University of Texas at El Paso

ScholarWorks@UTEP

2020-01-01

Multiple Change-point Modeling of Degradation Signal for Prognostics Improvement

Yuxin Wen
University of Texas at El Paso

MULTIPLE CHANGE-POINT MODELING OF DEGRADATION SIGNAL FOR

PROGNOSTICS IMPROVEMENT

YUXIN WEN

Doctoral Program in Electrical and Computer Engineering

APPROVED:

_____

Tzu-Liang (Bill) Tseng, Ph.D., Chair

_____

Jianguo Wu, Ph.D., Co-Chair

_____

Sergio Cabrera, Ph.D.

_____

Paras Mandal, Ph.D.

_____

Naijun Sha, Ph.D.

_____

Stephen L. Crites, Jr., Ph.D.
Dean of the Graduate School

MULTIPLE CHANGE-POINT MODELING OF DEGRADATION SIGNAL FOR

PROGNOSTICS IMPROVEMENT

by

YUXIN WEN, M.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Department of Electrical and Computer Engineering

THE UNIVERSITY OF TEXAS AT EL PASO

August 2020

Acknowledgements

The research presented in this dissertation benefited from the valuable insights and support of many people. It is my pleasure to express my sincere gratitude to all of them.

First and foremost, I would like to acknowledge my indebtedness and render my warmest thanks to my supervisor Prof. Bill Tseng for his guidance, consistent support, and encouragement throughout my doctoral study. I have been very lucky and honored to be supervised by him and have learned tremendously from him. The invaluable experience working with him will benefit me throughout my entire life.

I owe my deepest gratitude to my co-advisor, Prof. Jianguo Wu, whose unwavering enthusiasm keeps me constantly engaged with my research. His guidance brought a depth of knowledge to this dissertation, and more importantly, his dedication and encouragement over the past four years inspired me to work hard and aim to achieve my full potential.

In addition, I am extremely grateful to my doctoral committee members Prof. Sergio Cabrera, Prof. Paras Mandal and Prof. Naijun Sha for serving on my dissertation committee and for their constructive comments to improve this dissertation. Especially I would like to give my sincere gratitude to Prof. Paras Mandal for his full support in my research, job and scholarship applications.

I also would like to thank Prof. Miguel Velez-Reyes for giving me continued opportunity and support to succeed in the Electrical and Computer Engineering PhD program. My grateful thanks are also extended to all the other faculties and staff members in the Department of Electrical and Computer Engineering, Department of Industrial, Manufacturing and Systems Engineering who have taught me and helped me throughout my PhD studies.

My thanks also go to my friends in my research group: Dr. Zhonghua Hu, Dr. Satya Akundi, Dr. Hoejin Kim, Dr. Carlos A Garcia Rosales, Dr. Ivan Renteria Marquez, Honglun Xu, Md Fashiar Rahman, Anabel Renteria Marquez, Jeevarathinam Senthilkumar and Chien-Chun Chou. Thank you for being supportive and accompanying me in my PhD studies. You really make my life in UTEP enjoyable and memorable.

Last but not least, I would like to thank my parents and parents-in-law for their endless support and encouragement. Without their support, I would not have made this achievement in my life. I also want to give my special thanks to my sister for standing by me during these years. I owe tremendously to my husband Jinliang Wang, who has given me unconditional and unlimited love and support. To my precious little boy, Linus, thanks for stepping in my life. Your smiles are the sharpest sword to conquer any predicaments.

Abstract

To fulfill the increasing demand on functionality and quality, modern engineering systems are usually built with overwhelming complexities. The more complex functions the system has been built, the higher reliability required of the system. This is mainly due to the fact that a single failure can result in catastrophic consequences. Therefore, methods that can predict and prevent such catastrophes have long been explored. Prognostics refer to the process of evaluating the current health of a system or a component and then predicting the remaining useful life (RUL) based on the information collected through condition monitoring. The fast development of information and sensing technologies offer great opportunities for real-time health condition monitoring, ensuring the safety, availability, and efficiency of various engineering systems. The condition monitoring signals, collected from sensors, also called degradation signals, are commonly used for system reliability assessment due to their direct relation with underlying physical degradation processes. The commonly applied statistical approach for RUL prediction is to fit degradation signals using parametric regression models to describe and predict how the currently available degradation signal evolves. However, these parametric models are often too rigid and not adequate or flexible enough to model the real degradation signals during the whole life cycle. In fact, degradation signals often show multiple phases in many applications, where the conventional parametric degradation models are often inadequate.

Motivated by the issues, a novel Bayesian multiple change-point modeling approach to characterize degradation signals for prognostics is proposed. Under the Bayesian framework, two stages are often required for prognostics: the offline modeling of historical degradation signals, and the online Bayesian individual model updating and RUL prediction of a new unit. To

characterize the inherent unit-to-unit heterogeneity and make the model more flexible, In this dissertation, all the model parameters are assumed to be random in the model, including the number of change-points, their locations, and all model parameters of each segment. This assumption brings several challenges on how to effectively apply the multiple change-point model and RUL prediction. To address these challenges, we propose a series of approaches in both offline modeling and online model updating and RUL prediction. The main contributions of this research include: (1) An innovative stratified particle filtering algorithm with partial Gibbs resample-move strategy is developed to improve modeling and prognostics. To improve the prediction accuracy, the priors are specified with a novel stochastic process and the multiple change-point model is formulated to a novel state-space model to facilitate online monitoring and prediction; (2) To reduce the model complexity, an exact Bayesian inference is developed. where the closed form of all posterior distributions can be sequentially obtained at online stage. To further control the computational cost, a fixed-support-size strategy in the online model updating and a partial Monte Carlo strategy in the RUL prediction are proposed; (3) To better capture the temporal uncertainties that are inherent in the degradation process, a multiple change-point Wiener process modeling is proposed. The advantages and effectiveness of the proposed methods have been demonstrated through extensive numerical studies and real-world case studies.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 BACKGROUND

Reliability has always been an important aspect in the assessment of industrial products or engineering systems. Due to the increasing complexity of modern engineering systems, the reliability demand for expensive and critical products becomes more and more essential and crucial [1-3]. To ensure safe and reliable operation, systems must be periodically or even continuously monitored since even a single failure may result in catastrophic consequences. Therefore, effective maintenance strategies that has the capability of predicting the future states of systems and making decisions wisely are required to enhance the reliability of systems. Conventional researchers focus on using failure time data to estimate lifetime distribution. Many failure models, such as Poisson, Exponential, Weibull, and Lognormal distributions have been used to model machine reliability [4-6]. This conventional class of algorithms only estimate the distribution for the entire population of identical units, but usually they cannot provide failure prediction for an individual unit. These reliability approaches are more suitable for system design optimization in terms of reliability performance. In practical applications, however, as components become more reliable, few failure times may be available. On the other hand, due to the variability in the inner structures or the diversity in working environment, different components exhibit various degrading paths [7], prognosis on individual units becomes more and more important.

Recent advances in information and sensing technologies have provided a unique opportunity to monitor systems or components constantly without interrupting normal machine operation, which makes the maintenance strategy has evolved from traditional breakdown maintenance, preventive maintenance to condition-based maintenance (CBM) [8]. Condition-

based maintenance involves real-time analysis of sensor data to infer maintenance condition, or health. In comparison with traditional maintenance strategies, CBM can significantly reduce maintenance costs, improve replacement strategies and health management, as well as maximizing manufactured products availability [9, 10]. Diagnostics and prognostics are two important aspects for CBM. Diagnostics deals with fault detection, isolation and identification when it occurs, while prognostics deals with fault prediction before it occurs. Since prognostics is associated with predicting the future, which involves a large degree of uncertainty. Therefore, prognostics is much more challenging than the diagnostics. The most widely used prognostics is to predict how much time is left before a failure occurs given the current machine condition and past operation profile [8]. The "machine condition" collected from sensors are known as condition monitoring (CM) signals or degradation signals, they record machinery degradation process that deteriorates over time and eventually leads to a system/component failure. With this context, a failure is defined to be the time point when the magnitude first reaches a pre-specified threshold, which is also called soft failure determined by either engineering domain knowledge or accepted industrial standards [1, 11]. Some examples of degradation signals include vibration signals for monitoring excessive wear in rotating machinery, acoustic emissions for monitoring crack propagation, temperature changes and oil debris for monitoring engine lubrication, crack size changes in metallic structures, and many others. Figure 1-1 shows an example of three bearing degradation signals. Bearings are critical part of mechanical systems with rotational components. As it deteriorates over time, they tend to exhibit larger vibration. When the amplitude of a bearing's vibration exceeds the certain threshold (black dotted line), the bearing can be considered as no longer suitable for further operations [12].

2
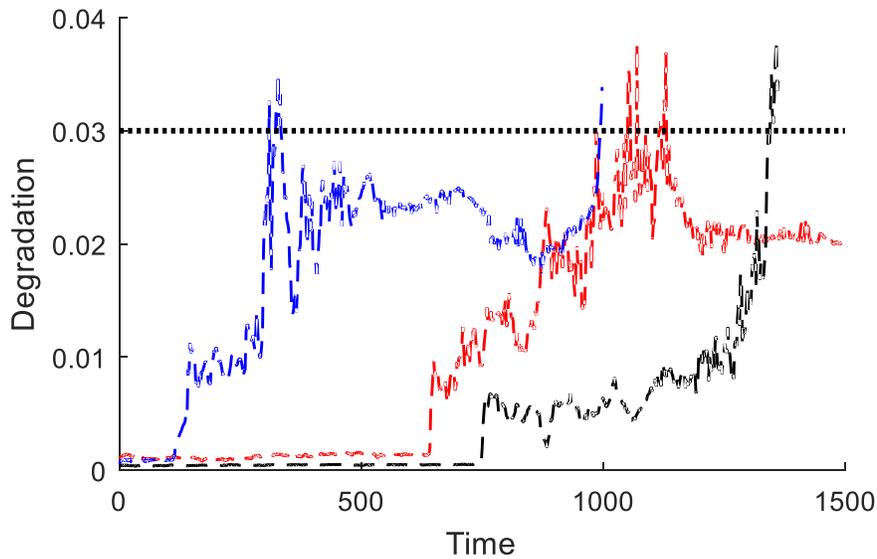
Figure 1-1: Illustration of degradation evolution

If we could build a model that is capable of describing the degradation phenomenon based on the current acquired condition monitoring data [13], we will be able to predict how much time is left before a failure occurs. The time left before observing a failure is usually called remaining useful life (RUL), that is, $T - t | T > t, X(t)$, where T is the failure time and t is the current time, $X(t)$ is the historical condition monitoring data. Due to randomness in the evolution paths of the degradation, the calculated RUL will be in the form of some probability distribution. Considerable attention has been focused on developing degradation models to improve the capability and accuracy of RUL prediction techniques in the past few years. The objective of this dissertation is to predict the RUL of an individual unit based on degradation modeling technologies. In the following sections, a literature review of the development of RUL prediction based on degradation modeling technologies is presented. Some issues involved in the current researches are addressed.

3

1.2 LITERATURE REVIEW

The prediction of RUL often requires a prognostic model, most of existing prognostics models can be categorized broadly into physics-based models, data-driven models [8, 10, 14, 15]. The physics-based degradation models are those which the degradation phenomenon can be described by accurate mathematical models such as the Paris' law, Forman law, Arrhenius law and corrosion initiation equation, or others. To establish these models, a thorough knowledge about the system mechanism, operating conditions, and life cycle loads applied to the system are required [2]. The main advantage of physics-based method is the ability to incorporate physical understanding of the system for prognostics. Another advantage is that, in many situations, the changes in degradation features are closely related to model parameters, an accurate functional mapping can be established between parameters and prognostic features [16]. A number of applications of physics-based models can be found in automotive, aerospace and defense industries. Li and Lee [17] used Paris-Erdogan law to model spur gear crack propagation and then to predict the RUL of a cracked gear. The Paris-Erdogan law also has been used by Kacprzynski et al. [18] for a helicopter crack propagation prognosis. Luo et al.[19] employed an Interacting Multiple Model (IMM) to track the degrading parameters and to estimate the remaining life of an automotive suspension system. Oppenheimer and Loparo [20] applied a physical model to predict the remaining useful life using linear elastic fracture mechanics for shaft cracking in a rotor. However, the development of physics-based models requires a sufficient understanding of the underlying physical processes that lead to system failure. Reaching to a perfect physical model that representing the multiple physical processes occurring for complex systems is very difficult [21]. The worst thing is, they are very expensive to implement, even small modification in the material or subcomponent of the system will require

remodeling [22]. Considering all of these reasons, data-driven approaches are preferable in most of applications where CM data can be easily obtained by sensors.
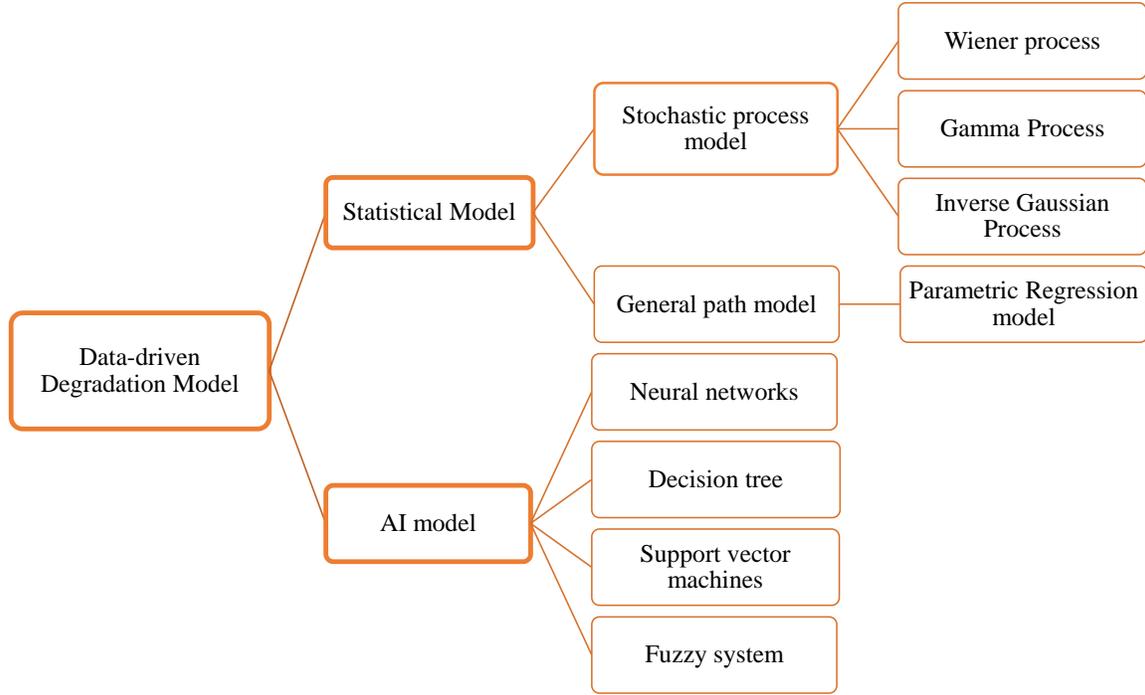
Figure 1-2: Categories of Data-driven methods

Data-driven methods directly utilize the collected CM data for health condition prediction, and do not require physics-of-failure models. According to [23-25], the existing scientific literature on data-driven modeling and prediction can be further classified into two categories as shown in Figure 1-2:

(a) Artificial intelligence (AI) approaches. The AI approaches typically include (1) *neural networks*. The feature of neural networks that they do not rely on priori principles or statistics models and can significantly simplify the model synthesized process. Therefore, neural networks are easy to implement accurate and fast on-line pattern recognition [26]. Lee [27] developed a neural network model based on cerebellar model articulation controllers (CMACs) in order to discriminate and quantify machine degradation. Gebraeel et al. [28, 29] developed ball bearing remaining life prediction methods based on feedforward neural networks. Dong et al. [30]

utilized a grey model and a back-propagation (BP) neural network to predict machine condition. Wang and Vachtsevanos [31] applied dynamic wavelet neural networks to predict the fault propagation process and estimate the RUL. Recently a recurrent neural network based on health indicator for RUL prediction of bearings was proposed by Guo et al. [32]; (2) *fuzzy logic*. Fussy logic provides a very human-like and intuitive way of representing and reasoning with incomplete and inaccurate information by using linguistic variables. A fuzzy-based data-driven similarity analysis is performed by Zio and Maio for predicting the remaining life of a newly developing failure trajectory [23]; (3) *expert system*. The process of building expert systems involve knowledge acquisition, knowledge representation, and the verification and validation of models. Butler proposed an expert system-based framework for incipient failure detection and predictive maintenance (FDPM) [33]. Except researches as mentioned above, some efforts have also been done to build combination models through combine two or more techniques together. Wang et al. [34] compared the results of applying recurrent neural networks and neural–fuzzy combination inference systems to predict the fault damage propagation trend. Xue et al. [35] developed a fuzzy mathematical model with radial basis function neural network to predict the potential faults of a coal-fired boiler. Baraldi et al. [36] recently predicted the RUL of ferritic steels by fuzzy similarity and belief function theory.

(b) Statistical approaches. The statistical approaches include general path models, stochastic processes (e.g., Wiener process, Gamma process, Inverse Gaussian process) and Markovian based models, etc.

General path model. The basic idea of general path model is to build a parametric evolution path to trend degradation signals. It assumes an underlying functional form for the degradation path of a specific unit as following

$$X_{ij}(t) = g(t_{ij}; \boldsymbol{\phi}, \boldsymbol{\theta}) + \varepsilon_{ij} \tag{1.1}$$

Where the fixed coefficients $\boldsymbol{\phi}$ are common for all units and random coefficients $\boldsymbol{\theta}$ are used to representing individual unit characteristics, ε_{ij} are normally distributed measurement errors with $\varepsilon_{ij} \sim N(0, \sigma^2)$. Parameters in the model should be estimated before we apply the model for decision-making and prognostics. In this model, the functional forms of $g(t_{ij}; \boldsymbol{\phi}, \boldsymbol{\theta})$ could be linear, polynomial, exponential, or combination of them. Tseng et al. [37] developed a simple linear random coefficients model to fit the luminosity degradation path of fluorescent lamps. Lu and Meeker [38] presented a general nonlinear regression model to characterize the degradation path of fatigue-crack-growth, A two-stage method was presented for estimating the model parameters. Chinnam [39] proposed a polynomial regression model and putted the model in thrust-force signal of drill-bits and fatigue crack growth signal. The result showed that a third-order model provided a better fit for modeling. Chakraborty et al. [40] developed a simulation-based algorithm and linear degradation model with gamma distribution as a prior. They concluded that using a gamma distribution which is the fact that its two parameters provides a better flexibility.

Stochastic process based models. In the stochastic process based model, for any time t, and $\Delta t > 0$, the increments $\Delta X(t) = X(t + \Delta t) - X(t)$ of degradation signal $X(t)$ in disjoint time intervals are independent [3]. If the increment is normally distributed, then $\{X(t)\}$ is the Wiener process. A basic Wiener process can be expressed as

$$X(t) = \lambda t + \sigma B(t) \tag{1.2}$$

Where λ is a drift parameter reflecting the rate of degradation, σ is a diffusion coefficient, $B(t)$ represents the standard Brownian motion. One of the main advantages of degradation modeling with Wiener processes is that the distribution of the failure time can be formulated

7

analytically by the first passage time (FPT), in which its PDF is an inverse Gaussian distribution. Due to its mathematical properties and physical interpretations, extended versions have also been made based on basic Wiener process to satisfy specific demand. One alternative is to add an error term onto the basic process to capture measurement errors in degradation signals [41-43]; the second way is to incorporate random-effects model in dealing with unobserved heterogeneities, specifically, assume that λ or σ or both follow some certain parametric distributions, see examples [44-46] among others. The third way is to incorporate nonlinear structure into this model to make the model more general. In particular, the more generalized model is defined as

$$X(t) = \lambda\Lambda(t; \boldsymbol{\theta}) + \sigma B(\tau(t; \boldsymbol{\gamma})) \tag{1.3}$$

Where $\Lambda(t; \boldsymbol{\theta})$ and $\tau(t; \boldsymbol{\gamma})$ are non-decreasing functions with parameter vectors of $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ [43, 47]. Wiener processes have been used extensively to model a variety of degradation processes encountered in real systems, such as bridge beams [46], fatigue crack dynamics [48], light-emitting diodes [44] and micro electro mechanical systems (MEMS) [49]. If the increment $\Delta X(t)$ follows Gamma distribution or inverse Gaussian distribution, the process $\{X(t)\}$ will be called Gamma process and inverse Gaussian process, respectively. Details of modeling degradation with these two processes are given by [50-52], The recent literature on stochastic modeling for high reliable products can be found in Ye and Xie [11].

Markovian Process based models. Note that although Markovian process still belongs to stochastic process, we list these models separately considering the big difference from the previously mentioned stochastic models. The rationale of Markovian process is that the future states of the process depends only on the present state, not the past states. Markovian process based models in degradation analysis can be categorized as Markov Chain model [53], Semi-

Markov model [54], Hidden Markov Model (HMM) [55], Hidden semi-Markov model (HSMM) [56, 57]. Semi-Markov processes are a generalization of Markov. An HMM consists of two stochastic processes: a Markov chain with finite number of states describing an underlying mechanism and an observation process depending on the hidden state. An HSMM is constructed by adding a temporal component into the well-defined hidden Markov-model (HMM) structures to cope with the inaccurate state duration modeling of HMM.

1.3 RESEARCH OBJECTIVES AND CHALLENGES

Selecting an accurate and effective prognostic model is the key for the prognostics. Due to the simplicity and flexibility, general path models are commonly used in industry and academic fields. Although many efforts have been done on general path modeling in the past few years, there still remain some issues to be improved for the accuracy and efficiency. Here, to be specific, some CM signals show two or more distinct phases in practice. For example, Bae and Kvam [58] demonstrated that the degradation path of vacuum fluorescent displays is not monotonic, showing obviously two phases or even three phases. Son et al. [59] also showed that the internal resistance degradation signal of vehicle batteries evolves more rapidly after the system has degraded down to a certain change point before failure occurs. This phenomenon has also been observed in many other CM signals, such as high-performance capacitors [60], the semiconductor laser diodes [61], the liquid coupling devices [62] and vibrational signals of rotational bearings [63, 64]. In this regard, general path models are often too rigid and not adequate or flexible enough to model the real CM signals in the whole time period. Some researchers [37] [65] chose to delete early degradation measurements at the first stage, under the assumption that the failure will not occur at the early stage, and then apply the parametric models to the second phase data for better model fitting and prediction. However, the truncated

9

measurements may contain valuable information about the degradation process or the prediction may need to be made at the early stage.

To avoid measurement truncation, change point models are usually used to deal with the system models that has changes suddenly during the whole life time period. Given a set of observations, Change point models separate observations into two or more segments assumed that the observations follow the same statistical model within each segment, but different models in different segments [66, 67]. Change point models have been widely spread to various areas, such as in finance, biology science, software reliability analysis. For example, Change-point detection techniques are used to detect changes in temperature and in precipitation [68], to detect abrupt shifts in carbon dioxide (CO_2) concentrations [69], to detect DNA segmentation [67]. Until recently, limited efforts devoted to degradation modeling with change-points. Li and Nilkitsaranont [70] employed a combination of a linear model in the first phase and a quadratic model in the second phase to estimate the remaining useful life of gas turbine engines, and used "compatibility check" to determine the transition point from one model to another. Son *et al.* [59] incorporated a change-point to the resistance signal in a joint prognostic model (joint modeling of reliability data and CM data) to predict the RUL of batteries. Bae and Kvam [71, 72] found that the prediction accuracy can be improved substantially by constructing a two-phase degradation process under general path models to account for burn-in effects of light displays. Gebraeel et al. [12] developed an exponential (i.e., log-linear) degradation model with a pre-set location of a change-point to illustrate the updating process of rolling element bearings, they assumed that the bearing has already entered the second phase and omitted the influence of the first phase. Later, Chen and Tsui [63] revisited Gebraeel's work, and applied a two-phase piecewise regression model with one change-point at unknown location to characterize both

10

phases of the bearing degradation signals. Then Wang and Tsui [73] optimized the piecewise model [63] to a unique model using an indicator function and extended Gebraeel's work to be more general. Wang et al. [62] proposed a two-stage stochastic process model, called as change-point gamma and Wiener process, to capture a degradation process of liquid coupling devices (LCD). All these aforementioned methods assume a two-phase pattern on CM signals. It means that add a change-point to divide CM signals into two phases and fit each phase with different models. In many situations, however, the degradation path may have three or even more phases during the whole life cycle. It would be difficult to select proper functional forms to characterize the degradation behavior with no change-point or only one change-point.

To fill such gap, this dissertation endeavors to develop a novel multiple change-point modeling approach for condition monitoring and RUL prediction. To the best of our knowledge, very few efforts have been made to address the degradation analysis with multiple change points. Note that, some authors also refer multiple change points as "multi-phase" or "multi-stage". In the existing literature, the definition of "phase" (or "stage") can be classified into two categories: (1) it is commonly referred to as different operational conditions or states, such as the working state and storage state for missiles, multiple consecutive phases of operations required to finish the service for phased-mission systems (PMS). For example, Si et al. [74] estimated the real-time reliability of an individual phased-mission system (PMS) by using a stochastic filtering model to model the phase duration, and Brownian motion was employed to account for the dependency of mission phase-dependency. Si et al. [75] predicted the residual storage life of critical systems by applying a two-state continuous-time homogeneous Markov process to approximate the switches between the working state and storage state. (2) It is also frequently used to denote health conditions with different characteristics (e.g., normal working stage and irreversible degradation

11

stage with defects occurred for bearings), or different patterns shown on CM signals, which may not have specific physical meanings. For example, Feng et al. [60] proposed a multi-phase Wiener process model to predict the storage life of high-voltage-pulse capacitors. In our approach, "multiple-phase" is more related to the second category, though it can be easily applied to the first case as long as the degradation signals exist multiple patterns. The main difference between our work and the existing multiple change-point approaches in category (1) lies in the motivation and methodologies. In [74, 75], as mentioned earlier, in these methods, the multiple phases are used to model different operational states or stages, e.g., take-off, ascent, cruise, approach and landing phases of the on-board systems for the aided-guide of aircraft. Therefore, the number of phases are fixed, the phase index and the starting point of each phase before the current time can be exactly observed. In Feng's work [60] in category (2), the number of change-points and their locations are also deterministic and the same for all units, which is not realistic for real degradation signals with unit heterogeneity. The possible state changes in future time period are not considered in the life prediction. Comparing with those existing work, we innovatively apply the multiple change-point model to degradation signals to improve modeling and prognostics, which is fundamentally different from the existing multiple-phase modeling approaches in terms of motivation, methodology and applications. In this research, a piece-wise linear functional form is proposed to model CM signals. The rationale is the piecewise linear model with a proper number of change-points at proper locations is flexible enough to capture the non-linear and multiple-phase characteristics of various kinds of degradation signals in application. It could avoid the nontrivial selection of appropriate functional forms to model the CM signals. Besides, it makes more sense physically to define it as a phase when CM signal is degrading with a constant rate. To characterize the inherent unit-to-unit heterogeneity and make

the model more flexible, in our model all the model parameters are assumed to be random, including the number of change-points, their locations, and all model parameters of each segment. The Bayesian approach is naturally selected to incorporate the historical data and current information for model updating, prediction and uncertainty quantification. Under Bayesian framework, two stages are often required for real-time prognostics: i.e., offline and online stages. At the offline stage, all parameters need to be estimated based on historical CM data. At the online stage, the newly observed data from a specific unit will be merged with the offline information by using Bayesian inference for model updating and RUL prediction. There are several challenges on how to effectively apply the multiple change-point model for condition monitoring and RUL prediction under the Bayesian framework:

(1) Due to significantly increased dimensionality and complexity, how to specify reasonable priors (e.g., phase durations or number of change-points, model parameters of each phase) based on historical CM signals at the offline modeling stage.

