechanical devices operate under some sort of stress, load, and static and dynamic forces, they tend to deteriorate or degrade over time. In real-world systems, this gradual deterioration process eventually causes the systems to be unable to operate at their desired level of performance, reliability, and/or availability. Therefore, the overall health status of most mechanical systems gradually deteriorates over time. In conventional reliability analyses, systems are often assumed to be in either of two possible health states, namely, the working state and the failure state. However, most mechanical devices operating under a stress or a load condition deteriorate or degrade over time and may perform at several intermediate health states ranging between working perfectly to complete failure. Each state level may reflect certain operational performance, efficiency, and physical property of the device. Multistate degradation models provide more realistic representation of mechanical systems.

Multistate stochastic models are very common tools in degradation and failure analysis. In this thesis, the structure of a general multistate stochastic process called nonhomogeneous continuous-time semi-Markov process and its application in degradation modeling are discussed. Also, how to combine this structure with an observation process indirectly reflecting the degradation process is illustrated. As a result, the mathematical framework for the degradation and observation processes associated with a condition-monitored device under multistate degradation processes with unobservable states is developed. Due to its flexible structure, it can be used as an effective tool to simultaneously formulate degradation and observation processes of devices under condition monitoring where health states are not directly observable. This structure can cover many of the previously used structures for degradation modeling such as Markov and semi-Markov processes.

8.1.2 Training a Multistate Structure with Condition Monitoring Data

The primary step to use a stochastic process for modeling the degradation and observation processes associated with a device under condition monitoring is to find a reasonable structure that can represent the degradation process. Directly observing the health states may be too costly and technically complicated, and because of that, unsupervised estimation methods need to be developed for devices with unobservable states. In an unsupervised estimation method, the data required for estimation is only the observation process. In this thesis, the main elements of the multistate structure to be used for modeling the degradation and observation processes are first illustrated. Then, an unsupervised parameter estimation method is introduced, which can employ historical condition monitoring data for training a multistate degradation structure with unobservable states. This parameter estimation method can estimate the parameters that describe the stochastic behaviour of degradation transition between states as well as the elements of the observation probability matrix, which characterize the relationship between the degradation process and the observation process.

8.1.3 Diagnostic and Prognostic Measures for Online Degradation Monitoring

The ultimate objective of condition monitoring is to provide useful information on the current and future health status of a device under operation. Generally, this type of information is reported by some important characteristic and performance measure, which are easy to understand and directly or indirectly represent some important aspects of the degradation and observation processes associated with the device under study. These important measures are defined in the sense that they can finally be used for maintenance decision making. As in reality, stochastic properties of the degradation and observation processes result in different degradation and observation patterns, it is more reasonable to employ information available on the actual operation of the device for performance measures calculation. These types of measures are referred to as dynamic measures, that is, they change over the age of the device and therefore reflect the dynamic behavior of the degradation and observation processes. In this thesis, the mathematical formulas for some important dynamic measures of NHCTHSMP are introduced. Each measure is defined to reflect a certain aspect of the degradation process so that it can be employed as online tool for diagnostics and prognostics.

8.1.4 Condition-based Replacement Model

Maintenance activities are the set of actions during the life cycle of a system (or a device), which intend to keep it in working condition as much as possible. Condition-based maintenance is a type of preventive maintenance, which is based on monitoring a parameter of the condition of a device through methods such as direct measurement, inspection, and condition monitoring. The ultimate objective of condition monitoring is to enable providing some useful information that can be used for maintenance decision making in order to avoid unnecessary maintenance actions and prevent catastrophic and costly failures.

Maintenance decision making in the presence of condition monitoring data involves finding the answer for the key question, that is, when (under what condition) to replace a degraded device. When the device is under condition monitoring, decision makers can utilize the most updated information on the degradation and observation processes obtained through condition monitoring indicators for decision making. In this thesis, the mathematical framework for a dynamic replacement policy for a device with multistate health levels and unobservable states, where the degradation process follows a NHCTHSMP is introduced. Such framework can be used as an online tool that can employ condition monitoring data to determine when to replace a degraded device. It is shown in this thesis that employing condition monitoring data for maintenance decision making can actually result in the reduction of maintenance cost. Also by employing the most updated information on the degradation process, it can prevent unnecessary maintenance actions and catastrophic failures. In addition, it is shown that condition-based maintenance can be more effective when the stochastic behaviour of the degradation process is modeled through a multistate degradation process with hidden states indirectly observable through condition monitoring.

8.2 Problems to be Further Addressed and Future Work

Although the structure of this thesis is defined in the sense that important challenges and limitations of current models in stochastic degradation modeling of devices with multistate structures are covered, there are still some problems that need to be further addressed. Also, the proposed models have some new challenges, which need to be further described. For each challenge, topics for future work are discussed.

8.2.1 Computational Complexity and Estimation Error

The complexity of applying the unsupervised estimation problem presented in this thesis is sensitive to the number of states and transitions, the structure of the transition rate functions, the number of data histories used for parameter estimation, and the level of discretization used to convert continuous time points to discrete time points (the discretization step considered throughout this thesis is 1). The computational time required for training a single multistate structure can be very long depending on the above factors. Also, the proposed estimation method does not guarantee the global optimal solution for the parameters of the model. The above challenges verify that there is a need for developing efficient methods that can (1) work better when the size of the problem is large, and (2) minimize estimation error. A possible topic for future research would be to analytically investigate the convergence rate for the parameter estimation method, which has not been discussed in this thesis. Another shortcoming of the models introduced in Chapters 4-6 is the discretization error occurring when calculating some of the required measures. Developing more efficient approaches to be able to minimize the discretization error is another topic for future research.

8.2.2 Finding the Structure of the CM Feature

The assumption made throughout this thesis is that the condition monitoring indicator and its stochastic structure are known. However, such information is not always easy to obtain in real-world problems. Therefore, it is important to investigate an efficient approach, which can be used to select the best feature as the representative of the observation process, which has a stochastic relationship with the actual degradation levels.

8.2.3 Condition-based Maintenance for Systems with More Than One Device

The condition-based framework in this thesis is applicable to a single device under gradual degradation. However, in many real cases, the decision makers have to make maintenance decisions for a system with more than one device. In such cases, the failure of one device may affect the degradation pattern of other devices. Also, the replacement cost and failure cost of each device may also be affected by the degradation pattern of other devices in the system. Therefore, it is extremely important to develop a framework, which can utilize condition monitoring information obtained from multiple sensors for maintenance decision making for the whole system, not just one element of the system. Therefore, developing such frameworks that are efficient at handling real-world systems, where the degradation of a component is affected not only by its own failure modes, but also by other components in the system, is crucial. Another topic for future research work is to develop a more efficient (from the computational point of view) condition-based replacement policies that are optimal for any type of hazard functions (e.g. non-monotonic hazard functions).