(2) How to update the posterior distributions of model parameters sequentially for an in-service unit at the online model updating stage.

(3) How to determine the future possible change points, so as to incorporate the future possible change points into the RUL prediction, Since the degradation rate depends on the phase the degradation process at, the accuracy of the change-points' estimate has a direct influence on the accuracy of the RUL estimation.

Motivated by the aforementioned challenges of multiple change-point modeling. The goal of this dissertation is to develop a robust and efficient method for real-time RUL prediction. To achieve this goal, a series of approaches are developed to address these issues. (1) a simple yet effective two-stage empirical approach is proposed for offline model estimation. (2) The multiple

13

change-point model is reformulated as a non-standard state-space model, then a novel particle filtering (PF) algorithm is developed for online model updating and RUL prediction. We call the proposed method as stratified particle filtering (SPF) algorithm. (3) we develop an innovative recursive updating algorithm by using exact Bayesian inference. The closed form of all posterior distributions can be sequentially obtained by using conjugate priors. We also derive a closed form of the RUL prediction. To control the computational cost in the online stage, then a fixed-support-size strategy and a partial Monte Carlo (partial-MC) approach are proposed to control the computational cost of RUL estimation. (4) we propose a multiple change-point Wiener process degradation model to capture the temporal uncertainties that are inherent in the degradation process.

1.4 OUTLINE OF THE DISSERTATION

The remainder of the dissertation is organized as follows.

In chapter 2, a stratified Particle Filtering algorithm is described for RUL prediction. The detail of a multiple change-point general path model for the CM signals is presented. The technical details on how to sequentially update the posterior distributions of the phase index, latest change-point, and model parameters of the current phase, and how to predict the RUL under the Bayesian framework using the particle filtering algorithm are presented.

In chapter 3, a multiple change-point degradation modeling combined with exact Bayesian inference framework is developed, an innovative recursive updating and prognostic algorithm where the closed form of all posterior distributions and RUL distribution of derivation is presented. Then a fixed-support-size strategy and partial Monte Carlo (partial-MC) approach are proposed to control the computational cost of RUL estimation.

In chapter 4, to better capture the temporal uncertainties that are inherent in the degradation process, a multiple change-point Wiener process degradation model is proposed. In this model, at the offline stage, an empirical two-stage process is proposed for model estimation, and a cross-validation approach is adopted for model selection. At the online stage, an exact recursive model updating algorithm is developed for online individual model estimation, and an effective Monte Carlo simulation approach is proposed for RUL prediction.

The conclusions and future work are given in chapter 5.

Chapter 2: Stratified Particle Filtering Algorithm for RUL Prediction

2.1 INTRODUCTION

As discussed in chapter 1, to characterize the population-level trend as well as the individual heterogeneity, mixed-effects or random-effects models are most commonly selected in off-line modeling of historic CM signals. We develop multiple change-point based general path model. To predict the RUL for a new unit, the Bayesian approach is naturally selected for online model updating, prediction and uncertainty quantification, where the fitted parameters in the offline stage are used as priors. However, there are several challenges on how to effectively apply the multiple change-point model for condition monitoring and RUL prediction under the Bayesian framework. Due to significantly increased dimensionality and complexity, it is difficult to specify reasonable priors (e.g., phase durations or number of change-points, model parameters of each phase) in the offline modeling of historical CM signals. In addition, in the online model updating stage, the posterior distributions of model parameters need to be updated sequentially. However, the multiple change-point model is highly nonlinear and the conventional Kalman filtering techniques, which are commonly used for linear models, are not applicable. Besides, the RUL prediction given the posterior of the current model parameters is still very challenging due to the uncertainty of future change-points and model parameters. To address these challenges, in this chapter, we propose a series of approaches in both off-line modeling and online model updating and RUL prediction. In the off-line modeling, a novel stochastic process is proposed to specify and estimate priors in the off-line modeling. In the online stage, the multiple change-point model is formulated as a non-standard state-space model and a novel stratified particle filtering algorithm is developed for online model updating and RUL prediction.

The contribution of this chapter lies in the following three-fold: (1) we innovatively apply the multiple change-point model to degradation signals to improve modeling and prognostics, which is fundamentally different from the existing multiple-phase modeling approaches in terms of motivation, methodology and applications; (2) a full Bayesian framework is proposed for the multiple change-point model through a novel stochastic process; and (3) an efficient stratified particle filtering algorithm with partial Gibbs sampling strategy is developed for model updating and RUL prediction.

The rest of this chapter is organized as follows. In Section 2.2, a multiple change-point based general path model for the CM signals is presented. The prior parameters specification and state-space representation for the multiple change-point model are given in Section 2.3. Section 2.4 presents the technical details on how to sequentially update the posterior distributions of the phase index, latest change-point, and model parameters of the current phase, and how to predict the RUL using the particle filtering algorithm. Section 2.5 demonstrates the effectiveness and accuracy of the proposed method through numerical and case studies. The conclusion is given in Section 2.6.

2.2 MULTIPLE CHANGE-POINT GENERAL PATH MODELING

A degradation model that can adequately describe the degradation path is essential for prognostics. In this chapter, a multiple change-point based general path model is proposed to model CM signals. Specifically, we split the degradation signal into multiple phases using multiple change-points. Then each segment can be described by linear regression model. The piecewise linear model with a proper number of change-points at proper locations is flexible enough to capture the non-linear and multiple-phase characteristics of various kinds of degradation signals in application. It could avoid the nontrivial selection of appropriate

17

functional forms to model the CM signals. Here we use the bearing vibration signals [12] to demonstrate the superiority of the proposed method. Figure 2-1 shows an example of thrust bearing vibration signals with multiple-phase characteristic. Obviously, the bearing operates under a stable condition at first and then degrades rapidly with two distinct phases. The degradation signal could be appropriately modelled with three line segments. If only one change-point is incorporated, the degradation signal after the stable stage is poorly fitted (Figure 2-1a), which could consequently influence the prognostic accuracy (Figure 2-1c).
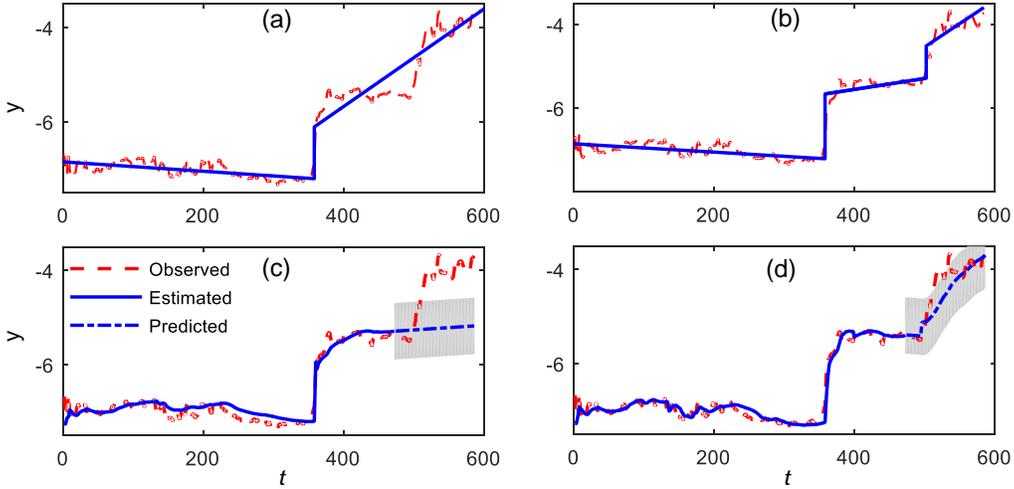
Figure 2-1: Modeling and prediction of degradation signal with two line segments (a and c) and three line segments (b and d). The dark regions are prediction confidence intervals.

Suppose there are I historical CM signals. Let Y_i denote the CM signal of the ith unit, and $Y_i = \{y_{i,1}, y_{i,2}, \dots, y_{i,j}, \dots, y_{i,n_i}\}, i = 1, \dots, I$ where $y_{i,j}$ is the jth observation of unit i at time $t_{i,j}$, and n_i is the total number of observations of unit i in the lifetime. Let k_i denote the total number of change-points of unit i before failure, which is modelled as a random variable to account for the unit-to-unit heterogeneity. Following a conventional notation of multiple change-point models [76], suppose the k_i change-points are the integer-valued indices $c_{i1}, c_{i2}, \dots, c_{ik_i}$. For notational convenience, we define $c_{i0} = 0$ and $c_{ik_i+1} = n_i$. Then $c_{i0} = 0 <$

18

$c_{i1} < c_{i2} < \cdots < c_{ik_i} < c_{ik_i+1} = n_i$. Consequently, the sequence of observations $\{y_{i,1}, y_{i,2}, \ldots, y_{i,j}, \ldots, y_{i,n_i}\}$ are partitioned into $k_i + 1$ contiguous segments $y_{i,c_{i0}+1:c_{i1}}, y_{i,c_{i1}+1:c_{i2}}, \ldots y_{i,c_{ik_i}+1:n_i}$. Mathematically, the multiple change-point model can be expressed as

$$
y_{i,j} = \begin{cases}
a_{i1} + b_{i1}t_{ij} + \sigma_{i1}\varepsilon_{ij}, & 0 < t_{ij} \leq c_{i1} \\
a_{i2} + b_{i2}t_{ij} + \sigma_{i2}\varepsilon_{ij}, & c_{i1} < t_{ij} \leq c_{i2} \\
\cdots \\
a_{ik_i} + b_{ik_i}t_{ij} + \sigma_{ik_i}\varepsilon_{ij}, & c_{ik_i-1} < t_{ij} \leq c_{ik_i} \\
a_{ik_i+1} + b_{ik_i+1}t_{ij} + \sigma_{ik_i+1}\varepsilon_{ij}, & c_{ik_i} < t_{ij} \leq T_i
\end{cases}
\tag{2.1}
$$

where ε_{ij} follows *i.i.d.* standard normal distribution, a_{ik}, b_{ik} and σ_{ik} are the intercept, slope and standard deviation of the k-th line segment, respectively. Give the position of a change-point, we assume that the observations before that change-point is independent of those after the change-point. For simplicity and without loss of generality, we assume that $t_{i,j} = j$ in the rest of the dissertation, i.e., the sampling intervals equal to 1 for all units.

Denote a multiple change-point model as $\boldsymbol{M} = \left(k, \{\delta^{(s)}\}_{s=1}^{k+1}, \{\boldsymbol{\theta}^{(s)}\}_{s=1}^{k+1}\right)$ where k is the number of change-points, $\delta^{(s)} = c_s - c_{s-1}$ is the duration of the sth segment, and $\boldsymbol{\theta}^{(s)} = \left(a^{(s)}, b^{(s)}, \sigma^{2(s)}\right)$ is the model parameters of the sth segment. Bayesian approach is commonly employed to integrate historical data with newly observed CM signal of a working unit for sequential model updating and RUL prediction. The prognostics often involves two stages under Bayesian framework. At the offline stage, the historical data provide prior information on the number of change-points, locations of change-points, and possible values of model parameters of each line segment, all these parameters are modelled with appropriate distributions, and the hyperparameters are estimated. At the online stage, the posterior distributions of the individual

model parameters are sequentially updated based on the prior information and observed CM signal of a working unit up to the current time. Denote $y_{1:t}$ as the observations of a working unit by the current time index t, and denote $\boldsymbol{x}_t = (a_t, b_t, \sigma_t^2, \tau_t, s_t, k)$ where a_t, b_t, σ_t^2 are the parameters of the current line segment, τ_t is the latest change-point, i.e., the starting time of the current line segment, s_t is the phase or stage index. In the model updating, the posterior of the current line segment can be expressed as

$$p(\boldsymbol{x}_t|y_{1:t}) \propto \pi(\boldsymbol{x}_t)p(y_{1:t}|\boldsymbol{x}_t) \tag{2.2}$$

where $\pi(\cdot)$ is the prior distribution obtained in the offline modeling of historical data. With the model with updated parameters, we can predict the future degradations. To predict the RUL, the posterior distribution of the future degradations $y_{t^*}, \forall t^* > t$ needs to be calculated based on the updated model

$$P(y_{t^*}|y_{1:t}) = \sum_{k,s_t,\tau_t} \int P(y_{t^*}|\boldsymbol{x}_t)P(\boldsymbol{x}_t|y_{1:t})d(a_t, b_t, \sigma_t^2) \tag{2.3}$$

where

$$P(y_{t^*}|\boldsymbol{x}_t) = \sum_{s_{t^*},\tau_{t^*}} \int P(y_{t^*}|\boldsymbol{x}_{t^*})\pi(\boldsymbol{x}_{t^*}|\boldsymbol{x}_t)d(a_{t^*}, b_{t^*}, \sigma_{t^*}^2) \tag{2.4}$$

Although the above three equations have simple formulation, they are generally intractable due to high dimensionality and high nonlinearity caused by the unknown change-points. To address this challenge, we reformulate the multiple change-point model to a nonstandard state-space model and use particle filtering techniques to approximate these posteriors. Section 2.3 will introduce the specification and calculation of priors based on the historical data, and the state-space representation of the multiple change-point model. Section 2.4 will give the technical details of the developed particle filtering algorithm.

2.3 PRIOR SPECIFICATION AND STATE-SPACE REPRESENTATION

Prior distribution plays an important role in Bayesian data analysis. Informative priors are often preferred if historical data is available, since they reflect the strong belief of a new unit and can lead to more accurate posterior inference of the degradation path. In this section, the priors for the multiple change-point model $\mathcal{M} = \left(k, \{\delta^{(s)}\}_{s=1}^{k+1}, \{\boldsymbol{\theta}^{(s)}\}_{s=1}^{k+1}\right)$ are specified and estimated.

In the Bayesian formulation of multiple change-point models with a *fixed* number of observations, the priors for the number of change-point k, the segment durations $\{\delta^{(s)}, s = 1, ..., k+1\}$ and the segment parameters $\{\boldsymbol{\theta}^{(s)}, s = 1, ..., k+1\}$ can be specified easily. For the change-points, a joint prior could be placed, i.e., $\pi\left(k, \{\delta^{(s)}\}_{s=1}^{k+1}\right) = \pi(k)\pi\left(\{\delta^{(s)}\}_{s=1}^{k+1}\big|k\right)$. More commonly, a Markov process could be assumed to simultaneously model the priors for the number of change-points and their occurrence intervals or equivalently their locations [77-81]. For example, a Poisson process could be used to model the occurrence of change-points, where the intervals $\delta^{(s)}$ between successive change-points follow an *i.i.d.* exponential distribution, and the last interval satisfies $\delta^{(k+1)} \geq T - c_k$. In such case, the prior density can be derived as

$$\pi\left(k, \{\delta^{(s)}\}_{s=1}^{k+1}\right) = \left[\prod_{s=1}^{k} f(\delta^{(s)}|\lambda)\right] P(\delta^{(k+1)} \geq T - c_k) = \lambda^k \exp(-\lambda T) \tag{2.5}$$

where $f(\cdot|\lambda)$ is the probability density function of an exponential distribution, λ is the Poisson rate and T is the total number of observations. In a Bernoulli process, each time step has the probability p to be a change-point and the intervals follow an *i.i.d.* geometric distribution [78, 82, 83]. The joint density is simply

$$\pi\left(k,\{\delta^{(s)}\}_{s=1}^{k+1}\right) = p^k (1-p)^{T-1-k} \tag{2.6}$$

where p is the parameter for the Bernoulli distribution. For the changing parameters $\left(\boldsymbol{\theta}^{(s)}, s = 1, \dots, k\right)$, *i.i.d.* Gaussian distribution is often assigned.

The aforementioned renewal process is often applied in the segmentation of time series data of a known and fixed length and the priors specified are often non-informative, i.e., the phase duration follows the same distribution. However, considering the phase heterogeneity of the CM signals, the prior distributions for the phase durations should be different to make the prior more informative for RUL prediction. Also, for a working unit, the number of observations to be collected before it fails is unknown. If a renewal process is applied to model the priors, an unlimited number of change-points beyond the current time has to be considered, which is unrealistic for informative prior specifications and RUL prediction. To solve this problem, we first place a prior distribution on the number of change-point k. Conditioning on k, the distribution of the phase interval lengths are modelled by a stochastic process where the first k interval lengths $\{\delta^{(s)}\}_{s=1}^{k}$ follow independent and non-identical distributions and the $k+1$ model parameters $\{\boldsymbol{\theta}^{(s)}\}_{s=1}^{k+1}$ follow independent and non-identical distributions. Then the prior could be factorized as

$$\pi(\mathcal{M}) = \pi(k) \prod_{s=1}^{k} \pi(\delta^{(s)}|k) \prod_{s=1}^{k+1} \pi(\boldsymbol{\theta}^{(s)}|k) \tag{2.7}$$

More specifically, we put a categorical distribution or generalized Bernoulli distribution on k, with $\pi(k = g) = p_g$ and $\sum_g p_g = 1$. For simplicity, we assume the phase durations follow normal distributions, $\delta^{(s)}|k \sim N(\delta_0^{(k,s)}, \sigma_0^{2(k,s)})$. For the changing parameters, the commonly

used normal and inverse Gamma are specified, i.e., $\boldsymbol{\beta}^{(s)}|k = \left(a^{(s)}, b^{(s)}|k\right)' \sim N(\boldsymbol{\mu}_0^{(k,s)}, \boldsymbol{\Sigma}_0^{(k,s)})$

and $\sigma^{2(s)}|k \sim IG\left(\alpha_1^{(k,s)}, \alpha_2^{(k,s)}\right)$. Since the CM signal often increases rapidly when it is

approaching the failure threshold in the last phase, we assume a truncated normal prior for the

last segment to make the prior more informative: $\boldsymbol{\beta}^{(k+1)}|k \sim TN\left(\boldsymbol{\mu}_0^{(k,k+1)}, \boldsymbol{\Sigma}_0^{(k,k+1)}|b^{(k+1)} > l_k\right)$

where l_k is a positive lower bound of the slope for k-change-point case. Note that here we

assume the model parameters are independent across different phases to reduce both the

computational complexity and the required number of historical CM signals.

To specify informative priors, the hyper-parameters of all these priors, denoted as $\boldsymbol{\psi}$,

need to be estimated based on the historical data. One common way is to estimate $\boldsymbol{\psi}$ by

maximizing the marginal likelihood [63] of I historical CM signals

$$\widehat{\boldsymbol{\psi}} = \arg\max_{\boldsymbol{\psi}} \prod_{i=1}^{I} \int P(\boldsymbol{Y}_i|\mathcal{M}_i)\pi(\mathcal{M}_i|\boldsymbol{\psi})\,d\mathcal{M}_i \qquad (2.8)$$

Unfortunately, the marginal likelihood is very complex and not tractable. An alternative

approach is a two-stage process where the model parameters $\widehat{\mathcal{M}}_i$ of each unit i are first

obtained through the maximum likelihood estimates (MLE) and then the hyper-parameters are

estimated through the MLE by treating these estimated models $\{\widehat{\mathcal{M}}_i, i = 1, \dots, I\}$ as

observations. In our case, however, MLE cannot be directly applied to each CM signal since

increasing k will also increase the fitting accuracy, and thus result in over-fitting. To address

this issue, we propose to use Bayesian information criterion (BIC) [84] for model selection and

parameter estimation of each CM signal

$$\widehat{\mathcal{M}} = \arg\min_{\mathcal{M}}(-2l(\mathcal{M}|Y) + n\log T) \tag{2.9}$$

where $n = 4k + 3$ is the number of model parameters to estimate (k change-points, $k + 1$ slopes, intercepts and noise variance), $l(\mathcal{M}|Y)$ is the log-likelihood given as

$$l(\mathcal{M}|Y) = \sum_{s=1}^{k+1}\left[-\frac{1}{2}\delta^{(s)}\log(2\pi\sigma^{2(s)}) - \frac{\left\|y_{c_{s-1}+1:c_s}^T - X_{c_{s-1}+1:c_s}\boldsymbol{\beta}^{(s)}\right\|^2}{2\sigma^{2(s)}}\right] \tag{2.10}$$

where

$$X_{c_{s-1}+1:c_s} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \tau_{c_{s-1}}+1 & \tau_{c_{s-1}}+2 & \cdots & c_s \end{bmatrix}^T \tag{2.11}$$

Given the number of change-points and their locations $\{k, c_1, \ldots, c_k\}$, the parameters $\{\boldsymbol{\beta}^{(s)}, \sigma^{2(s)}, s = 1, \ldots, k+1\}$ that minimize the BIC are just the MLE of the Gaussian linear models of each phase

$$\widehat{\boldsymbol{\beta}}^{(s)} = \left(X_{c_{s-1}+1:c_s}^T X_{c_{s-1}+1:c_s}\right)^{-1} X_{c_{s-1}+1:c_s}^T y_{c_{s-1}+1:c_s}^T,$$

$$\hat{\sigma}^{2(s)} = \left\|y_{c_{s-1}+1:c_s}^T - X_{c_{s-1}+1:c_s}\widehat{\boldsymbol{\beta}}^{(s)}\right\|^2 / \delta^{(s)} \tag{2.12}$$

If k and T are small, it is possible to try all combinations $\{k, c_1, \ldots, c_k\}$ to determine the optimal model. However, this method is not realistic for large k and T due to the exponentially increased computational cost. Instead we could use the PELT method [85], which is computationally efficient with a computational cost that is linear with T.

Based on the above prior specification, the multiple change-point model could be formulated to a non-standard state-space model with state vector $x_t = (\boldsymbol{\theta}^{(s_t)}, \tau_t, s_t, k)$ and prior state transition process

$$
x_{t+1} = \begin{cases} x_t, & p = 1 - p_{t+1}(x_t) \text{ if } s_t < k+1 \\ x_t, & p = 1 \quad\quad \text{if } s_t = k+1 \\ (\boldsymbol{\theta}^{(s_t+1)}, t, s_t + 1, k) & p = p_{t+1}(x_t) \quad \text{if } s_t < k+1 \end{cases} \tag{2.13}
$$

Here $p_{t+1}(x_t)$ is the transition probability of the stochastic process expressed by

$$
\begin{aligned}
p_{t+1}(x_t) &= p\big(\delta^{(s_t)} \le L + 1 \big| \delta^{(s_t)} \ge L\big) \\[2mm]
&= \frac{\Phi\left(L + 1 \big| \delta_0^{(k,s_t)}, \sigma_0^{2(k,s_t)}\right) - \Phi\left(L \big| \delta_0^{(k,s_t)}, \sigma_0^{2(k,s_t)}\right)}{1 - \Phi(L \big| \delta_0^{(k,s_t)}, \sigma_0^{2(k,s_t)})}
\end{aligned} \tag{2.14}
$$

where $L = t - \tau_t$ and $\Phi(\cdot)$ is the Gaussian cumulative distribution function. Note that when a hidden state is continuous-valued, the term *state-space model* is often used instead of *hidden Markov model*. Here we refer to our model as a *non-standard state-space model* in that its state vector contains both discrete and continuous-valued components, and the state x_{t+1} is not linearly correlated with x_t, which is different from standard state-space model.
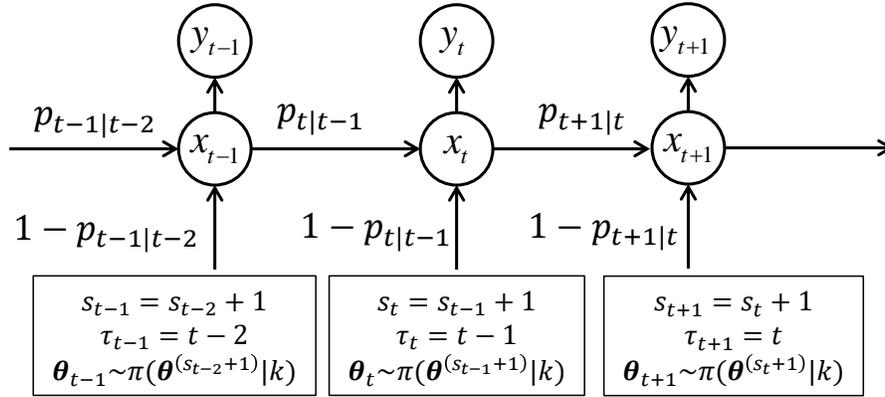
Figure 2-2: Illustration of the formulated state-space model.

The formulated state-space model is illustrated in Figure 2-2, where the transition probability from x_t to x_{t+1} can be expressed as

$$p_{t+1|t} = [1 - p_{t+1}(x_t)]^{\mathbf{1}(s_t < k+1)} \tag{2.15}$$

If there are no change-points, the formulated state-space model is a special linear state-space model with a constant state, which can be easily inferred using Kalman filters. However, due to the existence of unknown change-points, the formulated state-space is highly nonlinear, which makes the inference very challenging. The particle filtering techniques are particularly effective for nonlinear state-space models and have been widely applied in the prognosis area. Generally, the applications in the prognosis area can be classified into three categories based on the underlying state-space model: (1) nonlinear state transition model, linear observation model [86]; (2) linear state transition model, nonlinear observation model [87, 88] and (3) nonlinear state transition model and nonlinear observation model [89, 90]. The formulated state-space model in this paper falls into the third category. However, it is fundamentally different from the existing ones due to its special characteristics, i.e., high dimensionality, containing both discrete and continuous states, some states being constant (linear transition) across all life cycle while some states being constant between two successive change-points but changing once a new change-point occurs (nonlinear transition). To our best knowledge, none of the existing algorithms work well on our model. In the following section, a novel stratified particle filtering algorithm with partial Gibbs sampling strategy will be developed for sequential model updating and RUL prediction.

2.4 PARTICLE FILTERING ALGORITHM FOR ONLINE MODEL UPDATING AND RUL PREDICTION

Particle filters (PF) are effective sequential Monte Carlo (SMC) methods to solve the filtering problems. It is particularly useful for sequential Bayesian inference of linear/nonlinear Gaussian/non-Gaussian state-space models [91]. In this section, a customized particle filtering

algorithm is developed for sequential model estimation and RUL prediction of a working unit. For the sake of completeness, the basic theory of PF algorithm is first presented.

2.4.1 Review of Particle Filtering Algorithm

The basic idea of the PF technique is the sequential importance sampling (SIS). Consider a state-space model described as

$$x_1 \sim f(x), x_t|x_{t-1} \sim f(x_t|x_{t-1}), y_t|x_t \sim g(y_t|x_t) \tag{2.16}$$

where $f(\cdot)$ is the prior for the first state and $f(\cdot|\cdot)$ is the prior state transition probability density associated with state changing from x_{t-1} to x_t, and $g(\cdot|\cdot)$ is the density function of y_t conditioning on x_t. The observations $y_{1:T}$ are assumed to be conditionally independent given $x_{1:T}$. According to Bayes' theorem, the posterior density satisfies the following recursion

$$p(x_{1:t}|y_{1:t}) = p(x_{1:t-1}|y_{1:t-1}) \frac{f(x_t|x_{t-1})g(y_t|x_t)}{p(y_t|y_{1:t-1})} \tag{2.17}$$

Where

$$p(y_t|y_{1:t-1}) = \int p(x_{t-1}|y_{1:t-1}) f(x_t|x_{t-1})g(y_t|x_t)dx_{t-1:t} \tag{2.18}$$

In the filtering problem, $p(x_t|y_{1:t})$ is of interest and can be obtained by integrating out $x_{1:t-1}$ or directly based on Bayes' theorem

$$p(x_t|y_{1:t}) = \frac{p(x_t|y_{1:t-1})g(y_t|x_t)}{p(y_t|y_{1:t-1})} \tag{2.19}$$

where

$$p(x_t|y_{1:t-1}) = \int p(x_{t-1}|y_{1:t-1}) f(x_t|x_{t-1})dx_{t-1} \tag{2.20}$$

Equation (2.20) is known as the prediction step and (2.19) is called as the updating step. However, Equation (2.18) and (2.20) are often intractable analytically, and SIS is often used for

posterior approximation. If we select an important distribution that can be sequentially sampled with the following structure

$$q_t(\boldsymbol{x}_{1:t}) = q_1(\boldsymbol{x}_1) \prod_{i=2}^{t} q_i(\boldsymbol{x}_i|\boldsymbol{x}_{1:i-1}) \tag{2.21}$$

Then the unnormalized weight function can be expressed by

$$\omega_t(\boldsymbol{x}_{1:t}) = \frac{p(\boldsymbol{x}_{1:t}, y_{1:t})}{q_t(\boldsymbol{x}_{1:t})} = \frac{f(\boldsymbol{x}_1) \prod_{i=2}^{t} f(\boldsymbol{x}_i|\boldsymbol{x}_{i-1}) \prod_{i=1}^{t} g(y_i|\boldsymbol{x}_i)}{q_1(\boldsymbol{x}_1) \prod_{i=2}^{t} q_i(\boldsymbol{x}_i|\boldsymbol{x}_{1:i-1})} = \omega_1(\boldsymbol{x}_1) \prod_{i=2}^{t} w_i \tag{2.22}$$

where

$$\omega_1(\boldsymbol{x}_1) = f(\boldsymbol{x}_1)g(y_1|\boldsymbol{x}_1)/q_1(\boldsymbol{x}_1)$$
$$w_i = g(y_i|\boldsymbol{x}_i)f(\boldsymbol{x}_i|\boldsymbol{x}_{i-1})/q_i(\boldsymbol{x}_i|\boldsymbol{x}_{1:i-1}) \tag{2.23}$$

Equation (2.22) shows that the weight function can be calculated recursively, so that the posterior could be sequentially updated once a new observation is measured. The expectation of any function $\varphi(\boldsymbol{x}_{1:t})$ with respect to the posterior $p(\boldsymbol{x}_{1:t}|y_{1:t})$ can be estimated by

$$E(\varphi(\boldsymbol{x}_{1:t})|y_{1:t}) \approx \sum_{i=1}^{N} W_t^{(i)} \varphi\left(\boldsymbol{x}_{1:t}^{(i)}\right) \tag{2.24}$$

where $W_t^{(i)}$ is the normalized weight. In the PF algorithm, a resampling step based their updated weights is often added to obtain equally weighted particles which are approximately distributed as $p(\boldsymbol{x}_{1:t}|y_{1:t})$. It is a "Darwinian" procedure that can remove particles with low weights and carry on particles with high weights. The generic particle filtering algorithm with a resampling step is given in Table 2-1 as follows.

Table 2-1: Generic Particle Filtering Algorithm

At time $t = 1$,

1. Sample $\boldsymbol{x}_1^{(i)} \sim q_1(\boldsymbol{x}_1)$ for $i = 1, \dots, N$

2. Compute weights $\omega_1(\boldsymbol{x}_1^{(i)})$ and normalized weights $W_1^{(i)} = \omega_1(\boldsymbol{x}_1^{(i)}) / \sum_{i=1}^{N} \omega_1\left(\boldsymbol{x}_1^{(i)}\right)$.

3. Resample $\{W_1^{(i)}, \boldsymbol{x}_1^{(i)}\}$ according to their weights to obtain N equally weighted particles $\{\frac{1}{N}, \bar{\boldsymbol{x}}_1^{(i)}\}$ and set $\left\{W_1^{(i)}, \boldsymbol{x}_1^{(i)}\right\} \leftarrow \left\{\frac{1}{N}, \bar{\boldsymbol{x}}_1^{(i)}\right\}$.

   At time $t \geq 2$:

1. Sample $\boldsymbol{x}_t^{(i)} \sim q_t\left(\boldsymbol{x}_t \middle| \boldsymbol{x}_{1:t-1}^{(i)}\right)$, set $\boldsymbol{x}_{1:t}^{(i)} \leftarrow (\boldsymbol{x}_{1:t-1}^{(i)}, \boldsymbol{x}_t^{(i)})$ for $i = 1, \dots, N$

2. Compute $w_t\left(\boldsymbol{x}_{1:t}^{(i)}\right)$ and normalized weights $W_t^{(i)}$

3. Resample $\{W_t^{(i)}, \boldsymbol{x}_{1:t}^{(i)}\}$ to obtain N equally weighted particles $\{\frac{1}{N}, \bar{\boldsymbol{x}}_{1:t}^{(i)}\}$ and set $\left\{W_t^{(i)}, \boldsymbol{x}_{1:t}^{(i)}\right\} \leftarrow \left\{\frac{1}{N}, \bar{\boldsymbol{x}}_{1:t}^{(i)}\right\}$

2.4.2 Stratified Particle Filtering Algorithm for Model Updating

In the development of PF algorithm, the importance function needs to be specified. The optimal importance function should be the one that minimizes the variances of the importance weight of sampled particles [91]. It can reduce the particle degeneracy issue, i.e., the weights concentrate on only a few particles and most particles have negligible weights. However, the optimal importance function is often not obtainable in practice. Instead, we propose to use the prior transition density as the importance density function

$$q_t(\boldsymbol{x}_t | \boldsymbol{x}_{1:t-1}) = f(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}) \tag{2.25}$$

Selecting the prior transition density is the most common and convenient way in practical applications. According to Equation (2.23), with the resampling step implemented, the weight is simply

$$\omega_t(\boldsymbol{x}_{1:t}) = g(y_t|\boldsymbol{x}_t) \tag{2.26}$$

at each step, which greatly simplifies the computation.

Another important issue commonly faced in PF techniques is the particle impoverishment problem [82], where the number of unique particles or unique components of particles becomes less and less along iterations due to the resampling step. In our state-space model, each line segment between successive change-points is a special linear state-space model with a constant state, which makes the particle impoverishment problem even worse. Besides, the dimension of the state vector is relatively high, which may require a significantly large number of particles to guarantee the approximation accuracy, and thus result in high computational cost. However, for online model updating and RUL prediction, a low computational cost is often critically important. In this paper, we propose a Gibbs resample-move step to address both these issues. The resample-move strategy was first proposed by Gilks [92], where a "move" step is added after the resampling step to generate new particles through MCMC kernels with the posterior distribution as the invariant distributions. It can not only diversify the particles to reduce the particle impoverishment issue, but can also generate more particles with significant weights, thus reducing the particle degeneracy issue and reducing the required number of particles. In our algorithm, we propose a one-step partial Gibbs sampler to "move" the three continuous components $(\boldsymbol{\beta}_t, \sigma_t^2)$ through their conditional posterior distributions. The conditional posterior distributions are obtained based Lemma 1 as follows (the proof is included in the Appendix A).

Lemma 2.1. Suppose $\boldsymbol{\beta}^{(s)}|k = \left(a^{(s)}, b^{(s)}|k\right)' \sim N(\boldsymbol{\mu}_0^{(k,s)}, \boldsymbol{\Sigma}_0^{(k,s)})$ for $s = 1, \ldots, k$, $\boldsymbol{\beta}^{(k+1)}|k \sim TN\left(\boldsymbol{\mu}_0^{(k,k+1)}, \boldsymbol{\Sigma}_0^{(k,k+1)}|b^{(k+1)} > l_k\right)$ and $\sigma^{2(s)}|k \sim IG\left(\alpha_1^{(k,s)}, \alpha_2^{(k,s)}\right)$ for $s = 1, \ldots, k+1$, then

$$(\boldsymbol{\beta}_t | y_{1:t}, \sigma_t^2, \tau_t, s_t = s, k) \sim \begin{cases} N(\boldsymbol{\mu}_t^{(k,s)}, \boldsymbol{\Sigma}_t^{(k,s)}), & \text{if } s \le k \\ TN(\boldsymbol{\mu}_t^{(k,k+1)}, \boldsymbol{\Sigma}_t^{(k,k+1)} | b_t^{(k+1)} > l_k), & \text{if } s = k+1 \end{cases} \qquad (2.27)$$

$$(\sigma_t^2 | y_{1:t}, \boldsymbol{\beta}_t, \tau_t, s_t = s, k) \sim IG\left(\alpha_1^{(k,s)} + \frac{t - \tau_t}{2}, \alpha_2^{(k,s)} + \frac{\left\| y_{\tau_t+1:t}^T - \boldsymbol{X}_{\tau_t+1:t} \boldsymbol{\beta}_t \right\|^2}{2} \right) \qquad (2.28)$$

Where

$$\boldsymbol{\mu}_t^{(k,s)} = \left[\frac{\boldsymbol{X}_{\tau_t+1:t}^T \boldsymbol{X}_{\tau_t+1:t}}{\sigma_t^2} + \boldsymbol{\Sigma}_0^{-1(k,s)} \right]^{-1} \left[\frac{\boldsymbol{X}_{\tau_t+1:t}^T y_{\tau_t+1:t}^T}{\sigma_t^2} + \boldsymbol{\Sigma}_0^{-1(k,s)} \boldsymbol{\mu}_0^{(k,s)} \right], s = 1, \ldots, k+1 \qquad (2.29)$$

and

$$\boldsymbol{\Sigma}_t^{(k,s)} = \left[\frac{\boldsymbol{X}_{\tau_t+1:t}^T \boldsymbol{X}_{\tau_t+1:t}}{\sigma_t^2} + \boldsymbol{\Sigma}_0^{-1(k,s)} \right]^{-1}, s = 1, \ldots, k+1 \qquad (2.30)$$

The Gibbs "move" step could effectively diversify particles and generate more particles with significant weights. However, the introduction could result in extra computational cost as well as break the balance of the computational load at each time step. Based on Lemma 2.1, all the observations from the latest change-point to the current time are used for Gibbs move. It is intuitive that the longer the phase duration, the higher the computational cost the Gibbs move will take. To control the computational cost, we adopt the "partial move" strategy [93], where randomly drawn particles among the resampled particles are moved until the sum of their durations $t - \tau_t^{(i)}$ is larger than a controlling constant C.

Although the Gibbs move step has solved the particle impoverishment issue for the continuous components $(\boldsymbol{\beta}_t, \sigma_t^2)$, it could not handle the same problem with the discrete component k. Indeed, the discrete component k of each particle is generated at the first time step and kept constant across all the following time steps. That means the impoverishment issue is much worse than the other components of the state vector. As we observed, after only several iterations, there may be only one unique value for k among all particles, which will result in a

totally failed PF algorithm. To solve this problem, we propose to use a stratified approach. Specifically, for each category $k = g$ in the categorical distribution, the developed particle filtering algorithm with the same number of particles N is applied individually. In the posterior approximation, the extra group weight coefficient $W_t^{(g)}$ is applied to each category g or all particles of each group. The group weight coefficient $W_t^{(g)}$ can be calculated as

$$W_t^{(g)} = \frac{\sum_{i=1}^{N} \omega_t \left(\boldsymbol{x}_{1:t}^{(g,i)} \right)}{\sum_{g=1}^{\dim(k)} \sum_{i=1}^{N} \omega_t \left(\boldsymbol{x}_{1:t}^{(g,i)} \right)} \tag{2.31}$$

This strategy can effectively avoid the disappearing of certain k in the resampling process. We call this approach the *stratified particle filtering (SPF)*. In summary, the developed SPF algorithm for sequential model updating is given in Table 2-2.

Table 2-2: Stratified Particle Filtering Algorithm for Sequential Model Updating

At time $t = 1$:

For $g = 1: \dim\{k\}$

1. Set $k^{(g,i)} = k_g$, Sample $\boldsymbol{\beta}_1^{(g,i)} \sim N\left(\boldsymbol{\mu}_0^{(k^{(g,i)},1)}, \boldsymbol{\Sigma}_0^{(k^{(g,i)},1)}\right)$,

    $\sigma_1^{2(g,i)} \sim IG\left(\alpha_1^{(k^{(g,i)},1)}, \alpha_2^{(k^{(g,i)},1)}\right)$. Set $\tau_1^{(g,i)} = 0, s_1^{(g,i)} = 1$, and set $\boldsymbol{x}_1^{(g,i)} = $

    $(\boldsymbol{\beta}_1^{(g,i)}, \sigma_1^{2(g,i)}, \tau_1^{(g,i)}, s_1^{(g,i)}, k^{(g,i)})$ for $i = 1: N$

2. Compute weights $\omega_1(\boldsymbol{x}_1^{(g,i)})$ and normalized weights $W_1^{(g,i)} = \omega_1(\boldsymbol{x}_1^{(g,i)})/$

    $\sum_{i=1}^N \omega_1\left(\boldsymbol{x}_1^{(g,i)}\right)$ based on Eq. (2.26) for $i = 1: N$

3. Resample $\{W_1^{(g,i)}, \boldsymbol{x}_1^{(g,i)} | i = 1, \dots, N\}$ according to their weights $W_1^{(g,i)}$ to obtain N

    equally weighted particles $\left\{\frac{1}{N}, \bar{\boldsymbol{x}}_1^{(g,i)}\right\}$ and set $\left\{W_1^{(g,i)}, \boldsymbol{x}_1^{(g,i)}\right\} \leftarrow \left\{\frac{1}{N}, \bar{\boldsymbol{x}}_1^{(g,i)}\right\}$

End

4. Calculate the group weight $W_1^{(g)}$ based on Eq. (2.31)

5. Set $W_1^{(g,i)} = \frac{W_1^{(g)}}{N}$ for $g = 1: \dim\{k\}$ and $i = 1: N$

At time $t \geq 2$:

For $g = 1: \dim\{k\}$

1. Calculate the probability $p_{t|t-1}^{(g,i)}$ based on Eq. (2.15)

2. Sample $u^{(g,i)} \sim U(0,1)$

    - If $u^{(g,i)} \leq p_{t|t-1}^{(g,i)}$,

    i. Set $\tau_t^{(g,i)} = t - 1, s_t^{(g,i)} = s_{t-1}^{(g,i)} + 1$,

    ii. Sample $\boldsymbol{\beta}_t^{(g,i)} \sim N\left(\boldsymbol{\mu}_0^{\left(k^{(g,i)}, s_t^{(g,i)}\right)}, \boldsymbol{\Sigma}_0^{\left(k^{(g,i)}, s_t^{(g,i)}\right)}\right)$ if $s_t^{(g,i)} \leq k^{(g,i)}$, otherwise, sample

    $\boldsymbol{\beta}_t^{(g,i)} \sim TN\left(\boldsymbol{\mu}_0^{\left(k^{(g,i)}, s_t^{(g,i)}\right)}, \boldsymbol{\Sigma}_0^{\left(k^{(g,i)}, s_t^{(g,i)}\right)} | b_t^{(g,i)} > l_g\right)$

    iii. Sample $\sigma_t^{2(g,i)} \sim IG\left(\alpha_1^{\left(k^{(g,i)}, s_t^{(g,i)}\right)}, \alpha_2^{\left(k^{(g,i)}, s_t^{(g,i)}\right)}\right)$

    iv. Set $\boldsymbol{x}_t^{(g,i)} \leftarrow (\boldsymbol{\beta}_t^{(g,i)}, \sigma_t^{2(g,i)}, \tau_t^{(g,i)}, s_t^{(g,i)}, k^{(g,i)})$

- Otherwise, set $x_t^{(g,i)} \leftarrow x_{t-1}^{(g,i)}$

3. Compute $\omega_t\left(x_t^{(g,i)}\right)$ and normalized weights $W_t^{(g,i)}$ based on Eq.(2.26)

4. Resample $\{W_t^{(g,i)}, x_{1:t}^{(g,i)}\}$ to obtain N equally weighted particles $\left\{\frac{1}{N}, \bar{x}_t^{(g,i)}\right\}$ and set $\left\{W_t^{(g,i)}, x_t^{(g,i)}\right\} \leftarrow \left\{\frac{1}{N}, \bar{x}_t^{(g,i)}\right\}$

5. Gibbs move: select a subset S of $\{x^{(g,i)}|i = 1:N\}$ such that $\sum_{j \in S}(t - \tau^{(g,j)}) \leq C$

   - Sample $\beta_t^{(g,j)}$ based on Eq. (2.27)

   - Sample $\sigma_t^{2(g,j)}$ based on Eq. (2.28)

End

6. Calculate the group weight $W_t^{(g)}$ based on Eq. (2.31).

7. Set $W_t^{(g,i)} = \frac{W_t^{(g)}}{N}$ for $g = 1: \dim\{k\}$ and $i = 1: N$

2.4.3 RUL Prediction

After the degradation model of the working unit is updated using the observations up to the current time, the next step is to predict the future degradation magnitude and RUL for preventive maintenance. Due to the multiple change-point that may occur in future, the exact Bayesian inference is intractable, even if the current model is known. However, through the PF algorithm, the RUL prediction is proven to be very convenient. Denote R_t as the RUL at the current time t. Then the distribution of R_t can be expresses by

$$
\begin{aligned}
P(R > L|y_{1:t}) &= P(y_{t+1} < \Gamma, y_{t+2} < \Gamma, \cdots, y_{t+L} < \Gamma|y_{1:t}) \\
&= \int P(y_{t+1} < \Gamma, \cdots, y_{t+L} < \Gamma|x_{t+1:t+L})f(x_{t:t+L}|y_{1:t})dx_{t:t+L} \\
&= \int \left[\prod_{j=1}^{L} P(y_{t+j} < \Gamma|x_{t+j})f(x_{t+j}|x_{t+j-1})\right]f(x_t|y_{1:t})dx_{t:t+L}
\end{aligned}
\tag{2.32}
$$

where Γ is the failure threshold. The above equation is not tractable analytically. However, we can conveniently generate samples from the distribution $f(x_{t:t+L}|y_{1:t})$ based on the particles at the current time and the prior state transition process. Given the particles and their weights at the current time $\{W_t^{(g,i)}, x_t^{(g,i)}, g = 1: \dim(k), i = 1: N\}$ which approximately follow $f(x_t|y_{1:t})$, the samples of the future states $\{x_{t+1:t+L}^{(i)}, i = 1, ..., N\}$ can be generated through the prior state transition function $f(x_{t+j}|x_{t+j-1})$ given in Equation (2.13). Based on the generated samples, the RUL distribution can be approximated by

$$P(R > L|y_{1:t}) = P(y_{t+1} < \Gamma, \cdots, y_{t+L} < \Gamma|y_{1:t})$$

$$\approx \sum_{g=1}^{\dim(k)} \frac{W_t^{(g)}}{N} \sum_{i=1}^{N} \prod_{j=1}^{L} P\left(y_{t+j} < \Gamma|x_{t+j}^{(g,i)}\right) \tag{2.33}$$

Conditioning that there are $k + 1$ line segments or phases for a working unit, the failure will not occur before the $(k + 1)$-th phase. Therefore, the probability $P\left(y_{t+j} < \Gamma|x_{t+j}^{(g,i)}\right)$ can be calculated by

$$P\left(y_{t+j} < \Gamma|x_{t+j}^{(g,i)}\right) = \begin{cases} \Phi(\Gamma|a_{t+j}^{(g,i)} + b_{t+j}^{(g,i)}(t + j), \sigma_t^{2(g,i)}), & s_t^{(g,i)} = k^{(g,i)} + 1 \\ 1, & s_t^{(g,i)} \leq k^{(g,i)} \end{cases} \tag{2.34}$$

where $\Phi(\cdot)$ is the CDF of Gaussian distribution.

2.5 CASE STUDIES

2.5.1 Simulation Study

Table 2-3: Hyper-parameter Specification for Numerical Simulation

Variables	Two-phase Model	Three-phase Model
$\delta^{(s)}$	$\delta_0^{(1,1)} = 400, \sigma_0^{2(1,1)} = 225$ $\delta_0^{(1,2)} = 500, \sigma_0^{2(1,2)} = 100$	$\delta_0^{(2,1)} = 200, \sigma_0^{2(2,1)} = 100$ $\delta_0^{(2,2)} = 400, \sigma_0^{2(2,2)} = 100$ $\delta_0^{(2,3)} = 500, \sigma_0^{2(2,3)} = 100$
$\boldsymbol{\beta}^{(s)}$	$\boldsymbol{\mu}_0^{(1,1)} = [-15; 0.008]$ $\boldsymbol{\Sigma}_0^{(1,1)} = \begin{bmatrix} 0.015 & 0.0014 \\ 0.0014 & 0.00046 \end{bmatrix}$ $\boldsymbol{\mu}_0^{(1,2)} = [-30; 0.3]$ $\boldsymbol{\Sigma}_0^{(1,2)} = \begin{bmatrix} 0.024 & -0.0007 \\ -0.0007 & 0.0057 \end{bmatrix}$	$\boldsymbol{\mu}_0^{(2,1)} = [-10; 0.0005]$ $\boldsymbol{\Sigma}_0^{(2,1)} = \begin{bmatrix} 0.15 & 0.00014 \\ 0.00014 & 0.0009 \end{bmatrix}$ $\boldsymbol{\mu}_0^{(2,2)} = [-18; 0.02]$ $\boldsymbol{\Sigma}_0^{(2,2)} = \begin{bmatrix} 0.024 & -0.0007 \\ -0.0007 & 0.000048 \end{bmatrix}$ $\boldsymbol{\mu}_0^{(2,3)} = [-50; 0.08]$ $\boldsymbol{\Sigma}_0^{(2,3)} = \begin{bmatrix} 0.075 & -0.00008 \\ -0.00008 & 0.00025 \end{bmatrix}$
$\sigma^{2(s)}$	$\alpha_1^{(1,1)} = 1.4, \alpha_2^{(1,1)} = 2.5$ $\alpha_1^{(1,2)} = 2, \alpha_2^{(1,2)} = 4$	$\alpha_1^{(2,1)} = 3.64, \alpha_2^{(2,1)} = 2$ $\alpha_1^{(2,2)} = 0.6, \ \alpha_2^{(2,2)} = 0.5$ $\alpha_1^{(2,3)} = 3.6, \alpha_2^{(2,3)} = 5$

In this subsection, we evaluate the performance of the proposed method through simulated piecewise linear signals. For simplicity we assume that there are only two categories of degradation signals: two-phase and three-phase cases. The categorical distribution is given by

$$k = \begin{cases} 1, \text{with } p_1 = 0.3 \\ 2, \text{with } p_2 = 0.7 \end{cases} \tag{2.35}$$

We assume the unit will fail once the observation reaches the threshold $\Gamma = 20$. The slope lower bound of last phase is set to be $l_1 = l_2 = 0.003$, The hyper-parameters of $\delta^{(k,s)}, \boldsymbol{\beta}^{(k,s)}$ and $\sigma^{2(k,s)}$ are specified in Table 2-3.

In total 200 CM signals are simulated, among which 69 are two-phase signals and 131 are three-phase signals. The BIC based model selection method can accurately obtain the right number of change-points and their locations for each simulated signal. Due to page limitation,

the estimated hyper-parameters are not listed here. In the stratified particle filtering algorithm, the number of particles for each category is set to be $N = 5000$. Figure 2-3 shows the online monitoring of degradation signals with one and two change-points. From the top two panels we can see that the estimated signals (mean value) are very close to the true values. The second row of these panels shows the mean value of the current phase length. As we can see, the algorithm can rapidly detect the phase change. The bottom four panels show the probability mass function of the discrete components (s_t, k) of the state vector. As we can see, the algorithm can accurately detect the number of phases the degradation signal will have and the current phase the degradation signal is at.