8.2.4 Efficient Model Selection

Using a stochastic process for maintenance decision making requires efficient training of the selected process to be used for degradation modeling. In this thesis, a simple enumerative approach is proposed, which can be used to compare several multistate model alternatives. This approach is subject to several challenges. The first challenge is that the number of model alternatives can theoretically be very large, which can make the model selection process very time-consuming or even infeasible. The other challenge is that it can only be used to compare different options, that is, there is no guarantee that the final selected structure is optimal. Another challenge is that applying such model selection framework requires a large amount of training and validation data, which are not always available in practical cases. Therefore, it is important to develop efficient model selection techniques that can overcome the above challenges.

Bibliography

[1] http://www.pilebuckinternational.com, Accesed April 2013.

[2] R. Kothamasu, S.H. Huang, and W.H. Verduin. System health monitoring and prognostics - a review of current paradigms and practices. *International Journal of Advanced Manufacturing Technology*, 28(9):1012–1024, 2006.

[3] http://www.nola.com/news/gulf-oil-spill, Accesed April 2013.

[4] http://weibull.reliasoft.com, Accesed April 2013.

[5] http://cmore.mie.utoronto.ca/masterclass.html, Accesed April 2013.

[6] W. Kuo and M.J. Zuo. *Optimal Reliability Modeling: Principles and Applications*. Wiley, 2002.

[7] A. Lisnianski and G. Levitin. *Multi-State System Reliability: Assessment, Optimization and Applications*. Series on Quality, Reliability and Engineering Statistics, 6. World Scientific, 2003.

[8] Z. Li and K.C. Kapur. Continuous-state reliability measures based on fuzzy sets. *IIE Transactions (Institute of Industrial Engineers)*, 44(11):1033–1044, 2012.

[9] R. Moghaddass and M.J. Zuo. A parameter estimation method for a condition-monitored device under multi-state deterioration. *Reliability Engineering and System Safety*, 106:94–103, 2012.

[10] K. Yang and J. Xue. Continuous state reliability analysis. In *Proceedings of the Annual Reliability and Maintainability Symposium*, pages 251–257, 1996.

[11] G. Picciolo, F. Galli, A. Biamonti, and P. Magni. Determining the maximum periodic inspection interval for medium voltage motors using a Markov model. *Production Planning and Control*, 19(4):356–364, 2008.

[12] Z.S. Ye, L.C. Tang, and H.Y. Xu. A distribution-based systems reliability model under extreme shocks and natural degradation. *IEEE Transactions on Reliability*, 60(1):246–256, 2011.

[13] M.J. Kim, V. Makis, and R. Jiang. Parameter estimation for partially observable systems subject to random failure. *Applied Stochastic Models in Business and Industry*, 2012. Article in Press.

[14] J.A.M. Hontelez, H.H. Burger, and D.J.D. Wijnmalen. Optimum condition-based maintenance policies for deteriorating systems with partial information. *Reliability Engineering and System Safety*, 51(3):267–274, 1996.

[15] Y. Peng and M. Dong. A prognosis method using age-dependent hidden semi-Markov model for equipment health prediction. *Mechanical Systems and Signal Processing*, 25(1):237–252, 2011.

[16] H. Teng, J. Zhao, X. Jia, Y. Jia, X. Zhang, and L. Cai. Experimental study on gearbox prognosis using total life vibration analysis. In *Prognostics and System Health Management Conference, PHM-Shenzhen*, 2011.

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

[18] S. Katipamula and M.R. Brambley. Methods for fault detection, diagnostics, and prognostics for building systems - a review, part II. *HVAC and R Research*, 11(2):169–187, 2005.

[19] S. Katipamula and M.R. Brambley. Methods for fault detection, diagnostics, and prognostics for building systems - a review, part I. *HVAC and R Research*, 11(1):3–25, 2005.

[20] Y. Peng, M. Dong, and M.J. Zuo. Current status of machine prognostics in condition-based maintenance: A review. *International Journal of Advanced Manufacturing Technology*, 50(1-4):297–313, 2010.

[21] V.T. Tran, H. Thom Pham, B.-S. Yang, and T. Tien Nguyen. Machine performance degradation assessment and remaining useful life prediction using proportional hazard model and support vector machine. *Mechanical Systems and Signal Processing*, 32:320–330, 2012.

[22] X.S. Si, W. Wang, C.H. Hu, and D.H. Zhou. Remaining useful life estimation - a review on the statistical data driven approaches. *European Journal of Operational Research*, 213(1):1–14, 2011.

[23] J.Z. Sikorska, M. Hodkiewicz, and L. Ma. Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems and Signal Processing*, 25(5):1803–1836, 2011.

[24] A. Prajapati, J. Bechtel, and S. Ganesan. Condition based maintenance: A survey. *Journal of Quality in Maintenance Engineering*, 18(4):384–400, 2012.

[25] K.T. Huynh, A. Barros, and C. Brenguer. Maintenance decision-making for systems operating under indirect condition monitoring: Value of online information and impact of measurement uncertainty. *IEEE Transactions on Reliability*, 61(2):410–425, 2012.

[26] R. Moghaddass, M.J. Zuo, and X. Zhao. *Modeling Multi-State Equipment Degradation with Non-Homogeneous Continuous-Time Hidden Semi-Markov Process*, pages 151–181. Diagnostics and Prognostics of Engineering Systems: Methods and Techniques. IGI Global, 2013.

[27] R. Moghaddass and M.J. Zuo. An integrated framework for online diagnostic and prognostic health monitoring using a multistate deterioration process. Submitted to *Reliability Engineering and System Safety*, March 2013.

[28] G.D. Goel and K. Murari. Two-unit cold-standby redundant system subject to random checking, corrective maintenance and system replacement with repairable and non-repairable types of failure. *Microelectronics Reliability*, 30 (4):661–665, 1990.

[29] Y. Liu, Y. Li, H.Z. Huang, M.J. Zuo, and Z. Sun. Optimal preventive maintenance policy under fuzzy bayesian reliability assessment environments. *IIE Transactions (Institute of Industrial Engineers)*, 42(10):734–745, 2010.

[30] R. Moghaddass, M.J. Zuo, and W. Wang. Availability of a general k-out-of-n:g system with non-identical components considering shut-off rules using quasi-birthdeath process. *Reliability Engineering and System Safety*, 96(4):489–496, 2011.

[31] R. Moghaddass, M.J. Zuo, and J. Qu. Reliability and availability analysis of a repairable k-out-of-n : G system with r repairmen subject to shut-off rules. *IEEE Transactions on Reliability*, 60(3):658–666, 2011.

[32] B.M. Hsu and M.H. Shu. Reliability assessment and replacement for machine tools under wear deterioration. *International Journal of Advanced Manufacturing Technology*, 48(1-4):355–365, 2010.

[33] Y. Gu and J. Li. Multi-state system reliability: A new and systematic review. *Procedia Engineering*, 29:531–536, 2012.

[34] V. Barbu and N. Limnios. *Semi-Markov Chains and Hidden Semi-Markov Models toward Applications: Their Use in Reliability and DNA Analysis*. Springer Publishing Company, Incorporated, 2008.