Figure 2-4 shows the comparison of the SPF algorithm to three other PF algorithms without either partial Gibbs move or stratified strategy. The number of particles here is set to be 500. Clearly, without the stratified strategy (a and c), all the particles with discrete component $k = 2$ gradually diminish along iterations, which results in an inaccurate model with only two phases ($k = 1$). Without the partial Gibbs move (b and c), the degeneracy of the continuous components occurs, which significantly influence the model accuracy. The proposed SPF algorithm has effectively overcome the particle degeneracy and impoverishment issues and works quite well with only 500 samples.
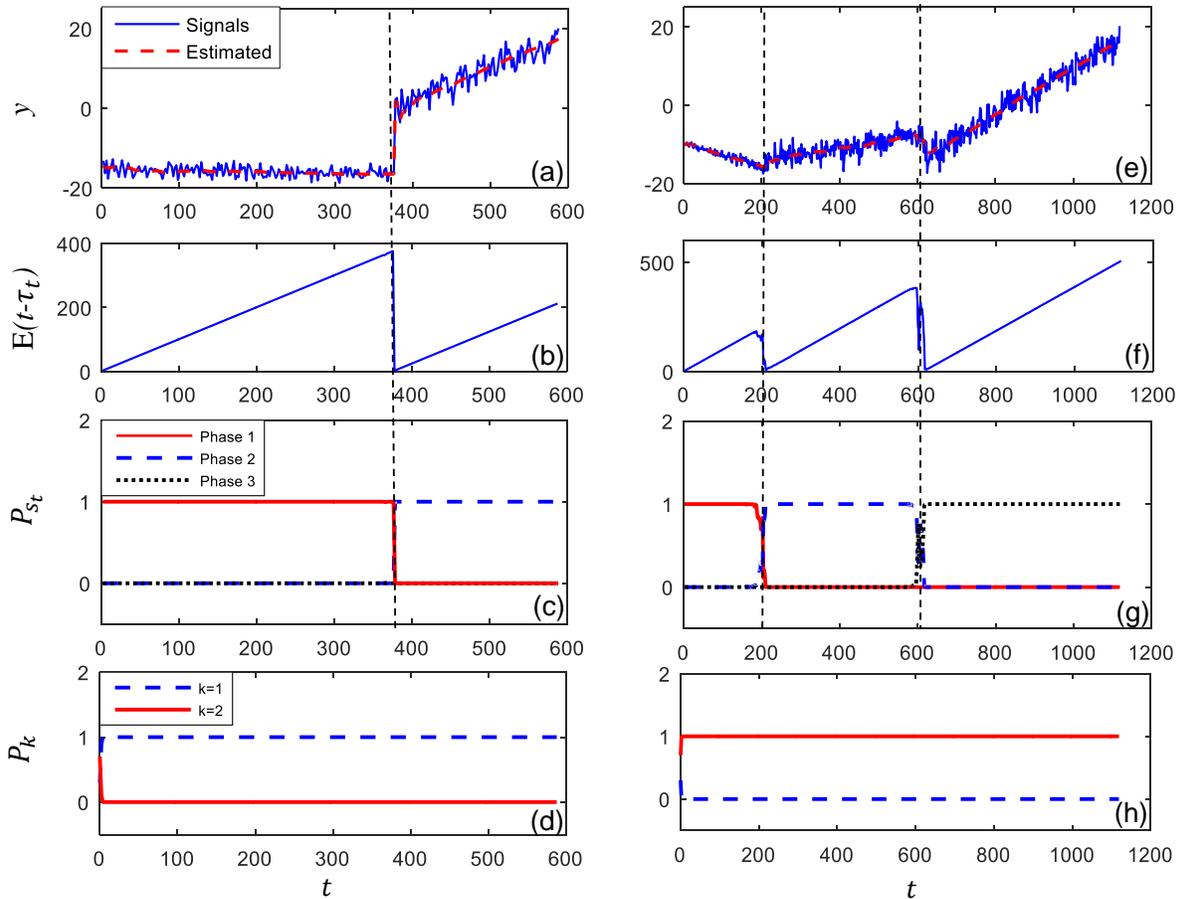
Figure 2-3: Illustration of the SPF based online monitoring of degradation signals with two phases (left panel) and three phases (right panel). (a) and (e): degradation signals and estimated signals; (b) and (f): the estimated duration of the current linear phase; (c) and (g): the probability mass function of the current phase; (d) and (h): the probability mass function of category. The vertical dashed lines are the true change-points.
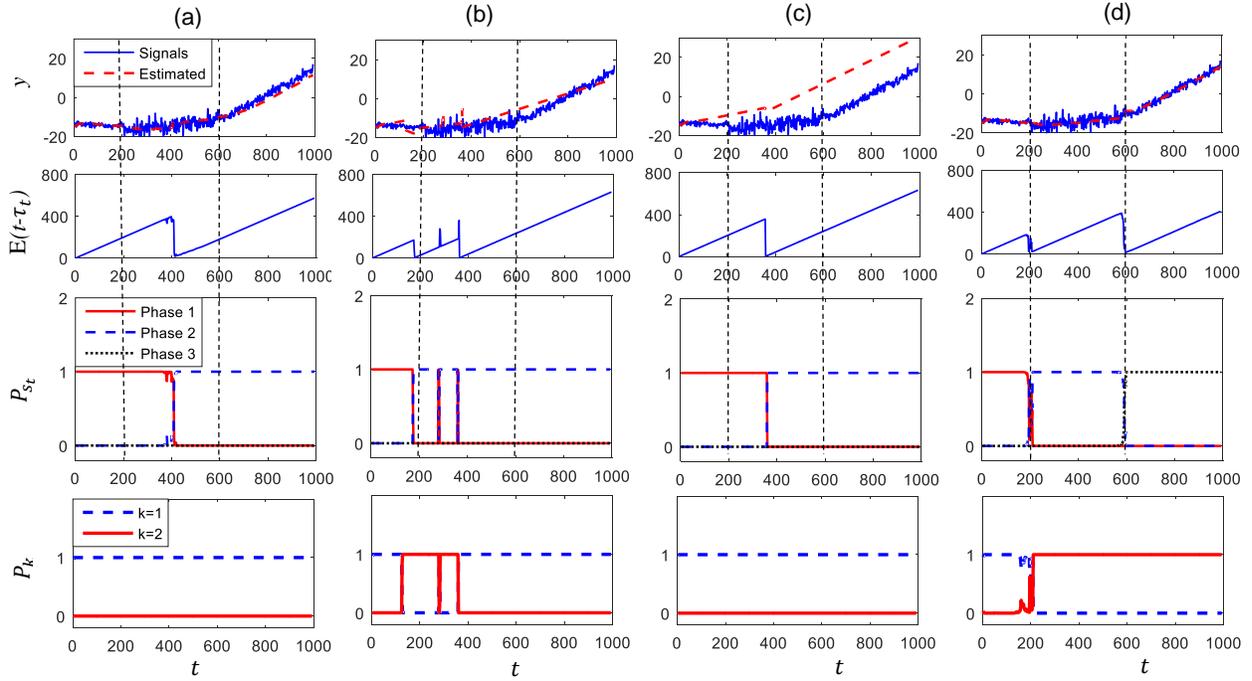
Figure 2-4: Comparison of the proposed SPF algorithm to other PF algorithms without either partial Gibbs move or stratified strategy: (a) no stratification, Gibbs move; (b) stratification, no Gibbs move; (c) no stratification, no Gibbs move and (d) proposed SPF with both particle Gibbs move and stratified approach.
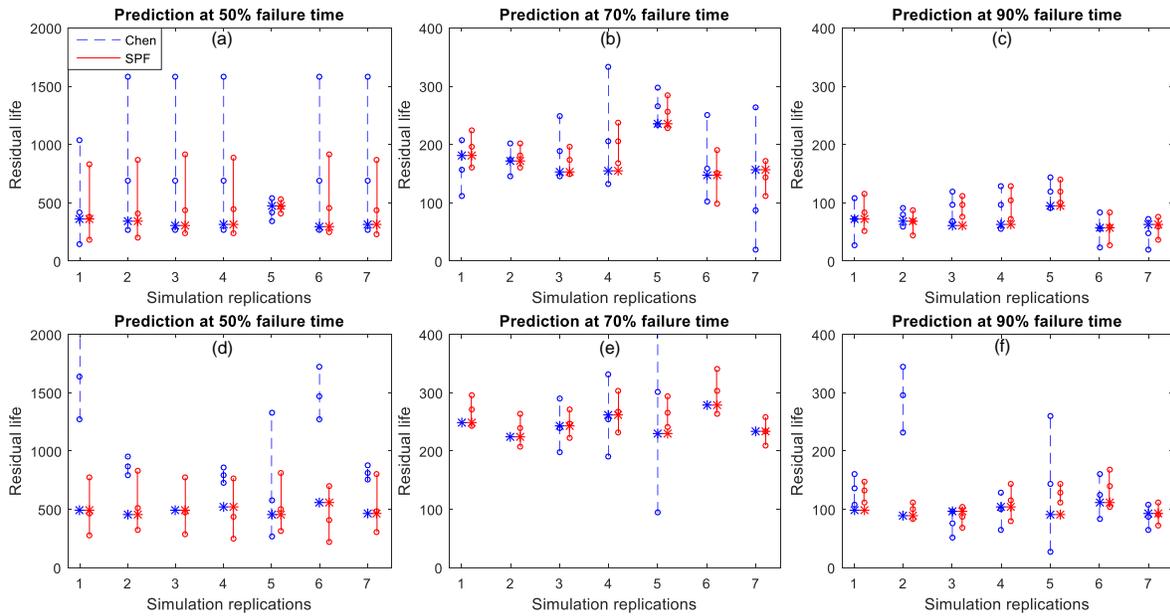
Figure 2-5: Prediction intervals of 7 simulated CM signals. The ∘ represents the 5%,50%,95% quantitles of the RUL distributions, ∗ is the actual RUL.(a)-(c) Prediction intervals for two-phase signals; (d)-(f) Prediction intervals for three-phase signals
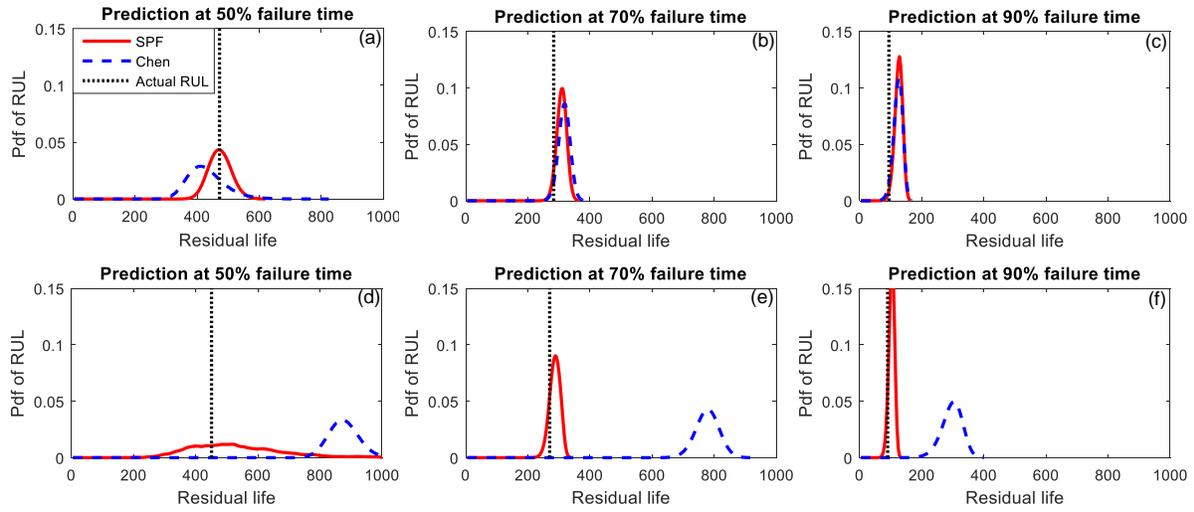
Figure 2-6: Comparison of the pdf of the RUL. (a)-(c) two-phase signal (d)-(f) three-phase signal

To evaluate the prediction performance, 200 new degradation signals are randomly generated as a testing dataset using the specified priors. We compare the SPF algorithm with Chen's two-phase model [63], where only one change-point is considered. For Chen's method, all the 200 training dataset with both two-phase and three-phase signals are used to estimate the priors of the two-phase model. Figure 2-5 shows the comparison of the prediction intervals of 7 simulated two-phase signals and 7 three-phase signals predicted at 50%, 70% and 90% of actual failure time. Figure 2-6 shows the detailed RUL prediction of the 5th and the 2nd unit of the seven signals of each category in Figure 2-5. From Figure 2-5 we can see that our prediction accuracy is much better than Chen's method in almost all the 14 cases. For two-phase signals, both methods work well. However, our method is slightly better at 70% and 90% of the failure time while much better at 50% of the failure time than Chen's method, which can also be seen from Figure 2-6(a-c). The main reason is that in Chen's method, the priors are estimated using all two-phase and three-phase signals, which will result in less accurate priors. At the 50% failure time, the prediction accuracy is mainly determined by the prior knowledge, while at the 70% and 90% of the failure time, the observations dominate the posterior distributions. Therefore at the early stage, our method with more accurate priors is much better than Chen's method while at the late stage, the performances of both methods are comparable. For three-phase signals, our method is much better than Chen's method at all the three prediction times, as shown in Figure 2-5(d-f) and Figure 2-6(d-f). It is expected since the two-phase model is inadequate to model signals with three phases.

To evaluate the overall performance, we use the root-mean-square-deviation (RMSD), which is defined as $\mathrm{RMSD} = \sqrt{E(R - R_{\mathrm{true}})^2}$, where R and R_{true} are the predicted and true RUL respectively. Since the proposed method is a Monte Carlo based method, there exists

inevitable randomness (though very small). So for each signal the SPF algorithm is repeated 10 times. Table 2-4 shows the RMSD of the proposed method and Chen's method using the 200 testing signals. As we can see, the proposed method is much more accurate than Chen's method, with the RMSD reduced by more than 70% at almost all six prediction times. As the prediction time approaches to the true failure time, the RMSD of the proposed method monotonically decreases. This is highly desirable since it becomes more and more important to get an accurate prediction when the RUL approaches zero. However, for Chen's method, RMSD first increases and then decreases. That means the prediction error at the second phase is even worse than making prediction at the first phase. The reason is that for three-phase signals, the second phase with a relatively small degradation rate is detected as the final phase in Chen's method. The more observations in the second-phase, the flatter the final phase of the updated model and thus the worse the prediction.

Table 2-4: Comparison of the RMSD at Six Prediction Times

Method	RMSD					
	40%	50%	60%	70%	80%	90%
Chen	593.4	649.4	696.7	548.7	188.6	58.6
SPF	222.1	194.6	133.8	74.4	23.4	21.9

The computational costs of the SPF and Chen's method using MATLAB running on an i7-6560U 2.21 GHz Intel processor are shown in Table 2-5. For the SPF method, the total number of particles is set 5000. In the model updating stage, we compare the computational costs of these two methods running 300 time steps. For the prediction stage, the costs of running different time steps are evaluated, since the cost of prediction in Chen's method nonlinearly increases with time steps. As we can see, the SPF method is much more expensive in the model

42

updating stage than Chen's method. However, in the prediction stage, the cost of Chen's method exponentially increases with the time steps, due to the CDF computation of a multivariate t distribution with an increasing dimension. For the SPF method, the computational cost of the prediction linearly increases with the time steps. Note that the selection of 5000 particles is quite conservative. From Figure 2-4 we can see that the model updating is quite accurate with only 500 particles.

Table 2-5: Comparison of the computational cost (UNIT: seconds)

Method	Updating	Prediction				
	300	25	50	75	100	125
Chen	0.024	2.1	7.4	16.1	28.8	45.7
SPF	179.4	6.2	11.7	16.9	22.0	27.0

2.5.2 Degradation Monitoring of the Rotational Bearings

In this section, the proposed method is applied to the real degradation signals of rotational bearings [12, 63, 94]. They are vibrational signals (log-transformed) of a set of identical thrust ball bearings captured by an accelerometer in the accelerated aging testing. There are in total 25 complete bearing signals available. The data sampling interval is 2 minutes for each signal. When the vibration magnitude exceeds the threshold $\Gamma = \log{(0.03)}$, which was computed from published industrial standards, the bearing is considered failed [12].

In the offline modeling and prior estimation process, we set the maximum number of phases for all 25 signals to be 3 to control the model complexity. It is found that all bearing signals with three phases have the minimum BIC. The estimated hyper-parameters of the prior distributions are shown in Table 2-6. It is observable that the slope is quite small at the first phase, indicating a stable operation process. The slopes of the following phases are larger than

the former ones. It indicates that when a new change-point occurs, the degradation rate of bearings increases.

Table 2-6: Estimated Hyper-parameters of the Prior Distributions

	$s = 1$	$s = 2$	$s = 3$
$\delta^{(s)}$	$\delta_0^{(1)} = 246,\ \sigma_0^{2(1)} = 163^2$	$\delta_0^{(2)} = 199,\ \sigma_0^{2(2)} = 123^2$	$\delta_0^{(3)} = 232, \sigma_0^{2(3)} = 123^2$
$\boldsymbol{\beta}^{(s)}$	$\boldsymbol{\mu}_0^{(1)} = [-7.14, 0.00027]$ $\boldsymbol{\Sigma}_0^{(1)}$ $= \begin{bmatrix} 0.15 & -0.0003 \\ -0.0003 & 1.30 \times 10^{-5} \end{bmatrix}$	$\boldsymbol{\mu}_0^{(2)} = [-6.42, 0.0028]$ $\boldsymbol{\Sigma}_0^{(2)}$ $= \begin{bmatrix} 2.05 & -0.009 \\ -0.009 & 7.36 \times 10^{-5} \end{bmatrix}$	$\boldsymbol{\mu}_0^{(3)} = [-6.8, 0.005]$ $\boldsymbol{\Sigma}_0^{(3)}$ $= \begin{bmatrix} 3.89 & -0.008 \\ -0.008 & 2.52 \times 10^{-5} \end{bmatrix}$ $l = 0.005$
$\sigma^{2(s)}$	$\alpha_1^{(1)} = 2.65, \alpha_2^{(1)} = 0.01$	$\alpha_1^{(2)} = 0.54, \alpha_2^{(2)} = 0.004$	$\alpha_1^{(3)} = 1.28, \alpha_2^{(3)} = 0.03$
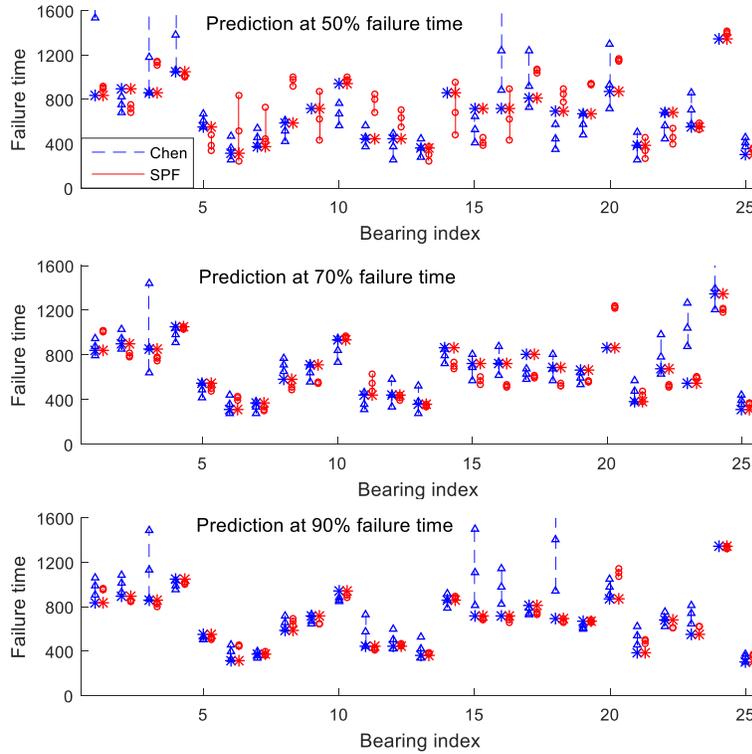
Figure 2-7: Prediction intervals of the 25 bearing signals

44
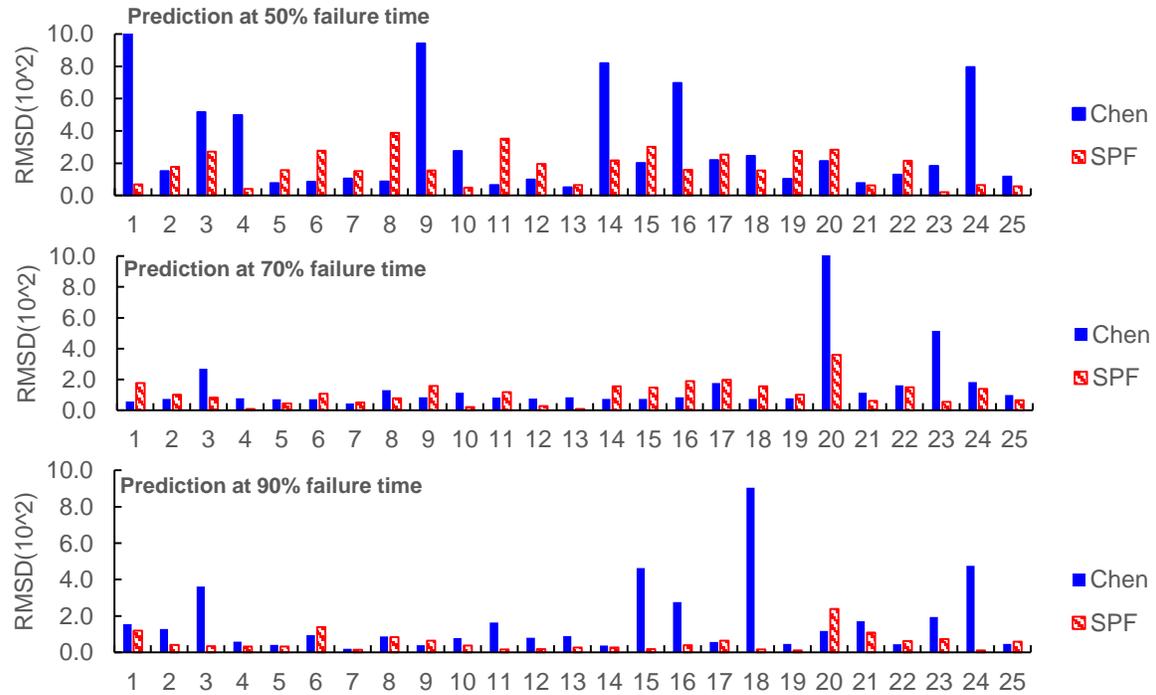
Figure 2-8: RMSD of the 25 bearing signals

Figure 2-7 shows the prediction intervals at 50%, 70% and 90% of failure time against the actual failure time. We can see that the prediction intervals at 90% failure time are much narrower than that of 70% failure time and 50% failure time. As the prediction time is closer to actual failure time, the intervals become smaller. It is obvious that the more observed data, the more accurate the prediction. Figure 2-8 shows the RMSD of the 25 bearing signals. Compared with Chen's method, the predictive accuracy of the proposed method is significantly improved. Table 2-7 shows the comparison of the SPF method with Chen's method, the GLLR method [12] and an extension of SPF method with three change-points (SPF-CP3) or four phases in terms of RMSD at the three time steps. In the GLLR method, the first phase with normal working condition is manually truncated first, and the remaining data are fitted using Bayesian simple linear regression. Clearly, our method outperforms Chen's method and the GLLR method at all three prediction times. The GLLR method has the largest prediction error on this dataset. Comparing

SPF and SPF-CP3, we find that adding one more phase will reduce the prediction performance at the early prediction stage (50% and 70% prediction time) while slightly improve the accuracy at 90% prediction time. The reason is that the addition of another phase could introduce extra uncertainty in change-point prediction at early stages, while at the late stage, the model of the last segment of the four-phase model is more accurate in modeling the degradation signal.

Table 2-7: Comparison of the SPF method with other methods

Method	RMSD		
	50%	70%	90%
GLLR	356.5	234.2	227.8
Chen	318.4	156.9	169.4
SPF	176.4	110.8	56.9
SPF-CP3	186	112.5	52.4

2.6 CONCLUSION

In this chapter, we propose a multiple-phase modeling of degradation signals for health condition monitoring and remaining useful life prediction. To integrate the historical data with in-situ observations of each new unit in the RUL prediction, the multiple change-point model is formulated under the Bayesian framework and a novel stochastic process is proposed as priors of the formulated model. To facilitate the online monitoring and RUL prediction, the multiple change-point model is first represented by a novel nonstandard state-space model and then a new particle filtering algorithm is developed for online model updating and RUL prediction. A stratified sampling approach and a partial Gibbs resample-move strategy are developed to overcome the particle impoverishment problem and reduce the computational burden. The

advantages of the proposed method have been demonstrated through extensive numerical studies

and real case studies*.

Chapter 3: Exact Bayesian Inference for Prognostic Improvement

3.1 INTRODUCTION

In chapter 2, a novel particle filtering algorithm is developed under multiple change-point framework for condition monitoring and RUL prediction. A stratified sampling approach and a partial Gibbs resample-move strategy are developed to overcome the particle degeneracy and impoverishment problem. However, due to the increased dimensionality of multiple change-point model, the computational efficiency is unsatisfactory which may limit their applications in online monitoring and prognostics. In this chapter, we develop a new recursive updating and prediction algorithm with less computational cost by using exact Bayesian inference method. In this algorithm, the explicit analytical form of the posterior and RUL distribution are obtainable by using conjugate priors, which is more efficient than using particle filtering techniques to approximate the posterior and RUL distribution. In this method, the multiple change-point model is used to model the degradation path and the exact Bayesian inference is used to calculate the posterior distribution of the latest change-point and model parameters sequentially. A closed form of the RUL prediction is derived and a partial Monte Carlo approach is proposed to control the computational cost for further improvement.

The remainder of this chapter is organized as follows. In Section 3.2, a multiple change-point model for the CM signals is presented. The prior specification and parameter estimation for the multiple change-point model is given in Section 3.3. Section 3.4 presents the technical details on how to sequentially update the posterior distributions of all necessary model parameters and how to perform RUL prediction. Section 3.5 demonstrates the effectiveness and accuracy of the proposed method through numerical and case studies. The conclusion and discussion are given in Section 3.6.

3.2 MULTIPLE CHANGE-POINT GENERAL PATH MODELING

In this chapter, we still use genera path model, For simplicity, we reformulate the multiple change-point model Eq.(2.1) as follows

$$
y_{i,j} = \begin{cases}
\boldsymbol{X}_{i,j,1}\boldsymbol{\beta}_{i1} + \sigma_{i1}\varepsilon_{ij}, & \text{if } 0 < t_{i,j} \le t_{i,c_{i1}} \\
\boldsymbol{X}_{i,j,1}\boldsymbol{\beta}_{i2} + \sigma_{i2}\varepsilon_{ij}, & \text{if } t_{i,c_{i1}} < t_{i,j} \le t_{i,c_{i2}} \\
\qquad\qquad \cdots \\
\boldsymbol{X}_{i,j}\boldsymbol{\beta}_{ik_i} + \sigma_{ik_i}\varepsilon_{ij}, & \text{if } t_{i,c_{ik_i-1}} < t_{i,j} \le t_{i,c_{ik_i}} \\
\boldsymbol{X}_{i,j}\boldsymbol{\beta}_{ik_i+1} + \sigma_{ik_i+1}\varepsilon_{ij}, & \text{if } t_{i,c_{ik_i}} < t_{i,j} \le t_{i,n_i}
\end{cases}
\tag{3.1}
$$

where $\boldsymbol{X}_{i,j,s}$ is the vector of polynomial basis functions, i.e., $\boldsymbol{X}_{i,j,s} = [1, (t_{i,j} - t_{i,c_{is-1}}), \cdots, (t_{i,j} - t_{i,c_{is-1}})^{q_{is}}]$ where q_{is} is the polynomial order of the s-th segment, $\boldsymbol{\beta}_i^{(s)}$ is a vector of regression parameter and $\sigma_i^{2(s)}$ is noise variance of the s-th segment, and ε_{ij} is a noise term following *i.i.d.* standard normal distribution. Note that the order of polynomial regression could vary across different segments. Give the position of a change-point, we assume that the observations before that change-point is independent of those after the change-point. For simplicity and without loss of generality, we assume that $t_{i,j} = j$ in the rest of the chapter, i.e., the sampling intervals equal to 1 for all units. Besides, given the total number of change-points k_i, we assume that the polynomial orders of the $k_i + 1$ segments are deterministic.

As mentioned in chapter 2, the prognostics often involves two stages, namely, the offline stage for modeling and estimation, and the online stage for sequential model updating and RUL prediction. To characterize both the population trend and the individual heterogeneity, all the model parameters are assumed random in the offline modeling of the historical data. Denote a multiple change-point model as $\mathcal{M} = \left(k, \{\delta^{(s)}\}_{s=1}^{k+1}, \{\boldsymbol{\theta}^{(s)}\}_{s=1}^{k+1}\right)$ where k is the number of change-points, $\delta^{(s)} = c_s - c_{s-1}$ is the duration of the sth segment, and $\boldsymbol{\theta}^{(s)} = \left(\boldsymbol{\beta}^{(s)}, \sigma^{2(s)}\right)$ is

the model parameters of the sth segment. In the offline modeling, all these parameters are modelled with appropriate distributions, and the hyperparameters are estimated. The estimated distributions are then used as priors in the online Bayesian model updating and RUL prediction. At the online stage, the posterior distributions of the individual model parameters are sequentially updated. The total number of change-points, the index of the current segment (or how many change-points have occurred), and the latest change-point (or the duration since the latest change-point) are three key parameters in Bayesian model updating and RUL prediction.

Let $y_{1:t}$ denote the observations of a working unit up to the current time t, and $\boldsymbol{x}_t = (\boldsymbol{\theta}_t, \tau_t, s_t, k)$ be the state vector where $\boldsymbol{\theta}_t = (\boldsymbol{\beta}_t, \sigma_t^2)$ are the model parameters of the current segment, τ_t be the latest change-point that has occurred ($\tau_t \leq t - 1$), and s_t be index of the current segment, e.g., $s_t = 1, 2, \ldots, k + 1$. At the online stage, the posterior distribution $p(\boldsymbol{x}_t | y_{1:t})$ and the predictive density $p(y_{t+L} | y_{1:t})$ for integer $L > 0$ have to be calculated. However, these distributions are generally intractable. As mentioned earlier, although the particle filtering techniques or sequential Monte Carlo techniques are capable of handling these nonlinear intractable problems, their efficiency is significantly affected by their notorious particle degeneracy and impoverishment issues, and the enormous computational cost. In this chapter, we find that by using conjugate priors, the closed form of the posterior distributions and the predictive density can be recursively and efficiently calculated. The overall prognostic framework with offline and online stages is summarized in Figure 3-1. The following two sections give the technical details about the prior specification and parameter estimation at the offline stage, and the posterior distribution estimation and RUL prediction at the online stage.
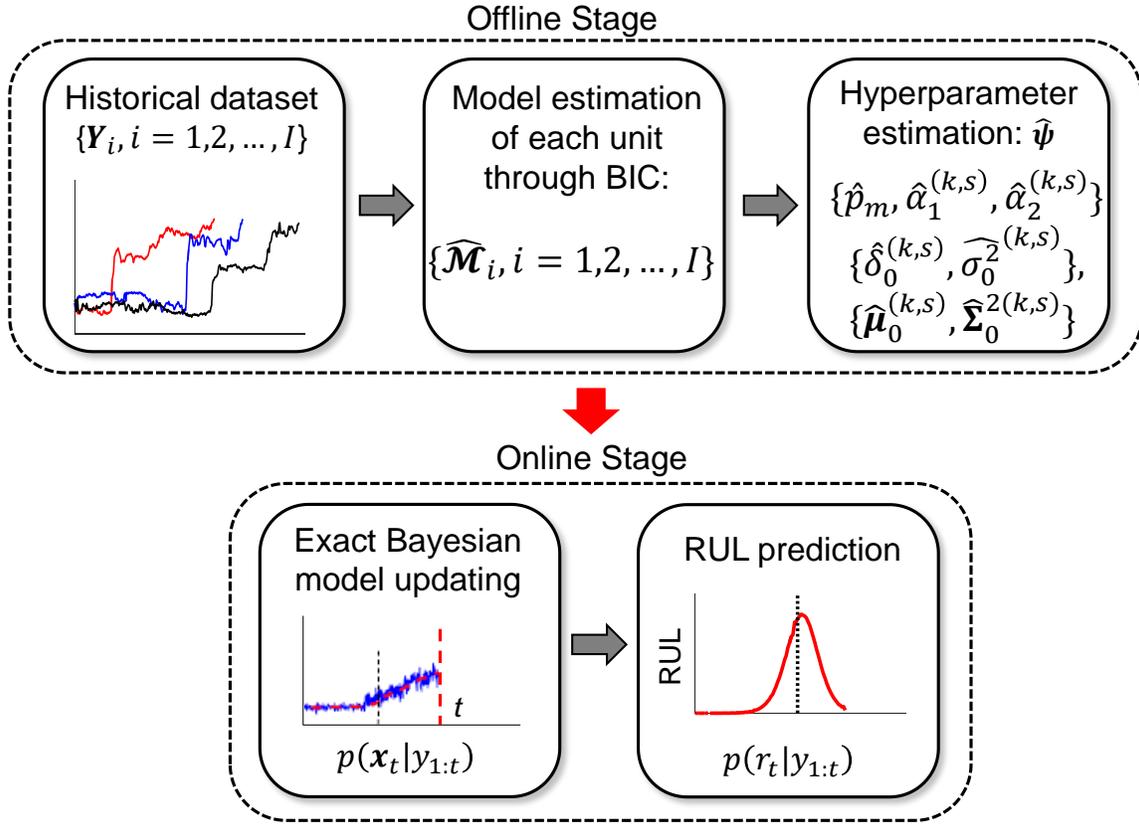
Figure 3-1: Illustration of the proposed prognostic framework.

3.3 PRIOR SPECIFICATION AND PARAMETER ESTIMATION

3.3.1 Specification of Priors

In this section, the priors for the multiple change-point model $\mathcal{M} = \left(k, \{\delta^{(s)}\}_{s=1}^{k+1}, \{\boldsymbol{\theta}^{(s)}\}_{s=1}^{k+1}\right)$ are specified and estimated. Follow the strategy in chapter 2, we use a nonhomogeneous Markov process where discrete distributions that are independent of the number of observations are selected for the number of change-points k, and then the positions of the change-points are modelled as a nonhomogeneous Markov process with durations between successive change-points depending on both k and the segment index s. The joint prior for both change-points and model parameters could be formulated as shown Eq.(2.7). We select a categorical distribution for k which is independent of n, i.e., $\pi(k = m) = p_m$ and $\sum_m p_m =$

1. The phase durations are approximately modelled with normal distributions, since they are more flexible in controlling mean and variance than most of the existing discrete probability distributions. Specifically, $\delta^{(s)}|k \sim N(\delta_0^{(k,s)}, \sigma_0^{2(k,s)})$. For the model parameters of each phase, in this chapter, the commonly used normal and inverse Gamma (IG) conjugate priors are assumed:

$$\pi(\boldsymbol{\beta}^{(s)}, \sigma^{2(s)}|k) = \pi(\sigma^{2(s)}|k)\pi(\boldsymbol{\beta}^{(s)}|\sigma^{2(s)}, k)$$

$$= IG(\sigma^{2(s)}|\alpha_1^{(k,s)}, \alpha_2^{(k,s)})N(\boldsymbol{\beta}^{(s)}|\boldsymbol{\mu}_0^{(k,s)}, \sigma^{2(s)}\boldsymbol{\Sigma}_0^{(k,s)}) \tag{3.2}$$

For notational convenience, in the rest of the chapter we use the double superscript (k, s) to denote the parameter or variable of s-th segment conditioning that there are in total k change-points, e.g., $\boldsymbol{\beta}^{(k,s)} = \boldsymbol{\beta}^{(s)}|k$.

3.3.2 Parameter Estimation from Historical Data

Informative priors can be obtained by estimating all the hyperparameters through historical data. Let $\boldsymbol{\psi}$ denote the vector of all hyperparameters, i.e., $\{p_m\}, \left\{\alpha_1^{(k,s)}, \alpha_2^{(k,s)}\right\}$ and Error! Bookmark not defined..

We use an empirical two-stage estimation method. Specifically, at the first stage, the model parameters $\widehat{\boldsymbol{M}}_i$ of each historical unit $i = 1, \ldots, I$ are obtained, we use the Bayesian information criterion (BIC) [84] for change-point model selection and the corresponding segment parameter estimation. For notational convenience, we ignore the subscript i for individual unit in the following four equations. The parameter estimation for each unit can be formulated as

$$\widehat{\boldsymbol{M}} = \arg\min_{\boldsymbol{M}}(-2l(\boldsymbol{M}|\boldsymbol{Y}) + K\log n) \tag{3.3}$$

where K is the total number of parameters, including change-points, regression parameters, and noise variances, and $l(\mathcal{M}|Y)$ is the log-likelihood function expressed by

$$l(\mathcal{M}|Y) = \sum_{s=1}^{k+1}\left[-\frac{1}{2}(c_s - c_{s-1})\log\left(2\pi\sigma^{2(k,s)}\right) - \frac{\left\|y_{c_{s-1}+1:c_s}^T - X_{1,c_s-c_{s-1}}\boldsymbol{\beta}^{(k,s)}\right\|^2}{2\sigma^{2(k,s)}}\right] \quad (3.4)$$

where $X_{1,c_s-c_{s-1}}$ is the design matrix, with X_{t_1,t_2} of order q defined as

$$X_{t_1,t_2} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_1+1 & \cdots & t_2 \\ \cdots & \cdots & \cdots & \cdots \\ t_1^q & (t_1+1)^q & \cdots & t_2^q \end{bmatrix}^T \quad (3.5)$$

Conditioning on the fixed change-points, i.e., $\left(k, \{\delta^{(k,s)}\}_{s=1}^{k+1}\right)$, the parameters of each segment that minimize Eq. (3.3) can be easily obtained through MLE of classical linear models

$$\begin{aligned} \hat{\boldsymbol{\beta}}^{(k,s)} &= \left(X_{1,c_s-c_{s-1}}^T X_{1,c_s-c_{s-1}}\right)^{-1} X_{1,c_s-c_{s-1}}^T y_{c_{s-1}+1:c_s}^T, \\ \widehat{\sigma^2}^{(k,s)} &= \left\|y_{c_{s-1}+1:c_s}^T - X_{1,c_s-c_{s-1}}\hat{\boldsymbol{\beta}}^{(k,s)}\right\|^2 / \delta^{(k,s)} \end{aligned} \quad (3.6)$$

Therefore, for each possible model defined by change-points, the BIC value can be easily evaluated. Suppose the estimated parameters are $\widehat{\mathcal{M}}_i = \left(\hat{k}_i, \left\{\hat{\delta}_i^{(\hat{k}_i,s)}\right\}_{s=1}^{\hat{k}_i}, \left\{\hat{\boldsymbol{\beta}}_i^{(\hat{k}_i,s)}, \widehat{\sigma^2}_i^{(\hat{k}_i,s)}\right\}_{s=1}^{\hat{k}_i+1}\right)$

for $i = 1, \dots, I$. The second stage is to estimate the hyperparameters based on $\widehat{\mathcal{M}}_i, i = 1, \dots, I$ through MLE approach. The MLE of hyperparameters p_m can be easily obtained as

$$\hat{p}_m = \frac{1}{I}\sum_{i=1}^{I} \mathbf{1}_{\hat{k}_i=m} \quad (3.7)$$

For the hyperparameters $(\alpha_1^{(k,s)}, \alpha_2^{(k,s)})$ in the inverse Gamma distribution, the MLE using observations $\left\{\widehat{\sigma^2}_i^{(\hat{k}_i,s)} \middle| \hat{k}_i = k, i = 1, \ldots, I\right\}$ can be estimated numerically through various optimization algorithms. The MLE of the hyperparameters $(\delta_0^{(k,s)}, \sigma_0^{2(k,s)})$ for the segment duration can be obtained straightforwardly as

$$\hat{\delta}_0^{(k,s)} = \frac{\sum_{i=1}^I \hat{\delta}_i^{(\hat{k}_i,s)} \mathbf{1}_{\hat{k}_i=k}}{\sum_{i=1}^I \mathbf{1}_{\hat{k}_i=k}}$$

$$\widehat{\sigma_0^2}^{(k,s)} = \frac{\sum_{i=1}^I \left(\hat{\delta}_i^{(\hat{k}_i,s)} - \hat{\delta}_0^{(k,s)}\right)^2 \mathbf{1}_{\hat{k}_i=k}}{\sum_{i=1}^I \mathbf{1}_{\hat{k}_i=k}}$$

(3.8)

For the hyperparameters $\boldsymbol{\mu}_0^{(k,s)}$ and $\boldsymbol{\Sigma}_0^{(k,s)}$, closed forms can be derived as

$$\hat{\boldsymbol{\mu}}_0^{(k,s)} = \frac{\sum_{i=1}^I \dfrac{\hat{\boldsymbol{\beta}}_i^{(\hat{k}_i,s)} \mathbf{1}_{\hat{k}_i=k}}{\widehat{\sigma^2}_i^{(\hat{k}_i,s)}}}{\sum_{i=1}^I \dfrac{\mathbf{1}_{\hat{k}_i=k}}{\widehat{\sigma^2}_i^{(\hat{k}_i,s)}}}$$

(3.9)

$$\hat{\boldsymbol{\Sigma}}_0^{(k,s)} = \frac{\sum_{i=1}^I \left(\hat{\boldsymbol{\beta}}_i^{(\hat{k}_i,s)} - \hat{\boldsymbol{\mu}}_0^{(k,s)}\right)\left(\hat{\boldsymbol{\beta}}_i^{(\hat{k}_i,s)} - \hat{\boldsymbol{\mu}}_0^{(k,s)}\right)^T \mathbf{1}_{\hat{k}_i=k}}{\sum_{i=1}^I \mathbf{1}_{\hat{k}_i=k}}$$

The details of the derivation can be found in Appendix B.

3.4 EXACT BAYESIAN ONLINE MODEL UPDATING AND RUL PREDICTION

In this section we will discuss how to update the model sequentially through exact Bayesian inference and predict the RUL for a new in-service unit during the online stage.

3.4.1 Exact Bayesian Online Model Updating

The model updating of a working unit is an essential step for health condition monitoring and RUL prediction. It refers to the posterior distribution evaluation of all model parameters that

could capture the current health condition and future degradation evolution. In this chapter, the model updating is to calculate the posterior distribution $p(x_t|y_{1:t})$ where the state vector $x_t = (\theta_t, \tau_t, s_t, k)$. Generally, this posterior distribution is intractable and sequential Monte Carlo techniques are needed. Fortunately, due to the assignment of conjugate priors for θ, the posterior could be recursively calculated, which is shown as follows.

To calculate the joint posterior $p(x_t|y_{1:t})$, we first calculate the posterior distribution of the discrete components $P(\tau_t, s_t, k|y_{1:t})$, and then calculate the posterior distribution of the continuous components conditioning on the discrete components, i.e., $p(\theta_t|\tau_t, s_t, k, y_{1:t})$. That is

$$p(x_t|y_{1:t}) = P(\tau_t, s_t, k|y_{1:t})p(\theta_t|\tau_t, s_t, k, y_{1:t}) \tag{3.10}$$

The conditional posterior distribution of the continuous components $p(\theta_t|\tau_t, s_t, k, y_{1:t})$ can be calculated based on Theorem 3.1 as follows.

Theorem 3.1 Suppose the conjugate prior in Eq.(3.2) is assigned to β_t and σ_t^2.

$$\pi(\beta_t, \sigma_t^2|s_t = s, k) = IG\left(\sigma_t^2 \middle| \alpha_1^{(k,s)}, \alpha_2^{(k,s)}\right) N\left(\beta_t \middle| \mu_0^{(k,s)}, \sigma_t^2 \Sigma_0^{(k,s)}\right)$$

Then

$$(\sigma_t^2|\tau_t = j, s_t = s, k, y_{1:t}) \sim IG\left(\alpha_1^{(k,s)} + \frac{t-j}{2}, \alpha_2^{(k,s)} + \frac{H_{j+1,t}^{(k,s)}}{2}\right) \tag{3.11}$$

$$(\beta_t|\sigma_t^2, \tau_t = j, s_t = s, k, y_{1:t}) \sim N(\mu_{j+1,t}, \sigma_t^2 \Sigma_{j+1,t})$$

where

$$\Sigma_{j+1,t} = \left(X_{1,t-j}^T X_{1,t-j} + \left(\Sigma_0^{(k,s)}\right)^{-1}\right)^{-1}$$

$$N_{j+1,t} = \left(\left(\Sigma_0^{(k,s)}\right)^{-1} \mu_0^{(k,s)} + X_{1,t-j}^T y_{j+1:t}\right) \tag{3.12}$$

$$\mu_{j+1,t} = \Sigma_{j+1,t} N_{j+1,t}$$

$$H_{j+1,t}^{(k,s)} = y_{j+1:t}^T y_{j+1:t} + \left(\boldsymbol{\mu}_0^{(k,s)}\right)^T \left(\boldsymbol{\Sigma}_0^{(k,s)}\right)^{-1} \boldsymbol{\mu}_0^{(k,s)} - N_{j+1,t}^T \boldsymbol{\Sigma}_{j+1,t} N_{j+1,t}$$

The proof of Theorem 3.1 can be found in Appendix C. The calculation of $P(\tau_t, s_t, k|y_{1:t})$ is the main challenge. It can be recursively updated as

$$P(\tau_t = j, s_t = s, k|y_{1:t})$$

$$\propto P(\tau_t = j, s_t = s, k|y_{1:t-1})p(y_t|\tau_t = j, s_t = s, k, y_{1:t-1}) \tag{3.13}$$

Eq. (3.13) consists of two parts. The first part is the predictive probability mass function (PMF) $P(\tau_t = j, s_t = s, k|y_{1:t-1})$ which can be recursively calculated by

$$P(\tau_t = j, s_t = s, k|y_{1:t-1})$$

$$= \sum_{j',s'} P(\tau_{t-1} = j', s_{t-1} = s', k|y_{1:t-1})P(\tau_t = j, s_t = s|\tau_{t-1} = j', s_{t-1}$$

$$= s', k, y_{1:t-1}) \tag{3.14}$$

where $P(\tau_{t-1} = j', s_{t-1} = s', k|y_{1:t-1})$ is the posterior distribution obtained at the previous time step, and $P(\tau_t = j, s_t = s|\tau_{t-1} = j', s_{t-1} = s', k, y_{1:t-1})$ is the predictive Markov transition probability. Based on the specified nonhomogeneous Markov process for change-points, this predictive Markov transition probability can be derived as

$$P(\tau_t = j, s_t = s|\tau_{t-1} = j', s_{t-1} = s', k, y_{1:t-1})$$

$$= \begin{cases} \dfrac{1 - G^{(k,s')}(t-j')}{1 - G^{(k,s')}(t-j'-1)}, & \text{if } j = j' \text{ and } s = s' < k+1 \\ 1, & \text{if } j = j' \text{ and } s = s' = k+1 \\ \dfrac{G^{(k,s')}(t-j') - G^{(k,s')}(t-j'-1)}{1 - G^{(k,s')}(t-j'-1)}, & \text{if } j = t-1 \text{ and } s = s'+1 \le k+1 \\ 0, & \text{otherwise} \end{cases} \tag{3.15}$$

Where $G^{(k,s')}(\cdot)$ is the cumulative distribution function of s'-th segment duration for a CM signal with k change-points.

The second part of Eq. (3.13) is $p(y_t|\tau_t = j, s_t = s, k = m, y_{1:t-1})$, namely, the predictive density function of y_t. It is the only term that involves the newest observation y_t in the posterior updating of the discrete components. Once this density function is known, the posterior distribution of the discrete components can be recursively updated based on Eq. (3.13), (3.14) and (3.15). Therefore this part is critically important. It can be calculated based on Theorem 3.2 as follows.

Theorem 3.2 Denote $\boldsymbol{X}_t = [1, t, \cdots, t^q]$ where q is the polynomial order, then if $j < t - 1$,

$$\left(y_t|\tau_t = j, s_t = s, k, y_{j+1:t-1}\right) \sim$$

$$t_1\left(2\alpha_1^{(k,s)} + t - j - 1, \boldsymbol{X}_{t-j}\boldsymbol{\mu}_{j+1,t-1}, \frac{2\alpha_2^{(k,s)} + H_{j+1:t-1}^{(k,s)}}{2\alpha_1^{(k,s)} + t - j - 1}\left(1 + \boldsymbol{X}_{t-j}\boldsymbol{\Sigma}_{j+1,t-1}\boldsymbol{X}_{t-j}^T\right)\right) \tag{3.16}$$

And if $j = t - 1$,

$$(y_t|\tau_t = t - 1, s_t = s, k) \sim t_1\left(2\alpha_1^{(k,s)}, \boldsymbol{X}_1\boldsymbol{\mu}_0^{(k,s)}, \frac{\alpha_2^{(k,s)}}{\alpha_1^{(k,s)}}\left(1 + \boldsymbol{X}_1\boldsymbol{\Sigma}_0^{(k,s)}\boldsymbol{X}_1^T\right)\right) \tag{3.17}$$

The proof of Theorem 3.2 is given in Appendix D.

3.4.2 RUL Prediction

RUL prediction is to determine the time when the signal first hit the failure threshold Γ. For an operating unit, denote the remaining useful life as R_t at current time t. Then R_t can be defined as $R_t = \inf\{L: y_{t+L} \geq \Gamma | y_{1:t}\}$. The cumulative distribution function (CDF) of R_t conditional on available observations $y_{1:t}$ (also called conditional reliability function) can be expressed as

$$P(R_t > L|y_{1:t}) = \sum_k \sum_s \sum_j P(R_t > L|\tau_t = j, s_t = s, k, y_{1:t}) P(\tau_t = j, s_t = s, k|y_{1:t}) \quad (3.18)$$

where $P(\tau_t, s_t, k|y_{1:t})$ is recursively calculated through Eq. (3.13) in the model updating stage,

and $P(R_t > L|\tau_t, s_t, k, y_{1:t})$ can be reformulated as

$$P(R_t > L|\tau_t = j, s_t = s, k, y_{1:t})$$
$$= P(y_{t+1} < \Gamma, \dots, y_{t+L} < \Gamma|\tau_t = j, s_t = s, k, y_{1:t}) \quad (3.19)$$

The calculation of Eq. (3.19) depends on the segment index s_t. If $s_t = k + 1$, or the

degradation process is at the final segment, it can be shown that the vector $y_{t+1:t+L}$ follows a

multivariate t distribution of dimension L, as shown in Eq. (3.20) of Theorem 3.3.

Theorem 3.3 If $s_t = k + 1$,

$$\left(y_{t+1:t+L}|\tau_t = j, s_t = k + 1, k, y_{j+1:t}\right)$$

$$\sim t_L\left(2\alpha_1^{(k,k+1)} - j, X_{t+1-j,t+L-j}\mu_{j+1,t}, \frac{2\alpha_2^{(k,k+1)} + H_{j+1:t}^{(k,k+1)}}{2\alpha_1^{(k,k+1)} + t - j}\left(I + X_{t+1-j,t+L-j}\Sigma_{j+1,t}X_{t+1-j,t+L-j}^T\right)\right) \quad (3.20)$$

and

$$\left(y_{t+1:t+L}|\tau_{t+1} = j, s_{t+1} = k + 1, k\right)$$

$$\sim t_L\left(2\alpha_1^{(k,k+1)}, X_{t+1-j,t+L-j}\mu_0^{(k,k+1)}, \frac{\alpha_2^{(k,k+1)}}{\alpha_1^{(k,k+1)}}\left(I + X_{t+1-j,t+L-j}\Sigma_0^{(k,k+1)}X_{t+1-j,t+L-j}^T\right)\right) \quad (3.21)$$

The proof of Theorem 3.3 is similar to Theorem 3.2 and thus is not provided here. Based on

Theorem 3.3, if $s_t = k + 1$, i.e., the degradation is at the final stage, $P(R_t > L|\tau_t = j, s_t = k +$

$1, k, y_{1:t}) = \text{MT}_{t+1,t+L}(\Gamma)$ where $\text{MT}_{t+1:t+L}(\Gamma)$ is the CDF of L-dimensional t distribution

given in Eq. (3.20). If $s_t < k + 1$, however, the future change-points, especially the final

change-point, need to be predicted for RUL prediction. The calculation for the general case is

derived as follows. Denote the last or final change-point as c_k, then

58

$$P(R_t > L | \tau_t, s_t, k, y_{1:t}) = \sum_{c_k} P(R_t > L | c_k, \tau_t, s_t, k, y_{1:t}) P(c_k | \tau_t, s_t, k, y_{1:t}) \qquad (3.22)$$

where $P(c_k | \tau_t, s_t, k, y_{1:t})$ is the predictive PMF of the final change-point, which does not depend $y_{1:t}$ and thus can be recursively calculated in the off-line stage based on the defined Markov transition process in Eq.(3.15), and $P(R_t > L | c_k, \tau_t, s_t, k, y_{1:t})$ can be calculated based the model assumption that the degradation signal will not exceed the failure threshold before it reaches the final segment:

$$P(R_t > L | c_k, \tau_t, s_t, k, y_{1:t}) = \begin{cases} \mathrm{MT}_{t+1,t+L}(\Gamma), & \text{if } c_k = \tau_t \leq t - 1 \\ \mathrm{MT}_{c_k+1,t+L}(\Gamma), & \text{if } t - 1 < c_k < t + L \\ 1, & \text{if } c_k \geq t + L \end{cases} \qquad (3.23)$$

In Eq. (3.23), $\mathrm{MT}_{t+1,t+L}(\Gamma)$ is the CDF of t distribution given in Eq. (3.20) while $\mathrm{MT}_{c_k+1,t+L}(\Gamma)$ is the CDF of $\left(y_{c_k+1:t+L} | \tau_{c_k+1} = c_k, s_{c_k+1} = k + 1, k \right)$ given in Eq. (3.21). As we can see, the closed form of the conditional reliability function or the RUL distribution can also be exactly obtained based on Eq. (3.18)-(3.23).

3.4.3 Computational Issue and Approximation

Although the model updating and RUL prediction can be exactly calculated through recursion, both the computational and memory cost of each time step increase with time t. From Eq. (3.13)-(3.15) we can see that the computational and memory cost of the filtering recursion at time t is approximately linear with time t, since we need to calculate and store $t \sum_k (k + 1)$ probabilities for the posterior PMF $P(\tau_t = j, s_t = s, k | y_{1:t})$. In the RUL prediction, for each L in Eq. (3.18), the computational cost also increases with t. For large datasets, these computational and storage issues may become very prohibitive in real time applications and thus need to be solved. In practice, the posterior PMF $P(\tau_t = j, s_t = s, k | y_{1:t})$ is almost zero at most of the support points. Indeed, with more observations obtained in the current segment, the

posterior PMF would concentrate around the starting point c_{s-1} of the current segment, and for $\tau_t \ll c_{s-1}$, the posterior PMF is close to zero and thus can be negligible. Similar phenomenon can also be observed for s_t and k. A natural way to control the computational cost and memory issue is to approximate the posterior densities at each time step with a small set of support points of fixed size that have high probabilities, and set the posterior PMF to be zero at the remaining support points. However, this strategy may result in inaccurate approximation if directly applied to the 3-dimensional support points. The PMF can be temporally near zero for certain s_t and k, and then becomes dominant later as more observations are obtained. If the PMF is set zero at these support points, the PMF of these support points evaluated at the future time steps will also be zero, thus leading to an inaccurate approximation. To address this issue, we propose to select N most probable support points under each stratum (s_t, k) to approximate the posterior and set others to zero. Consequently, there are in total $N \sum_k (k + 1)$ non-zero support points. The details of the approximation algorithm are summarized in Table 3-1.

Table 3-1: Summary of the approximation updating algorithm

1. At time step $t = N + 1$,
   - Calculate $P(\tau_t = j, s_t = s, k|y_{1:t})$ at all the $(N + 1)\sum_k (k + 1)$ support points.
   - Within each stratum (s, k), select N time steps from $\{1, \dots, t\}$ with highest PMF $P(\tau_t = j, s, k|y_{1:t})$. Denote the selected time steps as $\boldsymbol{T}_N(t, s, k)$.
   - Normalize the probabilities of the selected $N \sum_k (k + 1)$ support points.
2. At time step $t > N + 1$
   - Calculate $P(\tau_t = j, s, k|y_{1:t})$ at $(N + 1)$ support points $\{\boldsymbol{T}_N(t - 1, s, k), t\}$ for each stratum (s, k).
   - Within each stratum (s, k), select N time steps from $\{\boldsymbol{T}_N(t - 1, s, k), t\}$ with highest PMF $P(\tau_t = j, s_t = s, k|y_{1:t})$. Update $\boldsymbol{T}_N(t, s, k)$.
   - Normalize the probabilities of the selected $N \sum_k (k + 1)$ support points.