[35] B. Castanier, C. Brenguer, and A. Grall. A sequential condition-based repair/replacement policy with non-periodic inspections for a system subject to continuous wear. *Applied Stochastic Models in Business and Industry*, 19(4): 327–347, 2003.

[36] R.H. Yeh. Optimal inspection and replacement policies for multi-state deteriorating systems. *European Journal of Operational Research*, 96(2):248–259, 1997.

[37] S. Bloch-Mercier. A preventive maintenance policy with sequential checking procedure for a Markov deteriorating system. *European Journal of Operational Research*, 142(3):548–576, 2002.

[38] M.J. Kim, R. Jiang, V. Makis, and C.G. Lee. Optimal Bayesian fault prediction scheme for a partially observable system subject to random failure. *European Journal of Operational Research*, 214(2):331–339, 2011.

[39] M.J. Carr and W. Wang. An approximate algorithm for prognostic modelling using condition monitoring information. *European Journal of Operational Research*, 211(1):90–96, 2011.

[40] J. Sun, H. Zuo, W. Wang, and M.G. Pecht. Application of a state space modeling technique to system prognostics based on a health index for condition-based maintenance. *Mechanical Systems and Signal Processing*, 28:585–596, 2012.

[41] A. Ghasemi, S. Yacout, and M.S. Ouali. Optimal condition based maintenance with imperfect information and the proportional hazards model. *International Journal of Production Research*, 45(4):989–1012, 2007.

[42] Y. Yang. *Contributions to the Analysis of Multistate and Degradation Data*. Phd thesis, The University of Michigan, 2011.

[43] S.Z. Yu. Hidden semi-Markov models. *Artificial Intelligence*, 174(2):215–243, 2010.

[44] L.F. Meira-Machado, J. de UnA-alvarez, C. Cadarso-Surez, and P.K. Andersen. Multi-state models for the analysis of time-to-event data. *Statistical Methods in Medical Research*, 18(2):195–222, 2009.

[45] Y. Liang, X. Liu, Y. Lou, and B. Shan. An improved noise-robust voice activity detector based on hidden semi-Markov models. *Pattern Recognition Letters*, 32(7):1044–1053, 2011.

[46] S. Chauhan, P. Wang, C. Sing Lim, and V. Anantharaman. A computer-aided MFCC-based HMM system for automatic auscultation. *Computers in Biology and Medicine*, 38(2):221–233, 2008.

[47] A. Pievatolo, F. Ruggeri, and R. Soyer. A Bayesian hidden Markov model for imperfect debugging. *Reliability Engineering and System Safety*, 103:11–21, 2012.

[48] J. Chen and Y.-C. Jiang. Development of hidden semi-Markov models for diagnosis of multiphase batch operation. *Chemical Engineering Science*, 66 (6):1087–1099, 2011.

[49] G. Morcous, Z. Lounis, and M.S. Mirza. Identification of environmental categories for Markovian deterioration models of bridge decks. *Journal of Bridge Engineering*, 8(6):353–361, 2003.

[50] C. Scheffer, H. Engelbrecht, and P.S. Heyns. A comparative evaluation of neural networks and hidden Markov models for monitoring turning tool wear. *Neural Computing and Applications*, 14(4):325–336, 2005.

[51] F. De Rango, F. Veltri, and S. Marano. Channel modeling approach based on the concept of degradation level discrete-time Markov chain: UWB system case study. *IEEE Transactions on Wireless Communications*, 10(4):1098–1107, 2011.

[52] H. Qiu, H. Liao, and J. Lee. Degradation assessment for machinery prognostics using hidden Markov models. In *Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference - DETC2005*, volume 1 A, pages 531–537, 2005.

[53] C.C. Huang and J. Yuan. A two-stage preventive maintenance policy for a multi-state deterioration system. *Reliability Engineering and System Safety*, 95(11):1255–1260, 2010.

[54] M. Dong and D. He. Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis. *European Journal of Operational Research*, 178(3):858–878, 2007.

[55] Y. Peng and M. Dong. A hybrid approach of HMM and grey model for age-dependent health prediction of engineering assets. *Expert Systems with Applications*, 38(10):12946–12953, 2011.

[56] A. Chen and G.S. Wu. Real-time health prognosis and dynamic preventive maintenance policy for equipment under aging Markovian deterioration. *International Journal of Production Research*, 45(15):3351–3379, 2007.

[57] D. Lin and V. Makis. On-line parameter estimation for a failure-prone system subject to condition monitoring. *Journal of Applied Probability*, 41(1):211–220, 2004.

[58] M.J. Kim, V. Makis, and R. Jiang. Parameter estimation in a condition-based maintenance model. *Statistics and Probability Letters*, 80(21-22):1633–1639, 2010.

[59] M. Muhammad and M. Amin Abd Majid. Reliability and availability evaluation for a multi-state system subject to minimal repair. *Journal of Applied Sciences*, 11(11):2036–2041, 2011.

[60] I.W. Soro, M. Nourelfath, and D. At-Kadi. Performance evaluation of multi-state degraded systems with minimal repairs and imperfect preventive maintenance. *Reliability Engineering and System Safety*, 95(2):65–69, 2010.

[61] S.H. Sim and J. Endrenyi. Failure-repair model with minimal & major maintenance. *IEEE Transactions on Reliability*, 42(1):134–140, 1993.

[62] M. Dong and Y. Peng. Equipment PHM using non-stationary segmental hidden semi-Markov model. *Robotics and Computer-Integrated Manufacturing*, 27(3):581–590, 2011.

[63] C.T. Chen. Dynamic preventive maintenance strategy for an aging and deteriorating production system. *Expert Systems with Applications*, 38(5):6287–6293, 2011.

[64] Y.F. Li, E. Zio, and Y.H. Lin. A multistate physics model of component degradation based on stochastic petri nets and simulation. *IEEE Transactions on Reliability*, 61(4):921–931, 2012.

[65] Y. Liu and H.Z. Huang. Optimal replacement policy for multi-state system under imperfect maintenance. *IEEE Transactions on Reliability*, 59(3):483–495, 2010.

[66] M. Dong, D. He, P. Banerjee, and J. Keller. Equipment health diagnosis and prognosis using hidden semi-Markov models. *International Journal of Advanced Manufacturing Technology*, 30(7-8):738–749, 2006.

[67] C. Teresa Lam and R.H. Yeh. Optimal replacement policies for multistate deteriorating systems. *Naval Research Logistics*, 41(3):303–315, 1994.

[68] O. Chryssaphinou, N. Limnios, and S. Malefaki. Multi-state reliability systems under discrete time semi-Markovian hypothesis. *IEEE Transactions on Reliability*, 60(1):80–87, 2011.

[69] J.P. Kharoufeh, C.J. Solo, and M.Y. Ulukus. Semi-Markov models for degradation-based reliability. *IIE Transactions (Institute of Industrial Engineers)*, 42(8):599–612, 2010.

[70] M.H. Shu, B.M. Hsu, and K.C. Kapur. Dynamic performance measures for tools with multi-state wear processes and their applications for tool design and selection. *International Journal of Production Research*, 48(16):4725–4744, 2010.

[71] J.K. Vaurio. Reliability characteristics of components and systems with tolerable repair times. *Reliability Engineering and System Safety*, 56(1):43–52, 1997.

[72] P. Nelson and S. Wang. Dynamic reliability via computational solution of generalized state-transition equations for entry-time processes. *Reliability Engineering and System Safety*, 92(9):1281–1293, 2007.

[73] A. Ghasemi, S. Yacout, and M.S. Ouali. Parameter estimation methods for condition-based maintenance with indirect observations. *IEEE Transactions on Reliability*, 59(2):426–439, 2010.

[74] A. Ghasemi, S. Yacout, and M.S. Ouali. Evaluating the reliability function and the mean residual life for equipment with unobservable states. *IEEE Transactions on Reliability*, 59(1):45–54, 2010.

[75] G. Becker, L. Camarinopoulos, and G. Zioutas. A semi-Markovian model allowing for inhomogenities with respect to process time. *Reliability Engineering and System Safety*, 70(1):41 – 48, 2000.

[76] N. Limnios. Reliability measures of semi-Markov systems with general state space. *Methodology and Computing in Applied Probability*, 14(4):895–917, 2012.

[77] C.T. Chen, Y.W. Chen, and J. Yuan. On a dynamic preventive maintenance policy for a system under inspection. *Reliability Engineering and System Safety*, 80(1):41–47, 2003.

[78] J.H. Chiang and J. Yuan. Optimal maintenance policy for a Markovian system under periodic inspection. *Reliability Engineering and System Safety*, 71(2): 165–172, 2001.

[79] V. Barbu and N. Limnios. Nonparametric estimation for failure rate functions of discrete time semi-Markov processes. In Mikhail Nikulin, Daniel Commenges, and Catherine Huber, editors, *Probability, Statistics and Modelling in Public Health*, pages 53–72. Springer US, 2006.

[80] D. Lin and V. Makis. Filters and parameter estimation for a partially observable system subject to random failure with continuous-range observations. *Advances in Applied Probability*, 36(4):1212–1230, 2004.

[81] T. Heyns, P.S. Heyns, and J.P. De Villiers. Combining synchronous averaging with a Gaussian mixture model novelty detection scheme for vibration-based condition monitoring of a gearbox. *Mechanical Systems and Signal Processing*, 32:200–215, 2012.

[82] J. Yu. Machine tool condition monitoring based on an adaptive Gaussian mixture model. *Journal of Manufacturing Science and Engineering, Transactions of the ASME*, 134(3), 2012.

[83] J. Yu. A hybrid feature selection scheme and self-organizing map model for machine health assessment. *Applied Soft Computing Journal*, 11(5):4041–4054, 2011.

[84] S. Yella, N.K. Gupta, and M.S. Dougherty. Comparison of pattern recognition techniques for the classification of impact acoustic emissions. *Transportation Research Part C: Emerging Technologies*, 15(6):345–360, 2007.

[85] F. Zorriassatine, A. Al-Habaibeh, R.M. Parkin, M.R. Jackson, and J. Coy. Novelty detection for practical pattern recognition in condition monitoring of multivariate processes: A case study. *International Journal of Advanced Manufacturing Technology*, 25(9-10):954–963, 2005.

[86] X. Wu, Y. Li, and W. Teng. Modified hidden semi-Markov models for motor wear prognosis. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, 226(2):174–179, 2012.

[87] X. Wu, Y. Li, T.D. Lundell, and A.K. Guru. Integrated prognosis of ac servo motor driven linear actuator using hidden semi-Markov models. In *IEEE International Electric Machines and Drives Conference*, pages 1408–1413, 2009.

[88] Q. Huo and C. Chan. Contextual vector quantization for speech recognition with discrete hidden Markov model. *Pattern Recognition*, 28(4):513–517, 1995.

[89] P. Baruah and R.B. Chinnam. HMMs for diagnostics and prognostics in machining processes. *International Journal of Production Research*, 43(6):1275–1293, 2005.

[90] A. Saxena, K. Goebel, D. Simon, and N. Eklund. Damage propagation modeling for aircraft engine run-to-failure simulation. In *International Conference on Prognostics and Health Management*, 2008.

[91] R. Huang, L. Xi, X. Li, C. Richard Liu, H. Qiu, and J. Lee. Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods. *Mechanical Systems and Signal Processing*, 21(1): 193–207, 2007.

[92] L.R. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257 –286, 1989.

[93] M. Bladt and M. Srensen. Efficient estimation of transition rates between credit ratings from observations at discrete time points. *Quantitative Finance*, 9(2):147–160, 2009.

[94] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):pp. 1–38, 1977. ISSN 00359246.

[95] X.D. Huang and M.A. Jack. Semi-continuous hidden markov models for speech signals. *Computer Speech and Language*, 3(3):239–251, 1989.

[96] A.A.R. Sa, A.O. Andrade, A.B. Soares, and S.J. Nasuto. Estimation of hidden Markov models parameters using differential evolution. In *AISB 2008 Convention: Communication, Interaction and Social Intelligence - Proceedings of the AISB 2008 Symposium on Swarm Intelligence Algorithms and Applications*, pages 51–56, 2008.

[97] N. Giordana and W. Pieczynski. Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(5):465–475, 1997.

[98] L.A. Liporace. Maximum likelihood estimation for multivariate observations of Markov sources. *IEEE Transactions on Information Theory*, IT-28(5 pt 1): 729–734, 1982.

[99] K. Bijak. Genetic algorithms as an alternative method of parameter estimation and finding most likely sequences of states of hidden Markov chains for HMMs and hybrid HMM/ANN models. *Fundamenta Informaticae*, 86(1-2): 1–17, 2008.

[100] D. Lin and V. Makis. Recursive filters for a partially observable system subject to random failure. *Advances in Applied Probability*, 35(1):207–227, 2003.

[101] Y. Gudon. Computational methods for discrete hidden semi-Markov chains. *Applied Stochastic Models in Business and Industry*, 15(3):195–224, 1999.

[102] Y. Gudon. Estimating hidden semi-Markov chains from discrete sequences. *Journal of Computational and Graphical Statistics*, 12(3):604–639, 2003.

[103] S-Z. Yu and H. Kobayashi. A hidden semi-Markov model with missing data and multiple observation sequences for mobility tracking. *Signal Processing*, 83(2):235 – 250, 2003.

[104] S.E. Levinson. Continuously variable duration hidden Markov models for automatic speech recognition. *Computer Speech and Language*, 1(1):29–45, 1986.

[105] D.A. Tobon-Mejia, K. Medjaher, N. Zerhouni, and G. Tripot. A mixture of Gaussians hidden Markov model for failure diagnostic and prognostic. In *IEEE International Conference on Automation Science and Engineering, CASE 2010*, pages 338–343, 2010.

[106] Y.-W. Liu and K.C. Kapur. Reliability measures for dynamic multistate nonrepairable systems and their applications to system performance evaluation. *IIE Transactions (Institute of Industrial Engineers)*, 38(6):511–520, 2006.