Another computational issue is the calculation of the conditional reliability function through Eq. (3.18) and (3.22) at the online RUL prediction stage. It involves a large number of evaluations of the CDF of multivariate t distribution, e.g., $MT_{t+1,t+L}(\Gamma)$, whose computational cost increases enormously with the dimension L. To control the computational cost, we could alternatively use the Monte Carlo (MC) simulation approach. One MC approach is to directly generate the samples for the current state vector x_t through the posterior distribution $p(x_t|y_{1:t})$ and simulate future state vectors $x_{t+L}, L = 1,2,\cdots,$ through the prior Markov state transition process. Specifically, to simulate x_{t+1} conditioning on the previous state x_t, we could first simulate the discrete components of x_{t+1} through Eq.(3.15), and then simulate the continuous component θ_{t+1} from the prior distribution if t is a change-point or let $\theta_{t+1} = \theta_t$ if t is not a change-point. Based on these simulated samples of current state and future states, the conditional reliability function can be easily calculated. However, due to the high dimensionality of the state vector, this MC approach requires a large number of samples to guarantee the approximation accuracy. To solve this issue, we propose to use partial MC simulation for only the calculation of $P(R_t > L|c_k, \tau_t, s_t, k, y_{1:t})$. Note that in Eq. (3.18) and (3.22), $P(c_k|\tau_t, s_t, k, y_{1:t})$ is independent of $y_{1:t}$ and can be calculated at the offline stage. $P(R_t > L|c_k, \tau_t, s_t, k, y_{1:t})$ is also independent of $y_{1:t}$ when $c_k \geq t$, and thus can also be calculated at the offline stage. Therefore we only need to estimate $P(R_t > L|c_k, \tau_t, s_t, k, y_{1:t})$ for $c_k = \tau_t \leq t - 1$ at the online RUL prediction stage. To estimate it, we generate S samples $\{\theta_{t1}, \ldots, \theta_{tS}\}$ from the posterior distribution $P(\theta_t|c_k, \tau_t, s_t, k, y_{1:t})$ given in Eq. (3.11), and then $P(R_t > L|c_k, \tau_t, s_t, k, y_{1:t})$ can be estimated by

$$P(R_t > L|c_k, \tau_t, s_t, k, y_{1:t}) = \frac{1}{S}\sum_{i=1}^{S}\prod_{l=1}^{L}\Phi(\Gamma|X_{t+l-c_k}\beta_{tl}, \sigma_{tl}^2) \qquad (3.24)$$

61

where $\Phi(\cdot | X_{t+l-c_k}\boldsymbol{\beta}_{tl}, \sigma_{tl}^2)$ is the CDF of Gaussian distribution with mean $X_{t+l-c_k}\boldsymbol{\beta}_{tl}$ and variance σ_{tl}^2. This strategy can significantly improve the computational efficiency yet without influencing the calculation accuracy. Another advantage of this strategy is that we can easily adopt truncated distributions for the last segment (e.g., only select samples with positive degradation rate) to avoid the occurrence of a temporary decreasing trend (unit would never fail and RUL prediction is infeasible) in the updated signal evolution path for RUL prediction, which is common in real degradation signals due to measurement noises [95]. Note that in partial MC method, only $P(R_t > L | c_k, \tau_t, s_t, k, y_{1:t})$ with $c_k = \tau_t \leq t - 1$ (i.e., degradation is at the final phase) is calculated by MC method. The approximation accuracy decreases with L for a fixed sample size. However, in many practical applications, the final phase is often very steep and short in duration, so that $P(R_t > L | c_k, \tau_t, s_t, k, y_{1:t})$ converges rapidly to 0 as L increases. Therefore, using a small sample size can often achieve a very accurate approximation.

3.5 CASE STUDIES

In this section, we first use simulated signals to illustrate the effectiveness of the proposed method, and then apply the method to vibrational signals of rotational bearings for performance evaluation.

3.5.1 Simulation Study

To demonstrate the effectiveness of the proposed algorithm, a simulation study is conducted. In the simulation model, we assume that there are two types of degradation signals in terms of the number of change-points, namely, one-change-point (one-CP) and two-change-points (two-CP) signals. The corresponding probabilities are given by

$$k = \begin{cases} 1, \text{with } p_1 = 0.2 \\ 2, \text{with } p_2 = 0.8 \end{cases} \tag{3.25}$$

Table 3-2: Hyperparameters for the Bayesian multiple change-point model

Variables	$k = 1$	$k = 2$
$\delta^{(k,s)}$	$\delta_0^{(1,1)} = 500, \sigma_0^{2(1,1)} = 30^2$ $\delta_0^{(1,2)} = 500, \sigma_0^{2(1,2)} = 30^2$	$\delta_0^{(2,1)} = 500, \sigma_0^{2(2,1)} = 50^2$ $\delta_0^{(2,2)} = 400, \sigma_0^{2(2,2)} = 30^2$ $\delta_0^{(2,3)} = 500, \sigma_0^{2(2,3)} = 30^2$
$\boldsymbol{\beta}^{(k,s)}$	$\boldsymbol{\mu}_0^{(1,1)} = [-10; 0.005]$ $\boldsymbol{\Sigma}_0^{(1,1)} = \begin{bmatrix} 0.2 & 0.0015 \\ 0.0015 & 0.0008 \end{bmatrix}$ $\boldsymbol{\mu}_0^{(1,2)} = [-40; 0.05]$ $\boldsymbol{\Sigma}_0^{(1,2)} = \begin{bmatrix} 0.5 & -0.0008 \\ -0.0008 & 0.006 \end{bmatrix}$	$\boldsymbol{\mu}_0^{(2,1)} = [-10; 0.003]$ $\boldsymbol{\Sigma}_0^{(2,1)} = \begin{bmatrix} 0.15 & 0.0014 \\ 0.0014 & 0.0009 \end{bmatrix}$ $\boldsymbol{\mu}_0^{(2,2)} = [-20; 0.02]$ $\boldsymbol{\Sigma}_0^{(2,2)} = \begin{bmatrix} 0.024 & -0.0009 \\ -0.0009 & 0.000055 \end{bmatrix}$ $\boldsymbol{\mu}_0^{(2,3)} = [-30; 0.09]$ $\boldsymbol{\Sigma}_0^{(2,3)} = \begin{bmatrix} 0.75 & -0.00008 \\ -0.00008 & 0.00045 \end{bmatrix}$
$\sigma^{2(k,s)}$	$\alpha_1^{(1,1)} = 2, \alpha_2^{(1,1)} = 4$ $\alpha_1^{(1,2)} = 1, \alpha_2^{(1,2)} = 4$	$\alpha_1^{(2,1)} = 3.6, \alpha_2^{(2,1)} = 3$ $\alpha_1^{(2,2)} = 3, \alpha_2^{(2,2)} = 5$ $\alpha_1^{(2,3)} = 3.6, \alpha_2^{(2,3)} = 5$

All the segments are modelled as simple linear regression or line segment for both one-CP and two-CP signals. The other hyperparameters of the Bayesian multiple change-point model are given in Table 3-2. In total $I = 300$ signals are simulated as the historical dataset. The failure threshold is set as $\Gamma = 30$. Based on simulation settings above, 64 signals are generated for $k = 1$ and 236 signals are generated for $k = 2$. The hyperparameters are then estimated by the empirical two-stage estimation approach introduced in section 3.3.2. The proposed method can detect the change points (the number and positions) accurately for each signal. Due to page

limitation, the estimated hyperparameters are not listed here. For illustration, we also consider another model where all the settings are the same with the model described above except that all the last segments are quadratic. The hyperparameters of this model are not provided here due to space limitation.

Figure 3-2 illustrates the online condition monitoring and model updating of individual unit for these two degradation models. Unless otherwise specified, the support size $N = 5$ for each stratum (s_t, k) for posterior approximation in this paper. Clearly, the sequentially calculated posterior distribution of the discrete components of the state vector could effectively detect the occurrence of change-points, track the index of the current segment, and infer how many segments the CM signal would have. The accurate estimation of the CM signals also indicates an effective updating of the posterior distribution of the continuous components or model parameters of each segment. Figure 3-3 shows an example of predicting the position of the final change-point at two different times using the degradation signal shown in Figure 3-2(b). At $t = 300$, the degradation signal is within its first segment. Hence, it is high likely that there are two future change-points to be predict, which increases the prediction uncertainty. At $t = 800$, the degradation signal has already transited to the second segment, and the final change-point can be accurately predicted. Clearly, as more data are observed, the prediction of the final change-point becomes more accurate.
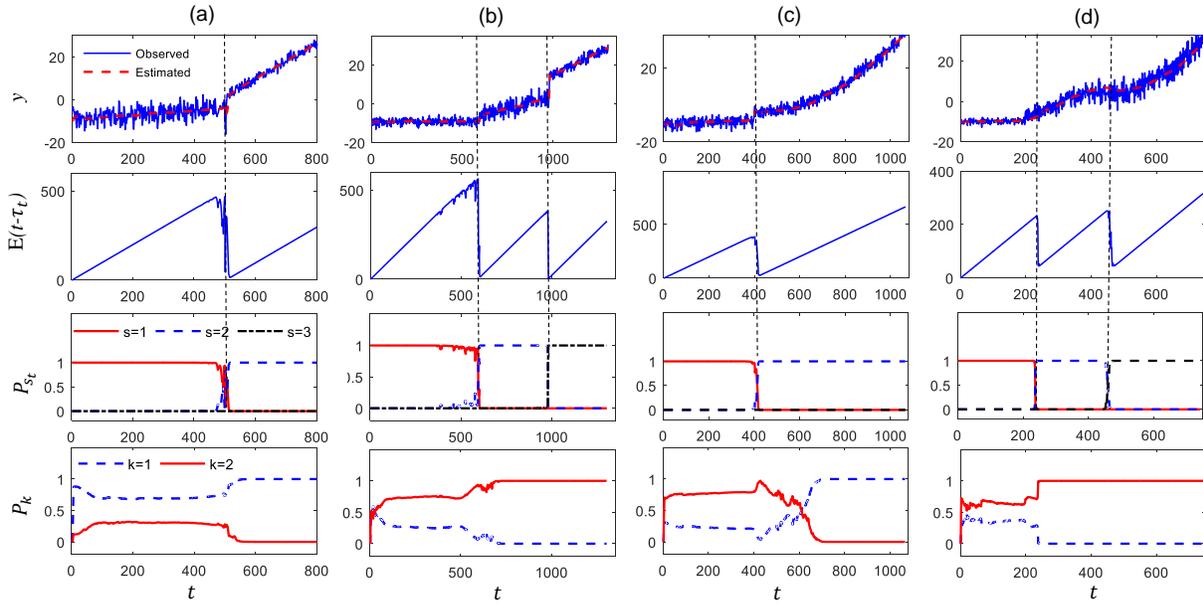
Figure 3-2: Illustration of the online monitoring and model updating for two models. (a) and (b): all segments are simple linear models; (c) and (d): all the last segments are quadratic; first row: raw and estimated or filtered CM signals; second row: the expected duration of the current segment; third row: the posterior PMF of the index of current segment; and bottom row: the posterior PMF of the total number of segments (or signal type). The vertical dashed lines are true change-points.
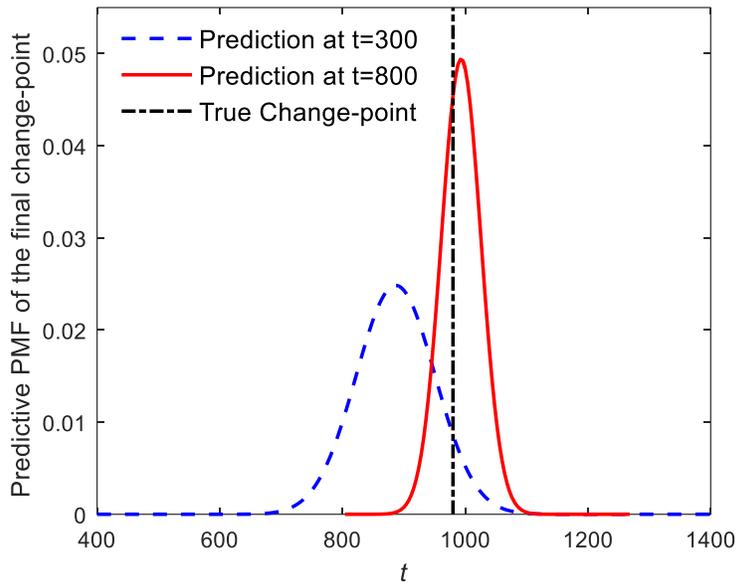
Figure 3-3: Predictive PMF of the final change-point with prediction time $t = 300$ and $t = 800$.
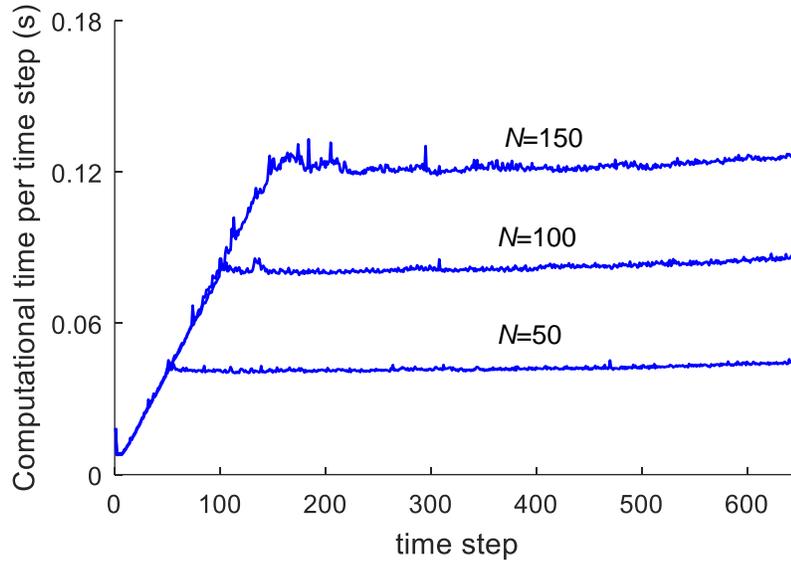
Figure 3-4: Computational time per time step for three different support sizes N in the model updating process.

Figure 3-4 shows the computational cost of each time step in the posterior model updating process using three different support sizes $N = 50, 100$ and 150. As we can clearly see, when $t < N$, the computational cost linearly increases with t, which is consistent with what we discussed in Section 3.4.3. Once the approximation strategy with a fixed support size N is applied, the computational cost of each step is fully controlled for $t \geq N$ with an almost constant computational time.

To evaluate the performance of RUL prediction, another 100 CM signals are simulated as testing dataset. The proposed method (denote it as EB) is compared with Chen's method [63], which models CM signals with two line segments and thus is an ideal method for comparison. For Chen's method, all 300 training CM signals (236 signals with three line segments and 64 with two line segments) are used to estimate the hyperparameters of the two line segments. For the EB method, the support size $N = 5$.

66
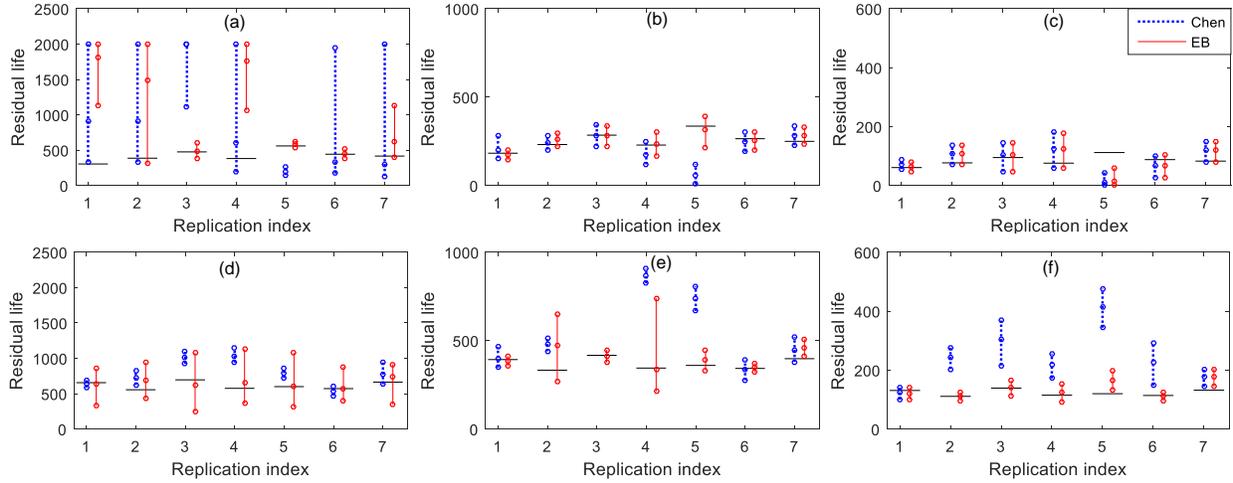
Figure 3-5: Prediction intervals for 7 one-CP signals (a, b and c) and 7 two-CP signals (d, e and f) at three different prediction times. (a) and (d): 50% of failure time; (b) and (e): 70% of failure time; (c) and (f): 90% of failure time. The ∘ represents the 5%, 50%, 95% quantiles of the predicted RUL distributions, and – denotes the actual RUL.
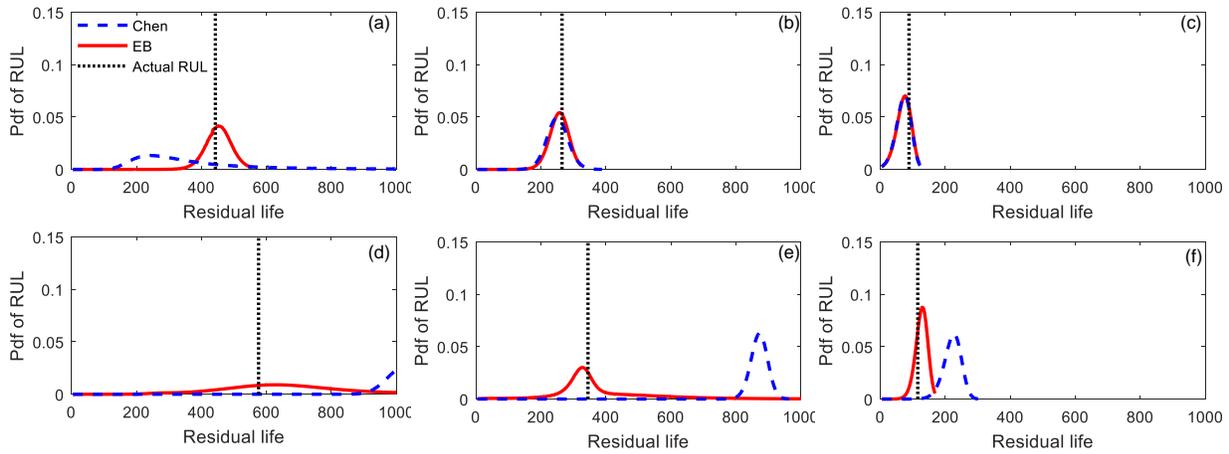
Figure 3-6: Comparison of the detailed pdf of the RUL. (a), (b) and (c) correspond to the 6th signal (one-CP) of Figure 3-5 (a), (b) and (c), respectively; (d), (e) and (f) correspond to the 4th signal (two-CP) of Figure 3-5 (d), (e) and (f) respectively

Figure 3-5 shows the prediction intervals of EB and Chen's method at three prediction times for 7 one-CP signals and 7 two-CP signals randomly selected from the testing dataset. Figure 3-6 shows the detailed pdf of the predicted RUL for the 6th signal of Figure 3-5(a-c) and the 4th signal of Figure 3-5(d-f). Unsurprisingly, the prediction for both methods becomes more and more accurate as more observations are available. Comparing these two methods, the

67

proposed EB method outperforms Chen's method for almost all the 14 signals. For one-CP signals, the EB method is slightly better at 70% and 90% of failure time, while at the prediction time 50%, the advantage of EB method is much more significant. The reason is that at the early degradation stage, the priors play a decisive role on the prediction accuracy. In Chen's method, the priors of two-line-segment model are estimated using all one-CP and two-CP signals, which results in inaccurate priors. At 70% and 90% of failure time, all degradation signals evolve into the second segment, and the posterior distribution of each model is dominated by the observations. Therefore, Chen's method can also accurately predict the RUL with inaccurate priors. For two-CP signals, the EB method is much better than Chen's method, which fulfills the purpose of the proposed method for degradation signals with two or even more segments.

To better quantify the prediction performance, the $\alpha - \lambda$ performance metric is calculated [96], where α specifies the error bound on the estimated RUL, i.e., $[(1 - \alpha)]R_{i,true} \leq \hat{R}_i \leq [1 + \alpha]R_{i,true}$, and λ specifies the relative distance, in time, of a given prediction point from the actual failure time, i.e., $\lambda = 0$ and $\lambda = 1$ correspond to the starting prediction time and the actual failure time, respectively. Figure 3-7 shows the $\alpha - \lambda$ performance metric for the 6th one-CP signal and 4th two-CP signal using the proposed EB method and Chen's method. The error bound α is set as 20%. It can be observed that, almost all the estimated RULs lie within the error bound for one-change-point case using both methods. However, for two-change-points case, Chen's method is much worse than the proposed EB method.
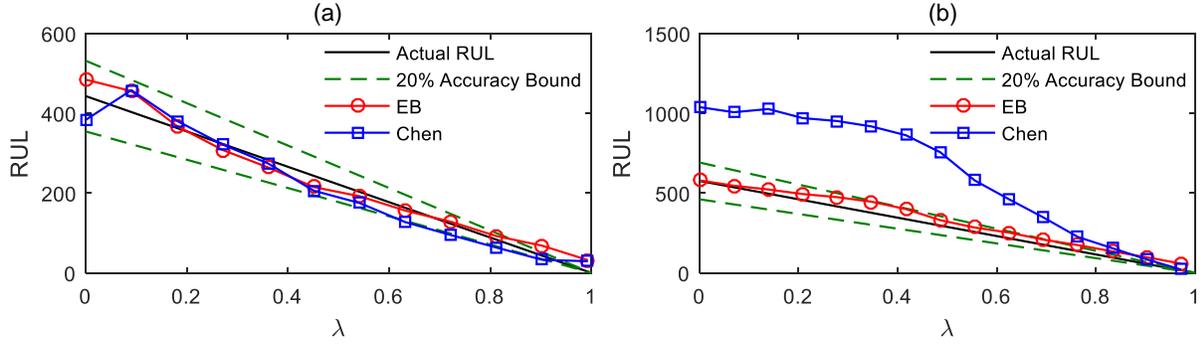
Figure 3-7: $\alpha - \lambda$ performance metric for (a) the 6th signal for one-CP case, and (b) the 4th signal for two-CP case.

To evaluate the overall performance of RUL prediction, we use the root-mean-square-deviation (RMSD) for these 100 testing signals defined as

$$\text{RMSD} = \sqrt{\frac{1}{100}\sum_{i=1}^{100} E\left(\hat{R}_i - R_{i,\text{true}}\right)^2} \tag{3.26}$$

where $\hat{R}_i$ and $R_{i,\text{true}}$ are the predicted true RUL of unit i, respectively.

Table 3-3: Comparison of the RMSD at six prediction times

Method	RMSD					
	40%	50%	60%	70%	80%	90%
Chen	319.5	271.6	218.3	240.6	211.0	102.0
EB	313.8	188.8	153.8	114.4	35.1	23.0
EB-partial MC	318.0	203.7	149.6	103.8	37.7	31.3

Table 3-3 shows the RMSD of the proposed EB method and Chen's method at six prediction times. We also add the EB method with partial Monte Carlo approximation (denoted as EB-partial MC) to see how partial MC influence the prediction accuracy. For EB-partial MC, a sampling size of 1000 is used. As we can see, the proposed EB method is much more accurate than Chen's method. At the early stage, e.g., 40% of the failure time, the advantage of the EB method is not significant, due to large uncertainty of model parameters and future change-points. As the prediction time approaches to the true failure time, the RMSD of the proposed method

decreases significantly, and the performance is much better than Chen's method. This is highly desirable since it becomes more and more important to get an accurate prediction when the RUL approaches zero. Comparing EB with EB-partial MC we can see that the prediction accuracy is not influenced much by partial MC sampling strategy.

Table 3-4 shows the computational costs of the EB method and EB-partial MC using MATLAB running on an i5-4690 CPU 3.50 GHz Intel processor at the prediction stage. In the RUL prediction, the computational times are calculated under different prediction steps. For example, if the prediction step is L, the conditional survival function $P(R_t > l|y_{1:t})$ is evaluated for $l = 1, 2, \cdots, L$, with in total L calculations. As we can see, the cost of EB method exponentially increases with the prediction step, due to the CDF computation of multivariate t distributions with increasing dimensions. For the EB-partial MC method, the computational cost of the prediction linearly increases with the prediction step. Therefore, using the partial MC strategy for the EB method could significantly reduce the computational cost, yet without influencing the prediction accuracy much.

Table 3-4: The computational cost of RUL prediction with and without partial Monte Carlo simulation (unit: seconds)

Method	Prediction Steps					
	10	20	30	40	50	60
EB	0.8	4.0	8.8	14.6	21.8	30.4
EB-partial MC	4.7	5.2	5.7	6.3	6.8	7.4

3.5.2 Application to Rotational Bearings

In this section, we apply the proposed EB method to the real degradation signals of rolling thrust bearings [12, 63, 94]. To generate these signals, a set of identical thrust bearings

was run at a constant rotational speed (2000 r/min) and a load of 200 lbs in an oil bath to provide continuous lubrication. Then the vibrations frequencies were acquired from an accelerometer, which was attached to the setup and connected to a vibration meter that measured the rms vibration level [28]. The amplitude of these frequencies increases as the bearing degrades. The degradation signal used in this paper consists of the average amplitude of the defective frequency and its first six harmonics frequencies. The degradation amplitudes are log-transformed. As the bearing degrades, the vibration becomes more and more severe and thus the degradation signal tends to increase. When the vibration magnitude reaches a threshold, the bearing is considered to have failed. Based on the published industrial standards, the failure threshold is set $\Gamma = \log(0.03)$ [12]. In total there are 25 historical signals. The sampling interval for all signals is 2 minutes.

Table 3-5: Estimated hyperparameters of the prior distributions

	$s = 1$	$s = 2$	$s = 3$
$\delta^{(s)}$	$\delta_0^{(1)} = 246,\ \sigma_0^{2(1)} = 163^2$	$\delta_0^{(2)} = 199,\ \sigma_0^{2(2)} = 123^2$	$\delta_0^{(3)} = 232,\ \sigma_0^{2(3)} = 123^2$
$\boldsymbol{\beta}^{(s)}$	$\boldsymbol{\mu}_0^{(1)} = [-7.28, 5.6 \times 10^{-6}]$ $\boldsymbol{\Sigma}_0^{(1)} = \begin{bmatrix} 20.77 & -0.072 \\ -0.072 & 0.0020 \end{bmatrix}$	$\boldsymbol{\mu}_0^{(2)} = [-6.97, 0.0015]$ $\boldsymbol{\Sigma}_0^{(2)} = \begin{bmatrix} 54.79 & -0.075 \\ -0.075 & 0.076 \end{bmatrix}$	$\boldsymbol{\mu}_0^{(3)} = [-4.76, 0.004]$ $\boldsymbol{\Sigma}_0^{(3)} = \begin{bmatrix} 3.74 & -0.0004 \\ -0.0004 & 0.0002 \end{bmatrix}$
$\sigma^{2(s)}$	$\alpha_1^{(1)} = 5.17, \alpha_2^{(1)} = 0.03$	$\alpha_1^{(2)} = 1.06, \alpha_2^{(2)} = 0.008$	$\alpha_1^{(3)} = 8.7, \alpha_2^{(3)} = 0.38$
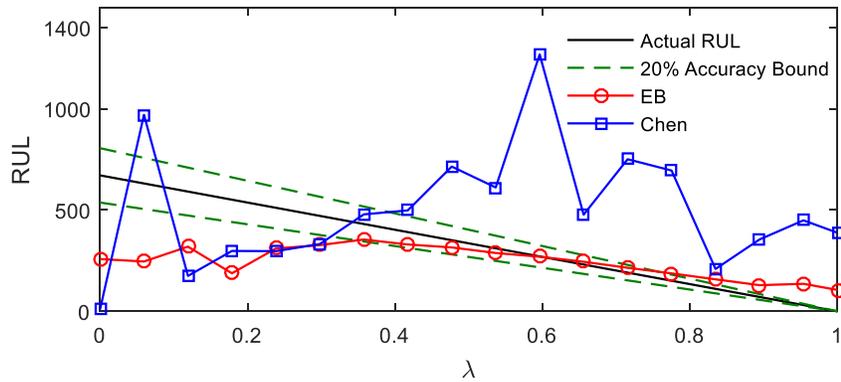
Figure 3-8: $\alpha - \lambda$ performance metric for the 24th bearing signal

In the offline modeling, we set the maximum number of change-points to be 2 to control the model complexity. We assume that all segments are line segments. Based on the BIC model selection, two-CP model is the best for all signals. The estimated hyperparameters are summarized in Table 3-5. The estimated means of three slopes show that degradation rate is almost zero at the first stage, indicating a stable operation, and then increases successively at the following two stages.

Figure 3-8 shows the α-λ performance metric for the 24th bearing signal. It can be observed that, although some estimated values are outside of the accuracy bound at early stage, the proposed method makes quite accurate prediction at later prediction stage. Apparently, the proposed method has a better performance. Figure 3-9 shows the prediction intervals at 50%, 70% and 90% of failure time for the 25 degradation signals. The prediction results of Chen's method is also provided for comparison. Some intervals by Chen's method are not shown since they are out of the y-axis range, e.g., the 9th and 14th signal on Figure 3-9 (a), the 20th signal on Figure 3-9 (b) and the 24th signal on Figure 3-9 (c). Figure 3-10 shows the RMSD of the 25 bearing signals for both EB and Chen's method. Clearly, as the prediction time is closer to the failure time, the intervals become narrower. The more observed data, the more accurate the

prediction. The prediction of the EB method is very stable and accurate across all units, while Chen's method is not stable and performs badly for some units. Table 3-6 shows the comparison of the proposed EB method with Chen's method, and the GLLR method [7], the GBPNN method [29] and EB method using three change-points (EB-CP3) in terms of the overall RMSD at the three time steps. In the GLLR method, the remaining data was fitted using Bayesian simple linear regression. In the GBPNN method, a back-propagation neural network-based model was developed for RUL prediction. It is noted that, in the GLLR and GBPNN methods, the first segment with normal working condition is manually truncated. Therefore, we just show the prediction at 70% and 90% prediction time for GLLR and GBPNN methods. Clearly, the proposed method outperforms all of other methods at all three prediction times. It should be mentioned that EB-CP3 is even worse. The reason is that most of bearings just have 1 or 2 obvious change-points, adding excessive change points will also introduce unnecessary uncertainty in RUL prediction, i.e., uncertainty of future change-point locations. Therefore, the number of change-points is critical for the prognostic model to generate an accurate prediction.

Table 3-6: RMSE of the proposed method in comparison with other methods

Method	RMSD		
	50%	70%	90%
GLLR	-	234.2	227.8
GBPNN	-	193.2	174.1
Chen	318.4	156.9	169.4
EB	236.8	106.0	41.8
EB-CP3	493.9	297.1	179.2
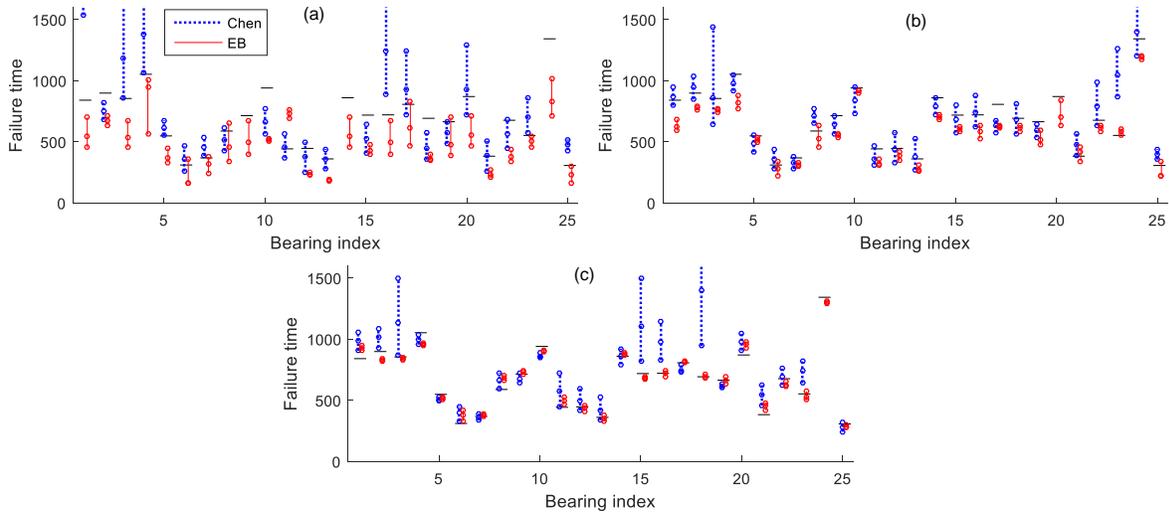
Figure 3-9: Prediction intervals for 25 bearing signals at three different prediction times. (a): 50%; (b): 70%; and (c): 90% of the failure time. The ∘ represents the 5%, 50%, 95% quantiles of the predicted RUL distributions, and – denotes the actual failure time.
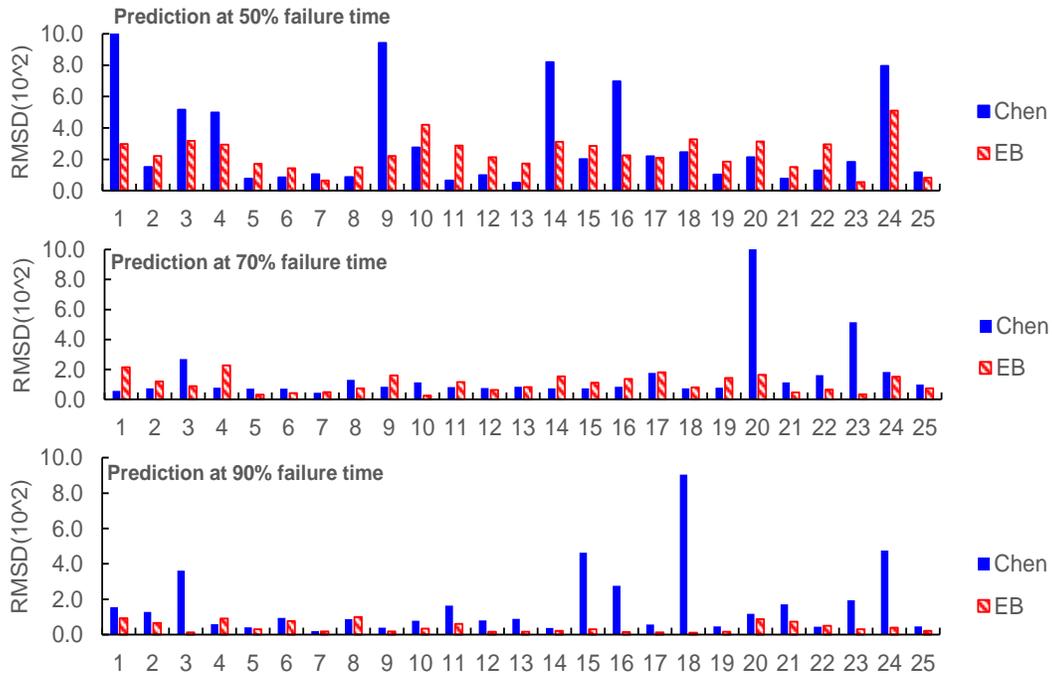
Figure 3-10: RMSD of 25 bearing signals

3.6 CONCLUSION

In this chapter we proposed a Bayesian multiple change-point modeling framework for degradation signals based condition monitoring and remaining useful life prediction. To capture the unit-to-unit heterogeneity and also to facilitate integration of historical data with in-situ observations of in-service unit for online prognostics, all model parameters are assumed to be random, including the number of change-points and their positions, and the model parameters of each linear segments. A novel stochastic process was proposed to model the joint prior of change-points and positions. A two-stage process was proposed to estimate all hyperparameters of priors. To facilitate online Bayesian model updating, a recursive updating algorithm was developed by which the posterior distribution of all state parameters can be exactly calculated. A closed-form of the RUL prediction is also derived. To control the computational cost in both model updating and RUL prediction process, a fixed-support-size strategy and a partial Monte Carlo strategy were proposed respectively, which significantly reduced the computational cost without influencing the prediction accuracy. The advantages of the proposed method have been demonstrated through thorough simulation studies and real case studies.

There are still open questions worthy of investigation. First of all, the current multiple change-point model assumes that all segments are independent. However, the degradation signals are often continuous in practice, indicating that all segments are connected and dependent. Incorporating such dependence could make the prior more informative and thus improve the prognostic accuracy. Secondly, adding more change-points may improve the model fitting and improve the prediction accuracy at the late degradation stage. However, it may reduce the prediction accuracy at the early degradation stage due to extra uncertainty by the added change-points. The strategy of using different models at different prediction stage may be

beneficial. Lastly, the segments of the multiple change-point model are modeled by parametric regression. It would be of interest to integrate the multiple change-point model to stochastic processes, e.g., Wiener process, for degradation modeling[*].

Chapter 4: Degradation Modeling and RUL Prediction using Wiener Processes subject to Multiple Change Points and Unit Heterogeneity

4.1 INTRODUCTION

In chapter 2 and chapter 3, both models are based on general path model. However, the inherent degradation path is deterministic once the regression parameters are known in the general path model [97]. It is often oversimplified and is not capable of capturing the temporal uncertainties that are inherent in the degradation process [98]. Therefore, the general path models are applicable only when the temporal uncertainties caused by unobserved internal or external factors are sufficiently small. The stochastic process models, on the other hand, are particularly effective in dealing with such unexplained randomness. The most popular stochastic process models include Gamma process [98], inverse Gaussian process [50], and Wiener process [99]. Due to nice mathematical properties and physical interpretations, Wiener processes have attracted widespread attention. Comprehensive reviews of Wiener process as degradation models can be found in Ye et al. [11]. this chapter aims to develop a multiple change-point Wiener process for degradation modeling and RUL prediction.

4.2 MULTIPLE CHANGE-POINT WIENER PROCESS MODELING

The general Wiener process can be represented as [43]

$$X(t) = \beta\Lambda(t) + \sigma B(\Lambda(t)) \tag{4.1}$$

where β is the drift parameter reflecting the rate of degradation, $\sigma > 0$ is called the diffusion parameter or diffusion coefficient, $B(\cdot)$ is the standard Brownian motion that captures the stochastic dynamics of the degradation process, and $\Lambda(\cdot)$ is a monotone increasing function representing a general time scale. When $\Lambda(t) = t$, this formula is simplified to the conventional

linear Wiener process. Let $\Delta X(t) = X(t + \Delta t) - X(t)$ denote the degradation increments from time t to $t + \Delta t$. According to the property of Wiener process, the increments are independent and normally distributed as $\Delta X(t) \sim N(\beta[\Lambda(t + \Delta t) - \Lambda(t)], \sigma^2[\Lambda(t + \Delta t) - \Lambda(t)])$.

In this chapter, a multiple change-point Wiener process degradation model is proposed to characterize the degradation path of CM signals with multiple phases. Specifically, the change points segment the signal into several consecutive phases, where each phase is modelled as a Wiener process. For simplicity, we use simple linear model in each phase for the mean degradation path, i.e., $\Lambda(t) = t$. Suppose the system is inspected at times $t_0, t_1, \ldots, t_n$ with degradation observations $X_0 = X(t_0), X_1 = X(t_1), \ldots, X_n = X(t_n)$, and assume there are K change points, with index locations $c_1, c_2, \cdots, c_K$. For notational convenience, we define $c_0 = 0$ and $c_{K+1} = n$. Then $c_0 = 0 < c_1 < c_2 < \cdots < c_K < c_{K+1} = n$. Therefore, the observations are partitioned into $K + 1$ consecutive phases. Mathematically, the multiple change-point Wiener process can be expressed piecewisely as

$$
X(t_j) = \begin{cases}
\beta^{(1)}(t_j - t_0) + X(t_0) + \sigma^{(1)}\mathcal{B}(t_j - t_0), \text{if } t_0 \leq t_j \leq t_{c_1} \\
\beta^{(2)}(t_j - t_{c_1}) + X(t_{c_1}) + \sigma^{(2)}\mathcal{B}(t_j - t_{c_1}), \text{if } t_{c_1} < t_j \leq t_{c_2} \\
\cdots \\
\beta^{(K+1)}(t_j - t_{c_K}) + X(t_{c_K}) + \sigma^{(K+1)}\mathcal{B}(t_j - t_{c_K}), \text{if } t_{c_K} < t_j \leq t_n
\end{cases} \quad (4.2)
$$

where $\beta^{(k)}$ and $\sigma^{(k)}, k = 1, \ldots, K + 1$ are the drift parameter and diffusion parameter respectively for the kth phase. It is worth noting that in the above model, the mean degradation path is continuous at all change points, which is conventional in the existing literature. It can be easily extended to a general case by adding an extra intercept term for each phase if necessary. Besides, for the sake of simplicity, the starting time of each Wiener process or each phase is assumed to be exactly on the discrete inspection epochs.

To account for the inherent unit-to-unit heterogeneity, the model parameters, including change-point locations, drift rate and diffusion parameter of each phase, are assumed to be random. For the sake of simplicity, the number of change points K is assumed to be deterministic for all units, which is often sufficient for almost all multi-phase degradation signals in the existing literature. If necessary, however, it can be easily extended to a more general case with a random K, as is the case in Wen *et al.* [100].

Bayesian approach is a natural choice to integrate the current available data with historical data for RUL prediction. Under Bayesian framework, the prognostics involves two stages, namely, the offline stage for prior specification and estimation using historical data, and the online stage for sequential model updating and RUL prediction when new observations are available. There are several remaining challenging issues to address under the multiple change-point framework. First of all, the number of change points K needs to be selected appropriately, which plays a decisive role on the modeling and prediction accuracy. Secondly, the prior distributions for the random model parameters need to be specified, and the corresponding hyperparameters need to be estimated through the historical data. Thirdly, at the online monitoring stage, the posterior distributions of the phase index, the location of the latest change point occurred, and the Wiener process parameters (drift rate and diffusion parameter) have to be sequentially updated once a new observation is available, which is often very challenging. Denote the parameters that need to be updated at the current time index m as a state vector $\boldsymbol{\theta}_m = (\beta_m, \sigma_m^2, \tau_m, s_m)$, where β_m and σ_m^2 are the drift and diffusion parameters of the current phase, τ_m is the latest change point that has occurred ($\tau_m \leq m - 1$), and s_m is the index of the current phase, e.g., $s_m = 1, 2, \ldots, K + 1$. Mathematically, the online model updating is to compute the posterior $p(\boldsymbol{\theta}_m | X_{0:m})$, which is highly nonlinear and thus generally intractable.

Lastly, based on the updated posterior distributions, we need to predict the residual life. Due to the potential occurrence of future change points, the RUL prediction is very complex. The overall prognostic framework is illustrated in Figure 4-1. In the following sections, the technical details regarding the aforementioned challenges will be provided.

(a) Offline Model Fitting

Specify the number of CPs K for historical dataset

⇩

Model estimation for each unit through MLE

⇩

Hyper-parameter estimation

⇩

Select optimal number of CPs through cross validation

(b) Online Model Updating and RUL Prediction

Collect new observations from a in-service unit

⇩

Model updating: $p(\boldsymbol{\theta}_m | X_{0:m})$

⇩

Predict the RUL
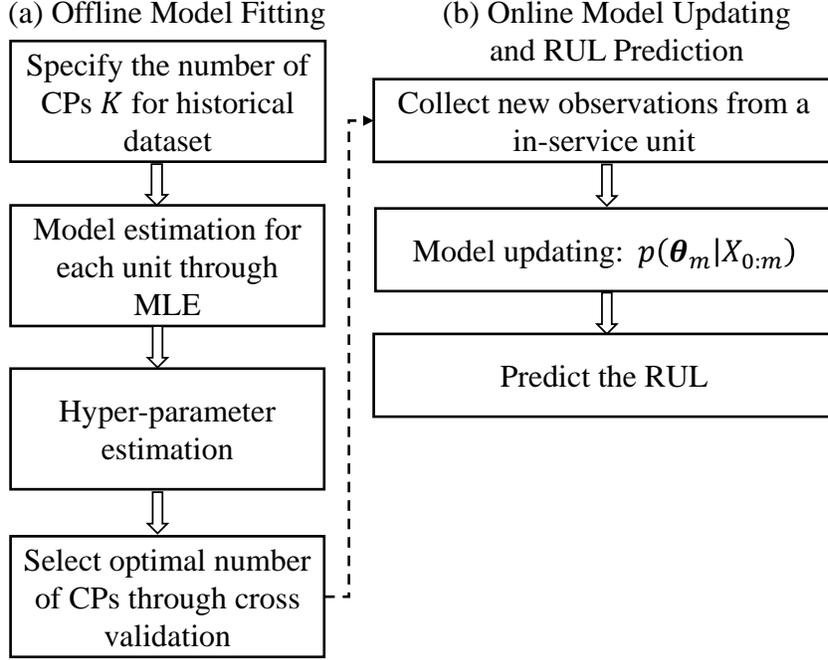
Figure 4-1: Illustration of the proposed prognostic framework.

4.3 OFFLINE PRIOR SPECIFICATION AND MODEL ESTIMATION

Denote the multiple change-point model as $\boldsymbol{\mathcal{M}} = \left(\{d^{(s)}\}_{s=1}^{K}, \{\beta^{(s)}\}_{s=1}^{K+1}, \{\sigma^{2(s)}\}_{s=1}^{K+1} \right)$ where K is the number of change points, $d^{(s)}$ is the duration of the sth phase, i.e., $d^{(s)} = t_{c_s} - t_{c_{s-1}}$. In the Bayesian scheme, appropriate priors for $\boldsymbol{\mathcal{M}}$ need to be specified. Prior information, which describes the population-based degradation characteristics, plays a critical role in the posterior inference of a new unit, especially when there are not sufficient observations at the early stage. Instead of directly specifying priors for the change-point locations, we focus on the

duration of each phase $d^{(s)}, s = 1, \dots, K$. Note that although there are $K + 1$ phases, only the first K phase durations are needed to identify the change-point locations. For simplicity, we assume that the Wiener process parameters are independent of phase durations, and all phases are independent, except that all phases are piecewise connected. Therefore, the joint prior for phase durations and Wiener process parameters can be formulated as

$$\pi(\mathcal{M}) = \prod_{s=1}^{K} \pi(d^{(s)}) \prod_{s=1}^{K+1} \pi(\beta^{(s)}, \sigma^{2(s)}) \tag{4.3}$$

Specifically, we assume that $d^{(s)}$ follows a normal distribution, i.e., $d^{(s)} \sim N(\mu_d^{(s)}, \sigma_d^{2(s)})$. For $\beta^{(s)}, \sigma^{2(s)}$, a commonly used normal and inverse Gamma (IG) conjugate priors are specified,

$$\pi(\beta^{(s)}, \sigma^{2(s)}) = \pi(\sigma^{2(s)})\pi(\beta^{(s)}|\sigma^{2(s)}) = IG\left(v_0^{(s)}, \gamma_0^{(s)}\right)N\left(\mu_0^{(s)}, \sigma^{2(s)}\kappa_0^{2(s)}\right) \tag{4.4}$$

Luckily, with the above conjugate priors, the joint posterior distribution of $\boldsymbol{\theta}_m$ can be exactly calculated sequentially through a recursive modeling updating method, which will be shown later. Let $\boldsymbol{\psi}$ be the vector of all unknown hyperparameters, i.e., $\boldsymbol{\psi} = $ Error! Bookmark not defined.. In the offline model fitting, all the hyperparameters have to be estimated. Suppose there are I units in the historical dataset. Naturally, the hyperparameters can be obtained from historical dataset by maximizing the following marginal likelihood [63]

$$\widehat{\boldsymbol{\psi}} = \arg\max_{\boldsymbol{\psi}} \prod_{i=1}^{I} \int p(\boldsymbol{X}_i|\mathcal{M}_i)\pi(\mathcal{M}_i|\boldsymbol{\psi})d\mathcal{M}_i \tag{4.5}$$

where $\boldsymbol{X}_i = \{X_{i,1}, \dots, X_{i,n_i}\}$ is the degradation signal for unit i, and n_i is the total number of observations. Unfortunately, the formula is too complex and very difficult to maximize directly.

In this paper, we adopt a commonly used empirical two-stage estimation method [59, 63], where the model parameter $\widehat{\mathcal{M}}_i$ for each unit i is estimated at the first stage, and then the hyperparameters are estimated through the maximum likelihood estimation (MLE) by treating $\{\widehat{\mathcal{M}}_i, i = 1, \dots, I\}$ as observations at the second stage.

To take advantage of the independent increment property of $X(t)$, define the increments of observations for unit i as $\delta_{i,1} = X_{i,1} - X_{i,0}, \delta_{i,2} = X_{i,2} - X_{i,1},\dots,\delta_{i,n_i} = X_{i,n_i} - X_{i,n_i-1}$, and time increments as $\lambda_{i,1} = t_{i,1} - t_{i,0}, \lambda_{i,2} = t_{i,2} - t_{i,1}, \dots, \lambda_{i,n_i} = t_{i,n_i} - t_{i,n_i-1}$. Then conditioning on $\mathcal{M}_i$, the increments $\boldsymbol{\delta}_i = \left(\delta_{i,1}, \delta_{i,2}, \dots, \delta_{i,n_i}\right)'$ follow independent normal distributions given by

$$f(\boldsymbol{\delta}_i|\mathcal{M}_i) = \prod_{s=1}^{K+1} \prod_{j=1}^{n_i^{(s)}} \left(2\pi\sigma_i^{2(s)}\lambda_{i,c_{s-1}+j}\right)^{-\frac{1}{2}} \exp\left(-\frac{\left(\delta_{i,c_{s-1}+j} - \beta_i^{(s)}\lambda_{i,c_{s-1}+j}\right)^2}{2\sigma_i^{2(s)}\lambda_{i,c_{s-1}+j}}\right) \quad (4.6)$$

where $n_i^{(s)} = c_{i,s} - c_{i,s-1}$ is the number of observations in the sth phase for unit i. The log-likelihood function can thus be expressed by

$$l(\mathcal{M}_i|\boldsymbol{\delta}_i) = \sum_{s=1}^{K+1} \left[-\frac{n_i^{(s)}}{2}\log(2\pi\sigma_i^{2(s)}) - \frac{1}{2}\sum_{j=1}^{n_i^{(s)}} \log\lambda_{i,c_{s-1}+j} - \frac{1}{2\sigma_i^{2(s)}}\sum_{j=1}^{n_i^{(s)}} \frac{\left(\delta_{i,c_{s-1}+j} - \beta_i^{(s)}\lambda_{i,c_{s-1}+j}\right)^2}{\lambda_{i,c_{s-1}+j}}\right] (4.7)$$

Given the change-point locations, the optimal drift and diffusion parameters that maximize Eq. (4.7) can be obtained as

$$\widehat{\beta}_i^{(s)} = \frac{\sum_{j=1}^{n_i^{(s)}} \delta_{i,c_{s-1}+j}}{\sum_{j=1}^{n_i^{(s)}} \lambda_{i,c_{s-1}+j}}, \qquad \widehat{\sigma^2}_i^{(s)} = \frac{1}{n_i^{(s)}} \sum_{j=1}^{n_i^{(s)}} \frac{\left(\delta_{i,c_{s-1}+j} - \widehat{\beta}_i^{(s)}\lambda_{i,c_{s-1}+j}\right)^2}{\lambda_{i,c_{s-1}+j}} \quad (4.8)$$

Plug in Eq. (4.8) into (4.7) we can get a likelihood function with $\{c_{i,s}, s = 1, \ldots, K\}$ being the only input variables. Denote $\boldsymbol{c}_i = \{c_{i,1}, c_{i,2} \ldots, c_{i,K}\}$, $\boldsymbol{\beta}_i = \{\beta_i^{(1)}, \ldots, \beta_i^{(K+1)}\}$ and $\boldsymbol{\sigma}_i^2 = \{\sigma_i^{2(1)}, \ldots, \sigma_i^{2(K+1)}\}$. The optimal change-point locations can be easily obtained by enumerating all possible values

$$\hat{\boldsymbol{c}}_i = \arg\max_{\boldsymbol{c}_i} l\left(\boldsymbol{c}_i, \widehat{\boldsymbol{\beta}}_i(\boldsymbol{c}_i), \widehat{\boldsymbol{\sigma}^2}_i(\boldsymbol{c}_i) \middle| \boldsymbol{\delta}_i\right) \tag{4.9}$$

At the second stage, the hyperparameters are estimated through MLE by treating the estimated $\{\widehat{\mathcal{M}}_i, i = 1, \ldots, I\}$ at the first stage as observations. The duration hyperparameters $(\mu_d^{(s)}, \sigma_d^{2(s)}, s = 1, \ldots, K)$ and drift rate hyperparameters $\left(\mu_0^{(s)}, \kappa_0^{2(s)}, s = 1, \ldots, K + 1\right)$ can be obtained analytically as

$$\hat{\mu}_d^{(s)} = \frac{\sum_{i=1}^{I} \hat{d}_i^{(s)}}{I}, \qquad \widehat{\sigma_d^2}^{(s)} = \frac{\sum_{i=1}^{I} \left(\hat{d}_i^{(s)} - \hat{\mu}_d^{(s)}\right)^2}{I} \tag{4.10}$$

And

$$\hat{\mu}_0^{(s)} = \frac{\sum_{i=1}^{I} \frac{\hat{\beta}_i^{(s)}}{\widehat{\sigma^2}_i^{(s)}}}{\sum_{i=1}^{I} \frac{1}{\widehat{\sigma^2}_i^{(s)}}}, \qquad \widehat{\kappa_0^2}^{(s)} = \frac{1}{I} \sum_{i=1}^{I} \frac{\left(\hat{\beta}_i^{(s)} - \hat{\mu}_0^{(s)}\right)^2}{\widehat{\sigma^2}_i^{(s)}} \tag{4.11}$$

For the hyperparameters v_0, γ_0 in the inverse Gamma distribution, the closed form is not tractable and instead can be estimated through nonlinear optimization techniques.

In the above model specification and estimation, the critical parameter K needs to be selected first. AIC [101] or BIC [84] is typically used for model selection in regression. However, the conventional AIC or BIC is not very effective for multiple change-point models

[77]. Besides, although increasing the number of change points may improve the model fitting accuracy, it may not necessarily result in better prognostic accuracy. In fact, increasing the number of change points will introduce extra uncertainties in RUL prediction (uncertainty of future change-point locations). Even if there is no over-fitting issue, it may still significantly reduce the prediction accuracy. Therefore there is tradeoff between model fitting accuracy and RUL predictability. To address this issue, we propose to use the cross validation technique for change-point model selection. Specifically, we apply leave-one-out-cross-validation approach. For each value K, the offline model estimation and online RUL prediction are performed, and the average prediction error is calculated. Then the optimal K is the one with minimal average prediction error. Note that for linear regression models without any change points, the cross validation approach is asymptotically equivalent to BIC based model selection [50]. However, for change-point models or other general models, cross-validation would be better since it is directly based on the model predictability.