[107] J.P. Kharoufeh and S.M. Cox. Stochastic models for degradation-based reliability. *IIE Transactions (Institute of Industrial Engineers)*, 37(6):533–542, 2005.

[108] M. Dong and D. He. A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology. *Mechanical Systems and Signal Processing*, 21(5):2248–2266, 2007.

[109] L. Doyen and O. Gaudoin. Modeling and assessment of aging and efficiency of corrective and planned preventive maintenance. *IEEE Transactions on Reliability*, 60(4):759–769, 2011.

[110] J. Nilsson and L. Bertling. Maintenance management of wind power systems using condition monitoring systems - life cycle cost analysis for two case studies. *IEEE Transactions on Energy Conversion*, 22(1):223–229, 2007.

[111] R. Ahmad and S. Kamaruddin. A review of condition-based maintenance decision-making. *European Journal of Industrial Engineering*, 6(5):519–541, 2012.

[112] J. Cibulka, M.K. Ebbesen, G. Hovland, K.G. Robbersmyr, and M.R. Hansen. A review on approaches for condition based maintenance in applications with induction machines located offshore. *Modeling, Identification and Control*, 33 (2):69–86, 2012.

[113] L. Zhang, X. Li, and J. Yu. A review of fault prognostics in condition based maintenance. In *Proceedings of SPIE - The International Society for Optical Engineering*, volume 6357 II, 2006.

[114] V. Makis and X. Jiang. Optimal replacement under partial observations. *Mathematics of Operations Research*, 28(2):382–394, 2003.

[115] L.M. Maillart. Maintenance policies for systems with condition monitoring and obvious failures. *IIE Transactions (Institute of Industrial Engineers)*, 38 (6):463–475, 2006.

[116] Z. Sinuany-Stern. Replacement policy under partially observed Markov process. *International Journal of Production Economics*, 29(2):159–166, 1993.

[117] D. Rosenfield. Markovian deterioration with uncertain information - a more general model. *Naval Research Logistics*, 23(3):389–405, 1976.

[118] M. Kurt and J.P. Kharoufeh. Monotone optimal replacement policies for a Markovian deteriorating system in a controllable environment. *Operations Research Letters*, 38(4):273–279, 2010.

[119] J.S. Ivy and S.M. Pollock. Marginally monotonic maintenance policies for a multi-state deteriorating machine with probabilistic monitoring, and silent failures. *IEEE Transactions on Reliability*, 54(3):489–497, 2005.

[120] C.L. Tomasevicz and S. Asgarpoor. Preventive maintenance using continuous-time semi-Markov processes. In *38th Annual North American Power Symposium, NAPS-2006 Proceedings*, pages 3–8, 2006.

[121] E. Byon and Y. Ding. Season-dependent condition-based maintenance for a wind turbine using a partially observed markov decision process. *IEEE Transactions on Power Systems*, 25(4):1823–1834, 2010.

[122] C. Bunks, D. McCarthy, and T. Al-Ani. Condition-based maintenance of machines using hidden markov models. *Mechanical Systems and Signal Processing*, 14(4):597–612, 2000.

[123] D.C. Brauer and G.D. Brauer. Reliability-centered maintenance. *IEEE Transactions on Reliability*, R-36(1):17–24, 1987.

[124] N.S. Arunraj and J. Maiti. Risk-based maintenance-techniques and applications. *Journal of Hazardous Materials*, 142(3):653–661, 2007.

[125] C.Joseph Lu and William O. Meeker. Using degradation measures to estimate a time-to-failure distribution. *Technometrics*, 35(2):161–174, 1993.

[126] I.T. Yu and C.D. Fuh. Estimation of time to hard failure distributions using a three-stage method. *IEEE Transactions on Reliability*, 59(2):405–412, 2010.

[127] L. Jiang, Q. Feng, and D.W. Coit. Reliability analysis for dependent failure processes and dependent failure threshold. In *ICQR2MSE 2011 - Proceedings of 2011 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering*, pages 30–34, 2011.

[128] G. Koole. A formula for tail probabilities of Cox distributions. *Journal of Applied Probability*, 41(3):935–938, 2004.

[129] F. Wang, F. Sun, and B.G. Cao. Feature fusion of mechanical faults based on evolutionary computation. *Insight - Non-Destructive Testing and Condition Monitoring*, 49(8):471–475, 2007.

[130] H. Jgou, M. Douze, and C. Schmid. Product quantization for nearest neighbor search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33 (1):117–128, 2011.

[131] J. Janssen and R. Manca. Numerical solution of non-homogeneous semi-Markov processes in transient case. *Methodology And Computing In Applied Probability*, 3:271–293, 2001. ISSN 1387-5841.

[132] E. Lpez Droguett, M. das Chagas Moura, C. Magno Jacinto, and M. Feliciano Silva Jr. A semi-Markov model with Bayesian belief network based human error probability for availability assessment of downhole optical monitoring systems. *Simulation Modelling Practice and Theory*, 16(10):1713 – 1727, 2008.

[133] G. D'Amico, J. Janssen, and R. Manca. Discrete time non-homogeneous semi-Markov reliability transition credit risk models and the default distribution functions. *Computational Economics*, 38(4):465–481, 2011.

[134] M.C. Moura and E.L. Droguett. Mathematical formulation and numerical treatment based on transition frequency densities and quadrature methods for non-homogeneous semi-Markov processes. *Reliability Engineering and System Safety*, 94(2):342–349, 2009.

[135] A. Blasi, J. Janssen, and R. Manca. Numerical treatment of homogeneous and non-homogeneous semi-markov reliability models. *Communications in Statistics - Theory and Methods*, 33(3):697–714, 2004.

[136] G. Becker, L. Camarinopoulos, and G. Zioutas. Semi-Markovian model allowing for inhomogenities with respect to process time. *Reliability Engineering and System Safety*, 70(1):41–48, 2000.

[137] H. Liao, E.A. Elsayed, and L.-Y. Chan. Maintenance of continuously monitored degrading systems. *European Journal of Operational Research*, 175(2): 821–835, 2006.

[138] E.J. Ward. A review and comparison of four commonly used bayesian and maximum likelihood model selection tools. *Ecological Modelling*, 211(1-2): 1–10, 2008.

[139] M.N. Park and J.Y. Ha. Model selection criterion using confusion models for HMM topology optimization. In *SICE-ICASE International Joint Conference*, pages 1004–1008, 2006.

[140] D. Li, A. Biem, and J. Subrahmonia. HMM topology optimization for handwriting recognition. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, volume 3, pages 1521–1524, 2001.

[141] T. Boutros and M. Liang. Detection and diagnosis of bearing and cutting tool faults using hidden markov models. *Mechanical Systems and Signal Processing*, 25(6):2102–2124, 2011.

[142] X. Zhao, M.J. Zuo, and T.H. Patel. Generating an indicator for pump impeller damage using half and full spectra, fuzzy preference-based rough sets and pca. *Measurement Science and Technology*, 23(4), 2012.

[143] Z. Shen, Z. He, X. Chen, C. Sun, and Z. Liu. A monotonic degradation assessment index of rolling bearings using fuzzy support vector data description and running time. *Sensors (Switzerland)*, 12(8):10109–10135, 2012.

[144] S.J. Loutridis. Instantaneous energy density as a feature for gear fault detection. *Mechanical Systems and Signal Processing*, 20(5):1239–1253, 2006.

[145] G. Corradi, J. Janssen, and R. Manca. Numerical treatment of homogeneous semi-Markov processes in transient case-a straightforward approach. *Methodology and Computing in Applied Probability*, 6(2):233–246, 2004.

[146] R. Moghaddass, M.J. Zuo, Y. Liu, and H.Z. Huang. Reliability analysis of a multistate deteriorating device using inspection data. Submitted to *IIE Transactions*, March 2013.

[147] D. Banjevic, A.K.S. Jardine, V. Makis, and M. Ennis. A control-limit policy and software for condition-based maintenance optimization. *INFOR*, 39(1): 32–50, 2001.

[148] H.C. Tijms. *A First Course in Stochastic Models*. Wiley, 2003.

[149] C.L. Tomasevicz and S. Asgarpoor. Optimum maintenance policy using semi-Markov decision processes. *Electric Power Systems Research*, 79(9):1286–1291, 2009.

[150] S.V. Amari, L. McLaughlin, and H. Pham. Cost-effective condition-based maintenance using Markov decision processes. In *Proceedings - Annual Reliability and Maintainability Symposium*, pages 464–469, 2006.

[151] S.M. Ross. *Introduction to probability models*. Introduction to Probability Models. Academic Press, 2003.

[152] J.A. DeCastro, J.S. Litt, and D.K. Frederick. A modular aero-propulsion system simulation of a large commercial aircraft engine. AIAA-2008-4579, 44th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit, July 21-23 2008.

[153] Z. Li, X. Yan, Z. Guo, Y. Zhang, C. Yuan, and Z. Peng. Condition monitoring and fault diagnosis for marine diesel engines using information fusion techniques. *Elektronika ir Elektrotechnika*, 123(7):109–112, 2012.

[154] Q. He, R. Yan, F. Kong, and R. Du. Machine condition monitoring using principal component representations. *Mechanical Systems and Signal Processing*, 23(2):446–466, 2009.

[155] A. Malhi and R.X. Gao. Pca-based feature selection scheme for machine defect classification. *IEEE Transactions on Instrumentation and Measurement*, 53 (6):1517–1525, 2004.

[156] C.T. Yiakopoulos, K.C. Gryllias, and I.A. Antoniadis. Rolling element bearing fault detection in industrial environments based on a K-means clustering approach. *Expert Systems with Applications*, 38(3):2888–2911, 2011.

[157] S. Yue and P. Pilon. A comparison of the power of the t test, Mann-Kendall and bootstrap tests for trend detection. *Hydrological Sciences Journal*, 49(1): 21–38, 2004.

[158] V. Makis and A.K.S. Jardine. Computation of optimal policies in replacement models. *IMA Journal of Management Mathematics*, 3(3):169–175, 1991.

[159] V. Makis and A.K.S. Jardine. Optimal replacement in the proportional hazards model. *INFOR*, 30(1):172183, 1992.

Appendix A

In this appendix, the steps needed to find the control-limit replacement policy are illustrated. These steps are very similar to the steps given in [41, 158] to find the control-limit replacement policy except that (1) the reliability function is directly conditional on the history of observed condition monitoring data, (2) the downtime cost and the cost and time of failure replacement and preventive replacement are also considered. While the focus of this thesis is on devices with non-decreasing hazard functions, an approximate approach to deal with the non-monotonic hazard functions is also provided. It should be noted that analytical evaluation of this approach is out of the scope of this thesis. Due to its computational complexities, a more efficient approach to deal with cases with a general form of hazard function (e.g. non-monotonic hazard rates) needs to be investigated in future work.

In order to compare the two maintenance options for policy γ at time t at each iteration of the policy improvement step, the difference between the associated value functions can be calculated as follows:

$$
\begin{aligned}
&V^2(\vartheta^t,\gamma) - V^1(\vartheta^t,\gamma) = \\
&\qquad \left[c_r + c_f + c_d \times (t_r + t_f) + V(\vartheta^0,\gamma)\right]\left[1 - R(t+\Delta|\vartheta^t)\right] - \\
&\qquad g(\gamma)\left(\tau_t(t+\Delta|\vartheta^t) + (t_r+t_f)(1 - R(t+\Delta|\vartheta^t))\right) \\
&+ \left[\sum_{\vartheta^{t+\Delta}\neq F} \Pr(\vartheta^t,\vartheta^{t+\Delta},0) \times V(\vartheta^{t+\Delta},\gamma)\right] - \left[c_r + c_d \times t_r - g(\gamma)\times t_r + V(\vartheta^0,\gamma)\right] = \\
&(c_f + c_d \times t_f)(1 - R(t+\Delta|\vartheta^t)) - g(\gamma)\left(\tau_t(t+\Delta|\vartheta^t) + t_f \times (1 - R(t+\Delta|\vartheta^t))\right) + \\
&\qquad \left[\sum_{\vartheta^{t+\Delta}\neq F} \Pr(\vartheta^t,\vartheta^{t+\Delta},0) \times V(\vartheta^{t+\Delta},\gamma) - V^1(\vartheta^{t+\Delta},\gamma) \times R(t+\Delta|\vartheta^t)\right]. \quad \text{(A.1)}
\end{aligned}
$$

For notational convenience, let $\varpi(\vartheta^t,\gamma)$ be:

$$\varpi(\vartheta^t,\gamma) = \left[\sum_{\vartheta^{t+\Delta}\neq F} \Pr(\vartheta^t,\vartheta^{t+\Delta},0)\times V(\vartheta^{t+\Delta},\gamma) - V^1(\vartheta^{t+\Delta},\gamma)\times R(t+\Delta|\vartheta^t)\right]. \quad \text{(A.2)}$$

It is obvious that if $\varpi(\vartheta^t,\gamma)+(c_f+c_d\times t_f)(1-R(t+\Delta|\vartheta^t))-g(\gamma)(\tau_t(t+\Delta|\vartheta^t)+t_f\times(1-R(t+\Delta|\vartheta^t))\geq 0$, then $V^2(\vartheta,\gamma)\geq V^1(\vartheta,\gamma)$ and therefore the optimal decision at time t under policy γ is to replace immediately. On the other hand, if $\varpi(\vartheta^t,\gamma^t)+(c_f+c_d\times t_f)(1-R(t+\Delta|\vartheta^t))-g(\gamma)(\tau_t(t+\Delta|\vartheta^t)+t_f\times(1-R(t+\Delta|\vartheta^t))<0$, then $V^2(\vartheta,\gamma)<V^1(\vartheta,\gamma)$ and the optimal decision at time t under policy γ is do nothing. As a result, the following replacement rule at time t can be defined:

$$\begin{cases} CI(\vartheta^t)\geq g(\gamma) & \to D(\vartheta^t,\gamma)=1 \\ \Psi^t=0 & \to D(\vartheta^t,\gamma)=1 \\ CI(\vartheta^t)<g(\gamma) & \to D(\vartheta^t,\gamma)=0 \end{cases}, \quad \text{(A.3)}$$

where

$$CI(\vartheta^t) = \frac{[c_f+c_d\times(t_f)]\left[1-R(t+\Delta|\vartheta^t)\right]}{\tau(t+\Delta|\vartheta^t)+t_f\left[1-R(t+\Delta|\vartheta^t)\right]} + \frac{\varpi(\vartheta^t,\gamma)}{\tau(t+\Delta|\vartheta^t)+t_f\left[1-R(t+\Delta|\vartheta^t)\right]}. \quad \text{(A.4)}$$

The term $CI(\vartheta^t)$ is the control index which can be calculated at time t based on the condition of the device. The first term on the left side of $CI(\vartheta^t)$ can be simply calculated by finding the conditional reliability function as given in Eq. (5.10) and the expected conditional operation time in the interval $(t,t+\Delta)$. The remaining part of calculating this control index is to evaluate $\varpi(\vartheta^t,\gamma)$.

It is obvious that for any policy γ, $V(\vartheta^{t+\Delta},\gamma)\leq V^1(\vartheta^{t+\Delta},\gamma)$ ($V(\vartheta^{t+\Delta},\gamma)=\min\{V^1(\vartheta^{t+\Delta},\gamma),V^2(\vartheta^{t+\Delta},\gamma)\}$). Therefore, we have
$\Pr(\vartheta^t,\vartheta^{t+\Delta},0)\times V(\vartheta^{t+\Delta},\gamma)-\Pr(\vartheta^t,\vartheta^{t+\Delta},0)\times V^1(\vartheta^{t+\Delta},\gamma)\leq 0$. Now, as it is known that $\sum_{\vartheta^{t+\Delta}\neq F}\Pr(\vartheta^t,\vartheta^{t+\Delta},0)=R(t+\Delta|\vartheta^t)$, then:
$\left[\sum_{\vartheta^{t+\Delta}\neq F}\Pr(\vartheta^t,\vartheta^{t+\Delta},0)\times V(\vartheta^{t+\Delta},\gamma)-V^1(\vartheta^{t+\Delta},\gamma)\times R(t+\Delta|\vartheta^t)\right]\leq 0$, which means $\varpi(\vartheta^t,\gamma)\leq 0$. Based on the above property, several scenarios can be considered for $CI(\vartheta^t)$. It is clear that when $\dfrac{[c_f+c_d\times(t_f)]\left[1-R(t+\Delta|\vartheta^t)\right]}{\tau(t+\Delta|\vartheta^t)+t_f\left[1-R(t+\Delta|\vartheta^t)\right]}<g(\gamma)$, then $CI(\vartheta^t)<g(\gamma)$, which means $\to D(\vartheta^t,\gamma)=0$, that is, the optimal decision is do

nothing. It can also be shown from the results in [159] that when the hazard function is non-decreasing (which is usually the case for devices under gradual degradation and no repair), if $\frac{[c_f + c_d \times (t_f)]\left[1 - R(t+\Delta|\vartheta^t)\right]}{\tau(t+\Delta|\vartheta^t) + t_f\left[1 - R(t+\Delta|\vartheta^t)\right]} \geq g(\gamma)$, then $CI(\vartheta^t) \geq g$ and therefore $D(\vartheta^t, \lambda) = 1$ and $V(\vartheta^t, \gamma) = V^1(\vartheta^t, \gamma)$, that is the optimal decision at time t is to replacement immediately. Under such a condition, the calculation of the $\varpi(\vartheta^t, \gamma)$ is unnecessary. Therefore, the control index can be simplified to:

$$CI(\vartheta^t) = \frac{[c_f + c_d \times (t_f)]\left[1 - R(t+\Delta|\vartheta^t)\right]}{\tau(t+\Delta|\vartheta^t) + t_f\left[1 - R(t+\Delta|\vartheta^t)\right]}. \tag{A.5}$$

In a general form of hazard function (e.g. when it does not have a monotonic trend), $\varpi(\vartheta^t, \gamma)$ should be evaluated for maintenance decision making. It is important to note that if $V(\vartheta^{t+\Delta}, \gamma) = V^1(\vartheta^{t+\Delta}, \gamma)$ for all $\vartheta^{t+\Delta} \in S$, that is, if the cost of replace immediately is less than the cost of do nothing at time $(t + \Delta)$ for all possible conditions $(\vartheta^t + \Delta)$, then $\varpi(\vartheta^t, \gamma) = 0$. However, if $V(\vartheta^{t+\Delta}, \gamma) = V^2(\vartheta^{t+\Delta}, \gamma)$ for any $\vartheta^{t+\Delta} \in S$, then $\varpi(\vartheta^t, \gamma)$ can be less than zero. The remainder of this appendix deals with cases where $\frac{[c_f + c_d \times (t_f)]\left[1 - R(t+\Delta|\vartheta^t)\right]}{\tau(t+\Delta|\vartheta^t) + t_f\left[1 - R(t+\Delta|\vartheta^t)\right]} \geq g(\gamma)$, but $CI(\vartheta^t)$ is still less than $g(\gamma)$. Ignoring $\varpi(\vartheta^t, \gamma)$ in such cases may result in early replacement. As stated in [147], it is very hard to find a practical decision rule, which is optimal in a general case of hazard function (e.g. non-monotonic hazard function). Two possible approaches can be employed to deal with non-monotonic hazard functions. If the hazard function of the device over time is expected to be non-decreasing, but noisy data may lead to a non-monotonic hazard function, then the multiple points control policy discussed in Section 6.3.7 can be used as an option to decrease unnecessary maintenance actions. However, if the hazard function of the device can theoretically be non-monotonic, then $\varpi(\vartheta^t, \gamma)$ needs to be carefully evaluated to avoid unnecessary maintenance actions. In the remainder of this appendix, an approach to deal with such cases is given.

At time t, the expected value of $V(\vartheta^{t+\Delta}, \gamma)$, given that the device survives the next observation interval, can be calculated as:

$$E(V(\vartheta^{t+\Delta}, \gamma)) = \left[\frac{\sum\limits_{\vartheta^{t+\Delta} \neq F} \Pr(\vartheta^t, \vartheta^{t+\Delta}, 0) \times V(\vartheta^{t+\Delta}, \gamma)}{R(t+\Delta|\vartheta^t)}\right]. \tag{A.6}$$

From the above equation and the definition of $\varpi(\vartheta^t, \gamma)$, we have:

$$\varpi(\vartheta^t, \gamma) = R(t+\Delta|\vartheta^t) \times \left(E(V(\vartheta^{t+\Delta}, \gamma)) - V^1(\vartheta^{t+\Delta}, \gamma)\right). \tag{A.7}$$

Now, let $E(\Pr(L < T_\gamma | L > t + \Delta, \vartheta^t))$ be the expected probability of a failure replacement, given that the device has not failed up to time $(t+\Delta)$ and condition ϑ^t is known. Also, let $E(\min(L, T_\gamma)|L > t+\Delta, \vartheta^t)$ be the expected time to replacement, given that the device has not failed up to time $(t + \Delta)$ and condition ϑ^t is known. Now, we have:

$$E(V(\vartheta^{t+\Delta}, \gamma)) - V^1(\vartheta^{t+\Delta}, \gamma) = \\ \min\{0, (c_f + c_d \times t_f) \times E(\Pr(L < T_\gamma | L > t + \Delta, \vartheta^t)) - \\ g(\gamma) \times \big(E(\min(L, T_\gamma)|L > t + \Delta, \vartheta^t) + t_f \times E(\Pr(L < T_\gamma | L > t + \Delta, \vartheta^t))\big)\}. \quad \text{(A.8)}$$

Eq. (A-8) presents the expected cost-saving of do nothing with respect to replace immediately, given that the device has not failed up to time $(t+\Delta)$ and condition ϑ^t is known. The last remaining steps to find $\varpi(\vartheta^t, \gamma)$ are to estimate $E(\Pr(L < T_\gamma | L > t + \Delta, \vartheta^t))$ and $E(\min(L, T_\gamma)|L > t + \Delta, \vartheta^t)$. Let us define $a(\vartheta^t, \gamma)$ as the time point where the value of control index will be equal to the average cost of replacement policy γ, given that the current condition of the device is ϑ^t. The time point $a(\vartheta^t, \gamma)$ can be estimated by solving the equation $CI(x|\vartheta^t) = g(\gamma)$, where $CI(x|\vartheta^t)$ is the expected value of the control index at time x, given that the current condition of the device is ϑ^t, the cost of the policy γ is $g(\gamma)$, and the device has survived until time point x. The element $CI(x|\vartheta^t)$ can be calculated using Eq. (A.4) by adding the condition of $L > x$ to $R(x + \Delta|\vartheta^t)$, $\tau(x + \Delta|\vartheta^t)$, and $\varpi(\vartheta^t, \gamma)$. The final step is to estimate $E(\Pr(L < T_\gamma | L > t + \Delta, \vartheta^t))$ and $E(\min(L, T_\gamma)|L > t + \Delta, \vartheta^t)$.

Let k be such an integer, that $(k - 1)\Delta \leq a(\vartheta^t, \gamma) < k\Delta$. Now, similar to the backward recursion approach given in [159], $E(\Pr(L < T_\gamma | L > t, \vartheta^t))$ and $E(\min(L, T_\gamma)|L > t, \vartheta^t)$ can be obtained by the following equations:

$$E\left(\Pr(L < T_\gamma | L > t, \vartheta^t)\right) = \\ \begin{cases} 0 & t \geq k\Delta \\ 1 - R(a(\vartheta^t, \gamma)|\vartheta^t) & t = (k-1)\Delta \\ (1 - R(t + \Delta|\vartheta^t)) + \\ \quad \sum\limits_{\vartheta^{t+\Delta} \neq F} \Pr(\vartheta^t, \vartheta^{t+\Delta}, 0) \times E(\Pr(L < T_\gamma | L > t + \Delta, \vartheta^{t+\Delta})) & t < (k-1)\Delta \end{cases}, \quad \text{(A.9)}$$

$$E\left(\min(L,T_\gamma)|L>t,\vartheta^t\right)=$$
$$\begin{cases} 0 & t\geq k\Delta \\ \int\limits_{t}^{a(\vartheta^t,\gamma)} R(\tau|\vartheta^t)d\tau & t=(k-1)\Delta \\ \int\limits_{0}^{\Delta} R(t+\tau|\vartheta^t)d\tau + & \\ \sum\limits_{\vartheta^{t+\Delta}\neq F} \Pr(\vartheta^{\vartheta^t,t+\Delta},0)\times E(\min(L,T_\gamma)|L>t+\Delta,\vartheta^{t+\Delta}) & t<(k-1)\Delta \end{cases}. \tag{A.10}$$

Unfortunately, directly solving Eqs. (A.9)-(A.10) is computationally very expensive because it requires enumerating over all possible combinations of future conditions. In the above equations, $\sum\limits_{\vartheta^{t+\Delta}} \Pr(\vartheta^t,\vartheta^{t+\Delta},0)\times E(\Pr(L<T_\gamma|L>t+\Delta,\vartheta^{t+\Delta}))$ can be estimated by $R(t+\Delta|\vartheta^t)\times E(\Pr(L<T_\gamma|L>t+\Delta,\vartheta^t))$ and $\sum\limits_{\vartheta^{t+\Delta}} \Pr(\vartheta^{t+\Delta},\vartheta^t,0)\times E(\min(L,T_\gamma)|L>t+\Delta,\vartheta^{t+\Delta})$ can be estimated by $R(t+\Delta|\vartheta^t)\times E(\min(L,T_\gamma)|L>t+\Delta,\vartheta^t)$. Now, for any $t\leq x$, we have:

$$E\left(\Pr(L<T_\gamma|L>x,\vartheta^t)\right)=$$
$$\begin{cases} 0 & x\geq k\Delta \\ 1-R(a(\vartheta^t,\gamma)|\vartheta^t,L>x) & x=(k-1)\Delta \\ (1-R(x+\Delta|\vartheta^t,L>x))+ & \\ R(x+\Delta|\vartheta^t,L>x)\times E(\Pr(L<T_\gamma|L>x+\Delta,\vartheta^t)) & x<(k-1)\Delta \end{cases}, \tag{A.11}$$

$$E\left(\min(L,T_\gamma)|L>x,\vartheta^t\right)=$$
$$\begin{cases} 0 & x\geq k\Delta \\ \int\limits_{t}^{a(\vartheta^t,\gamma)} R(\tau|\vartheta^t,L>x)d\tau & x=(k-1)\Delta \\ \int\limits_{0}^{\Delta} R(x+\tau|\vartheta^t,L>x)d\tau + & \\ R(x+\Delta|\vartheta^t,L>x)\times E(\min(L,T_\gamma)|L>x+\Delta,\vartheta^t) & x<(k-1)\Delta \end{cases}. \tag{A.12}$$

Now, Eqs. (A.11)-(A.12) can be recursively (backward) employed to estimate $E(\Pr(L<T_\gamma|L>t,\vartheta^t))$, $E(\Pr(L<T_\gamma|L>t+\Delta,\vartheta^t))$, $E(\min(L,T_\gamma)|L>t,\vartheta^t)$, and $E(\min(L,T_\gamma)|L>t+\Delta,\vartheta^t)$ for $t\geq 0$. Applying the above described approach to deal with non-monotonic hazard function is subject to the error of estimation.

Therefore, a more efficient method which can deal with non-monotonic hazard functions needs to be developed in future work.