4.4 ONLINE MODEL UPDATING AND RUL PREDICTION

Once the prior information is calculated based on historical data at the offline stage, as described in the previous section, it can be utilized for the RUL prediction of a new in-service unit at the online stage. To do this accurately requires a sequential updating the posterior distributions of certain key parameters, such as the location of the latest change point, the number of change-point occurred, and the drift and diffusion parameter of the current phase, which is the main challenge. In this section, we will first show the details of how to update the model recursively, and then present the RUL prediction method based on the updated model.

4.4.1 Exact Bayesian Model Updating

Assume that we have observed the degradation signal up to the current time step m for an in-service unit, denoted as $X_{0:m} = (X_0, X_1, \dots, X_m)$. The objective of Bayesian model updating is to incorporate the new observations to the estimated model by computing the posterior distribution of model parameters. The target distribution that needs to be updated is $p(\boldsymbol{\theta}_m | X_{0:m})$ where state vector $\boldsymbol{\theta}_m = (\beta_m, \sigma_m^2, \tau_m, s_m)$. In general, the analytic expression for this joint posterior distribution is intractable. A natural way is to use sequential Monte Carlo method to get an approximation. Wen *et al*. [100] developed a stratified particle filtering algorithm for online model updating of a general path model. This method can effectively handle intractable posteriors. However, to guarantee the approximation accuracy, this method requires a sufficient number of samples and has a relatively high computational cost. Fortunately, due to the assignment of conjugate priors for β and σ^2, the posterior could be exactly calculated through a novel recursive updating approach. Given the observed data, the joint posterior distribution of all parameters can be derived as

$$p(\boldsymbol{\theta}_m | X_{0:m}) = p(\beta_m, \sigma_m^2, \tau_m, s_m | X_{0:m}) = P(\tau_m, s_m | X_{0:m}) p(\beta_m, \sigma_m^2 | \tau_m, s_m, X_{0:m}) \qquad (4.12)$$

As we can see, the joint distribution can be factorized as the product of a posterior of the *discrete* components (τ_m and s_m), and the *continuous* components (β_m and σ_m^2). The discrete components are essential for phase tracking and future change-point prediction, while the continuous components are required to predict the degradation level at the end of the current phase. The details of how to calculate these two parts are given in the following paragraphs.

The conditional posterior distribution of the continuous components $p(\beta_m, \sigma_m^2 | \tau_m, s_m, X_{0:m})$ can be calculated based on Theorem 4.1 as follows.

Theorem 4.1. Given the conjugate priors shown in Eq. (4.4) for β and σ^2, the conditional posterior pdf $p(\beta_m, \sigma_m^2 | \tau_m = j, s_m = s, X_{0:m})$, can be calculated as

$$p(\beta_m, \sigma_m^2 | \tau_m = j, s_m = s, X_{0:m}) = p(\beta_m | \sigma_m^2, \tau_m = j, s_m = s, X_{0:m}) p(\sigma_m^2 | \tau_m = j, s_m = s, X_{0:m}) \quad (4.13)$$

Where

$$(\sigma_m^2 | \tau_m = j, s_m = s, X_{0:m}) \sim IG\left(v_0^{(s)} + \frac{m-j}{2}, \gamma_0^{(s)} + \frac{H_{j+1,m}^{(s)}}{2} \right) \quad (4.14)$$

$$(\beta_m | \sigma_m^2, \tau_m = j, s_m = s, X_{0:m}) \sim N(\mu_{j+1,m}^{(s)}, \sigma_m^2 \kappa_{j+1,m}^{2(s)})$$

and

$$H_{j+1,m}^{(s)} = \left[\frac{\mu_0^{2(s)}}{\kappa_0^{2(s)}} + \sum_{i=j+1}^{m} \frac{\delta_i^2}{\lambda_i} - \left(\frac{\mu_0^{(s)}}{\kappa_0^{2(s)}} + \sum_{i=j+1}^{m} \delta_i \right)^2 \left(\sum_{i=j+1}^{m} \lambda_i + \frac{1}{\kappa_0^{2(s)}} \right)^{-1} \right]$$

$$\kappa_{j+1,m}^{2(s)} = \left(\sum_{i=j+1}^{m} \lambda_i + \frac{1}{\kappa_0^{2(s)}} \right)^{-1} \quad (4.15)$$

$$\mu_{j+1,m}^{(s)} = \left(\frac{\mu_0^{(s)}}{\kappa_0^{2(s)}} + \sum_{i=j+1}^{m} \delta_i \right) \kappa_{j+1,m}^{2(s)}$$

The proof is given in Appendix E. For the discrete components, based on the Bayes' rule, the posterior can be derived as

$$P(\tau_m = j, s_m = s | X_{0:m}) \propto P(\tau_m = j, s_m = s | X_{0:m-1}) p(X_m | \tau_m = j, s_m = s, X_{0:m-1}) \quad (4.16)$$

In the above equation, $P(\tau_m = j, s_m = s | X_{0:m-1})$ is the predictive probability mass function (PMF), which can be recursively calculated by conditioning on the states of the previous time step:

$$P(\tau_m = j, s_m = s | X_{0:m-1})$$

$$= \sum_{j',s'} P(\tau_{m-1} = j', s_{m-1} = s' | X_{0:m-1}) P(\tau_m = j, s_m = s | \tau_{m-1} = j', s_{m-1} = s', X_{0:m-1}) \quad (4.17)$$

In Eq. (4.17), $P(\tau_{m-1} = j', s_{m-1} = s'|X_{0:m-1})$ is the posterior distribution of the discrete component obtained at the previous time step. $P(\tau_m = j, s_m = s|\tau_{m-1} = j', s_{m-1} = s', X_{0:m-1})$ is the prior state transition probability, which can be derived based on the Markov properties of the occurrence of the change points, i.e., the probability of the occurrence of a new change point or a new phase at the current time only depends on the duration and phase index at the previous time step. Based on the prior knowledge of the phase duration and total number of phases, the state transition probability can be obtained as

$P(\tau_m = j, s_m = s|\tau_{m-1} = j', s_{m-1} = s', X_{0:m-1})$

$$
= \begin{cases}
\dfrac{1 - G^{(s')}(t_m - t_{j'})}{1 - G^{(s')}(t_{m-1} - t_{j'})}, & \text{if } j = j' \text{ and } s = s' < K+1 \\
1, & \text{if } j = j' \text{ and } s = s' = K+1 \\
\dfrac{G^{(s')}(t_m - t_{j'}) - G^{(s')}(t_{m-1} - t_{j'})}{1 - G^{(s')}(t_{m-1} - t_{j'})}, & \text{if } j = m-1 \text{ and } s = s' + 1 \le K+1 \\
0, & \text{otherwise}
\end{cases}
\tag{4.18}
$$

where $G^{(s')}(\cdot)$ is the cumulative distribution function of the s'th phase duration. In Eq. (4.18), there are three nonzero probabilities corresponding to three scenarios. The first one refers to the probability of no occurrence of a new change-point or new phase (i.e., $\tau_m = \tau_{m-1}$, $s_m = s_{m-1}$) given that the degradation at the time step $m-1$ is not at the final phase (i.e., $s' < K+1$), and the duration of the s'-th phase is larger than $t_{m-1} - t_{j'}$ (i.e., $\tau_{m-1} = j'$). Therefore, it is equivalent to

$$
P\big(d^{(s')} \ge t_m - t_{j'}\big|d^{(s')} \ge t_{m-1} - t_{j'}\big) = \frac{1 - G^{(s')}(t_m - t_{j'})}{1 - G^{(s')}(t_{m-1} - t_{j'})}
$$

Similarly, we can easily get the other two probabilities.

The predictive density $p(X_m | \tau_m = j, s_m = s, X_{0:m-1})$ in Eq. (4.16) can be calculated based on Theorem 4.2 as follows.

Theorem 4.2. Suppose the conjugate priors shown in Eq. (4.4) are assumed for β and σ^2.

if $j < m - 1$, $(X_m | \tau_m = j, s_m = s, X_{0:m-1})$ follows a univariate t distribution given as

$$(X_m | \tau_m = j, s_m = s, X_{0:m-1}) \sim t_1 \left(2v_0^{(s)} + m - j - 1, \mu_{m|j+1:m-1}^{(s)}, \frac{2\gamma_{j+1,m-1}^{(s)} \eta_{m|j+1:m-1}^{(s)}}{2v_0^{(s)} + m - j - 1} \right) \quad (4.19)$$

where

$$\mu_{m|j+1:m-1}^{(s)} = \mu_{j+1,m-1}^{(s)} \lambda_m + X_{m-1},$$

$$\gamma_{j+1,m-1}^{(s)} = \gamma_0^{(s)} + \frac{H_{j+1,m-1}^{(s)}}{2} \quad (4.20)$$

$$\eta_{m|j+1:m-1}^{(s)} = \lambda_m + \kappa_{j+1,m-1}^{2(s)} \lambda_m^2$$

if $j = m - 1$, $(X_m | \tau_m = j, s_m = s, X_{0:m-1})$ follows a univariate t distribution given as

$$(X_m | \tau_m = j, s_m = s, X_{0:m-1}) \sim t_1 \left(2v_0^{(s)}, \mu_m^{(s)}, \frac{\gamma_0^{(s)} \eta_m^{(s)}}{v_0^{(s)}} \right) \quad (4.21)$$

where

$$\mu_m^{(s)} = \mu_0^{(s)} \lambda_m + X_{m-1}$$

$$\eta_m^{(s)} = \lambda_m + \kappa_0^{2(s)} \lambda_m^2 \quad (4.22)$$

The proof is included in Appendix F. As we can see, based on Eq. (4.12)-(4.22), the posterior distributions of all parameters of interest can be exactly calculated through a recursive updating approach.

4.4.2 Controlling the Computational Cost

Compared with sequential Monte Carlo methods, the exact inference for sequential model updating runs much faster when m is not large. However, as m becomes very large, the

algorithm may be very time-consuming. The reason is that it needs to calculate all the probabilities of $P(\tau_m = j, s_m = s | X_{0:m})$ for $j = 0, 1, \ldots, m-1$ and $s = 1, 2, \ldots, K+1$ at each time step m, which increases almost linearly with m. Therefore, the algorithm needs to be improved in a computationally efficient manner for real-time estimation. It is observed that as the number of observations increases, the PMF of the latest change point becomes more and more concentrated around the true change point, whereas the probabilities at other locations are close to zero. To reduce the computational load and balance the computational cost for all time steps, an approximation strategy with a set of support point of fixed size can be applied. The basic idea is to approximate the posterior distribution of the discrete components by a small set of support points of a fixed size with significant probabilities. In other words, we set the posterior PMF to be zero at those with negligible values and keep others with high probabilities. However, this strategy may result in zero PMFs for certain phases, i.e., $P(s_m = s | X_{0:m}) = 0$ or $P(\tau_m = j, s_m = s | X_{0:m}) = 0$ for all $j = 0, 1, \ldots, m-1$, whose true values may be just temporally small and will become significant as more observations are obtained. To avoid this situation, we propose to use a stratified sampling method, where for each s, N locations with highest PMF are selected for the latest change points. The algorithm is summarized in Table 4-1. With this strategy, the maximum computational cost for each time step m can be controlled effectively.

Table 4-1: The fixed support size strategy for model updating

For each time step m:

If $m \leq N - 1$:

- Calculate $P(\tau_m = j, s_m = s | X_{0:m})$ for $j = 0, 1, m - 1$, $s = 1, \ldots, K + 1$.

If $m > N$:

- For each s, select $N - 1$ support points with highest probabilities $P(\tau_{m-1} = j, s_{m-1} = s | X_{0:m-1})$

- Calculate $P(\tau_m = j, s_m = s | X_{0:m})$ at all these selected $(N - 1)$ support points and at the current time step m for $s = 1, \ldots, K + 1$, set others to be zero.

- Normalize the probabilities.

4.4.3 RUL prediction

Once the parameters in the model have been updated, the RUL of the in-service unit can be predicted. A failure is typically defined as the event that the degradation signal first hits the failure threshold Γ. Denote the RUL at the current time t_m as R_m. Based on the concept of first passage time (FPT), the RUL can be formulated as $R_m =$ infError! Bookmark not defined.. For the conventional Wiener process degradation model, RUL has been proven to follow an inverse Gaussian distribution. For an in-service unit with observations $X_{0:m}$, if β and σ^2 are fixed, the pdf of the residual life can be derived as [48]

$$f_{R_m}(l | X_m, \beta, \sigma^2) = \frac{\Gamma - X_m}{\sqrt{2\pi\sigma^2 l^3}} \exp\left(-\frac{(\Gamma - X_m - \beta l)^2}{2\sigma^2 l}\right) \tag{4.23}$$

However, in our model, due to the unknown change points and randomness of β, σ^2 at each phase, the RUL is very complicated and intractable analytically. For model consistency, we assume that the degradation amplitude will not exceed the failure threshold before the last change point. In RUL prediction, we need to first predict the location of the final change point

t_{fc} and the degradation amplitude $X(t_{\text{fc}})$, and then conditioning on them to predict when the last

phase will hit the failure threshold. Mathematically, the pdf of RUL can be represented as

$$f_{R_m}(l|X_{0:m}) = \iint p(t_{\text{fc}}, X(t_{\text{fc}})|X_{0:m}) f_{R_m}(l|t_{\text{fc}}, X(t_{\text{fc}}), X_{0:m}) d\, X(t_{\text{fc}}) dt_{\text{fc}} \qquad (4.24)$$

where

$$p(t_{\text{fc}}, X(t_{\text{fc}})|X_{0:m}) = \sum_{s_m, \tau_m} P(s_m, \tau_m|X_{0:m}) p(t_{\text{fc}}, X(t_{\text{fc}})|s_m, \tau_m, X_{0:m})$$

$$= \sum_{s_m, \tau_m} P(s_m, \tau_m|X_{0:m}) \int p(t_{\text{fc}}, X(t_{\text{fc}})|\boldsymbol{\theta}_m) p(\beta_m, \sigma_m^2|s_m, \tau_m, X_{0:m}) d\beta_m d\sigma_m^2$$

$$(4.25)$$

and

$$f_{R_m}(l|t_{\text{fc}}, X(t_{\text{fc}}), X_{0:m})$$

$$= \iint f_{R_m}(l|t_{\text{fc}}, X(t_{\text{fc}}), X_{0:m}, \beta^{(K+1)}, \sigma^{2(K+1)}) p(\beta^{(K+1)}, \sigma^{2(K+1)}|t_{\text{fc}}, X(t_{\text{fc}}), X_{0:m}) d\beta^{(K+1)} d\sigma^{2(K+1)}$$

$$(4.26)$$

It is worth noting that here t_{fc} may be less than t_m, therefore $f_{R_m}(l|t_{\text{fc}}, X(t_{\text{fc}}), X_{0:m})$ may

depend on $X_{0:m}$ or may be independent of $X_{0:m}$. Clearly, the RUL is intractable due to multiple

complex integrations. A natural way to address this issue is to use Monte Carlo simulation

approach. Specifically, we can first generate M samples for the current state vector $\boldsymbol{\theta}_m$ through

the updated posterior distribution $p(\boldsymbol{\theta}_m|X_{0:m})$, and then conditioning on each sample, simulate

the remaining change points, β and σ^2 of each phase, and degradation levels at remaining

change points. The duration of the last phase can be directly sampled from inverse Gaussian

distribution based on Eq. (4.23). The details of the Monte Carlo approach for RUL sampling

$(T_i, i = 1, \dots, M)$ is illustrated in Table 4-2. It should be mentioned that for multiple change-point

based model, the probability of being at the last phase at early prediction stage is inevitably

nonzero, i.e., $P(s_m = K + 1|X_{0:m}, \tau_m = j) > 0$ while the actual phase $s_m < K + 1$. This

probability may even be significant for certain signals due to inherent randomness of Wiener process. If the actual degradation rate is very small at the current time, the sampled failure time may be significantly larger than the actual value. Besides, for phases before the last phase, negative drift parameters may be sampled, which may also significantly increase the residual life. To make it more accurate, we apply lower bounds $l^{(s)}$ for all the drift parameters as constraints in the sampling process, which can be obtained by selecting the minimal drift parameter of each phase of all historical signals.

<div align="center">Table 4-2: Monte Carlo Simulation for RUL Prediction</div>

Generate samples from $p(\boldsymbol{\theta}_m|X_{0:m})$:

For $i = 1:M$

    Sample $(\tau_{m,i}, s_{m,i}) \sim P(\tau_m, s_m|X_{0:m})$ based on Eq. (4.16)

    Sample $(\beta_{m,i}, \sigma_{m,i}^2) \sim p(\beta_m, \sigma_m^2|\tau_{m,i}, s_{m,i}, X_{0:m})$ based on Eq. (4.14)

End

Simulate RUL

For $i = 1:M$

    If $s_{m,i} = K + 1$, then

    Sample $T_i \sim f_{R_m}\left(l|X_m, \beta_{m,i}, \sigma_{m,i}^2\right)$ based on Eq. (4.23)

    else

    For $s = s_{m,i}:K$

      If $s = s_{m,i}$

        Sample $t_{c_s,i} \sim t_{c_{s-1},i} + N\left(d\Big|\mu_d^{(s)}, \sigma_d^{2(s)}, d \geq t_m - t_{\tau_{m,i}}\right)$

        Sample $X(t_{c_s,i}) \sim X(t_m) + N\left(X|\beta_{m,i}(t_{c_s,i} - t_m), (t_{c_s,i} - t_m)\sigma_{m,i}^2\right)$

      else

        Sample $t_{c_s,i} \sim t_{c_{s-1},i} + N\left(d\Big|\mu_d^{(s)}, \sigma_d^{2(s)}\right)$

        Sample $(\beta_i^{(s)}, \sigma_i^{2(s)}) \sim IG\left(v_0^{(s)}, \gamma_0^{(s)}\right) N\left(\mu_0^{(s)}, \sigma_0^{2(s)}\kappa_0^{2(s)}\right)$

        Sample $X(t_{c_s,i}) \sim X(t_{c_{s-1},i}) + N\left(X\Big|\beta_i^{(s)}(t_{c_s,i} - t_{c_{s-1},i}), (t_{c_s,i} - t_{c_{s-1},i})\sigma_i^{2(s)}\right)$

      End

    End

    Sample $\left(\beta_i^{(K+1)}, \sigma_i^{2(K+1)}\right) \sim IG\left(v_0^{(K+1)}, \gamma_0^{(K+1)}\right) N\left(\mu_0^{(K+1)}, \sigma^{2(K+1)}\kappa_0^{2(K+1)}\right)$

    Sample $L_i \sim f_{R_{c_K}}\left(l\Big|X(t_{c_K,i}), \beta_i^{(K+1)}, \sigma_i^{2(K+1)}\right)$ based on Eq. (4.23)

    $T_i = (t_{c_K,i} - t_m) + L_i$

    End

End

4.5 CASE STUDIES

In this section, we first use simulation study to demonstrate the robustness and effectiveness of our model. Then, the proposed model is applied to real case study of rotational bearings.

4.5.1 Simulation Study

In this subsection, the prediction is illustrated and the performance is evaluated through numerical simulations. For simplicity, we only consider one-change-point and two-change-point scenarios in the simulation model. The hyperparameters for each scenario are specified in Table 4-3. The failure threshold is set to be $\Gamma = 20$. For each scenario, 200 degradation signals are generated as training dataset. Another 200 signals in each category are generated as testing dataset. The support size for the PMF approximation at the updating stage is specified as $N = 20$. The number of samples for RUL prediction is set as $M = 3000$. Due to the randomness of Monte Carlo method, the prediction procedure is repeated 10 times for each signal.

Due to space limitation, the estimated hyperparameters are not provided here. Using the leave-one-out-cross-validation approach, we find that the identified optimal change-point number for each dataset is equivalent to the true value, which demonstrates the effectiveness of the proposed approach for model selection (see Figure 4-5). Figure 4-2 shows the online model updating process for two degradation signals, each of which is randomly selected from each dataset. The first and the second column shows the results by assuming $K = 1$ and $K = 2$, respectively, while the true value $K_{\text{true}} = 1$. Similarly, the third and fourth column assume $K = 1$ and $K = 2$, respectively, while $K_{\text{true}} = 2$. Here we select two $K's$ for each signal to study the consequence if K is specified inappropriately. Clearly, if an appropriate K is specified (Figure 4-2 (a) and (d)), the recursive updating algorithm can accurately detect the occurrence of

change points and track the phase index. However, if K is larger (or less) than the actual value, more phases will occur (or some change points may not be detected), which will thus affect the model fitting and RUL prediction.

Table 4-3: Monte Carlo Simulation for RUL Prediction

Variables	$K_{\text{true}} = 1$	$K_{\text{true}} = 2$
$d^{(s)}$	$\mu_d^{(1)} = 300, \sigma_d^{2(1)} = 10^2$ $\mu_d^{(2)} = 400, \sigma_d^{2(2)} = 10^2$	$\mu_d^{(1)} = 200, \sigma_d^{2(1)} = 10^2$ $\mu_d^{(2)} = 300, \sigma_d^{2(2)} = 10^2$ $\mu_d^{(3)} = 500, \sigma_d^{2(3)} = 10^2$
$\beta^{(s)}$	$\mu_0^{(1)} = 0.01, \kappa_0^{2(1)} = 0.008$ $\mu_0^{(2)} = 0.06, \kappa_0^{2(2)} = 0.006$	$\mu_0^{(1)} = 1 \times 10^{-5}, \kappa_0^{2(1)} = 9 \times 10^{-4}$ $\mu_0^{(2)} = 0.02, \kappa_0^{2(2)} = 0.002$ $\mu_0^{(3)} = 0.09, \kappa_0^{2(3)} = 0.002$
$\sigma^{2(s)}$	$v_0^{(1)} = 2, \gamma_0^{(1)} = 0.05$ $v_0^{(2)} = 2, \gamma_0^{(2)} = 0.1$	$v_0^{(1)} = 2, \gamma_0^{(1)} = 0.06$ $v_0^{(2)} = 2, \gamma_0^{(2)} = 0.08$ $v_0^{(3)} = 2, \gamma_0^{(3)} = 0.1$
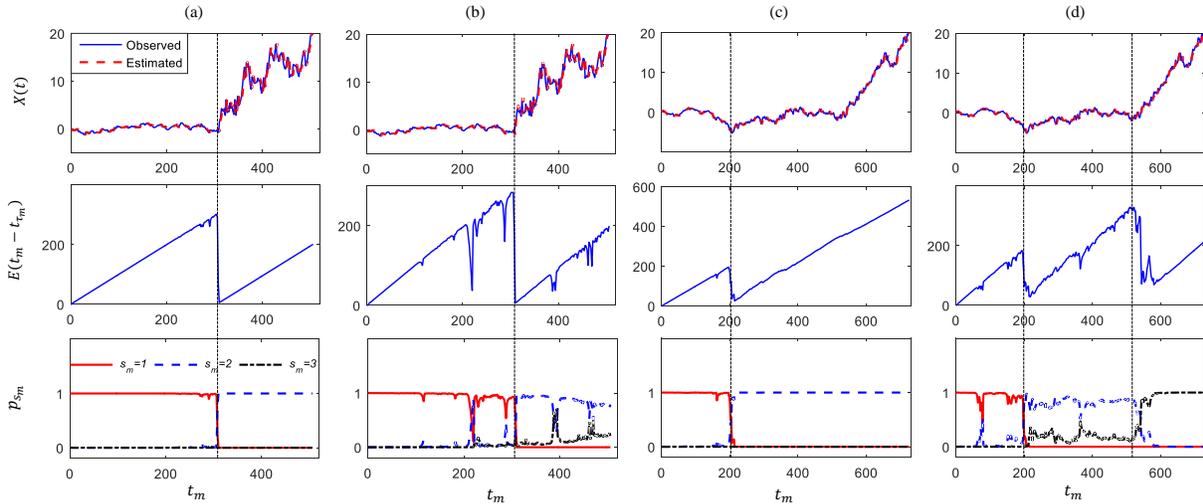
Figure 4-2: Illustration of the online model updating process. (a) $K_{\text{true}} = 1, K = 1$; (b) $K_{\text{true}} = 1, K = 2$ (c) $K_{\text{true}} = 2, K = 1$ and (d) $K_{\text{true}} = 2, K = 2$; top panel: observed and filtered signals; middle panel: the expected duration of the current phase; bottom panel: the posterior PMF of the index of current phase. The vertical dashed lines are true change-point locations.
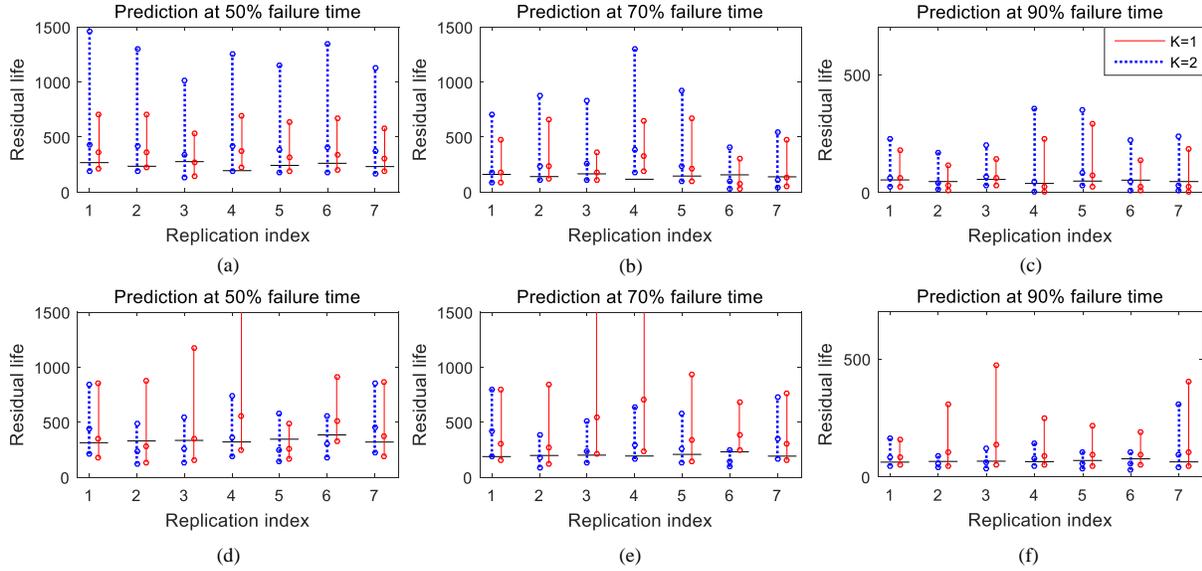
Figure 4-3: Prediction intervals of RUL for 14 simulated signals predicting at 50%, 70%, 90% of failure time. (a-c): $K_{true} = 1$; (d-f): $K_{true} = 2$. The "°" denotes the 5%, 50% and 95% quantile of RUL prediction; " − " represent the actual failure time.

Figure 4-3 shows the prediction intervals of 7 randomly selected signals from each dataset at three different prediction times, i.e., 50%, 70%, 90% of actual failure time. Figure 4-4 illustrates the pdf of RUL of the third and the second unit of the seven signals in each category. To compare the prediction performance of different models, we select $K = 1$ and $K = 2$ for both two types of signals. As expected, for both model specifications, the prediction becomes more and more accurate for all signals as more and more observations are collected. This characteristic is highly desirable since it becomes more important to get an accurate RUL prediction as the unit approaches failure. Comparing the two model selections, we can see that if the true K is selected, the prediction performance is much better than if K is selected inappropriately. Note that in the simulation, the right model was effectively selected through the cross-validation approach.
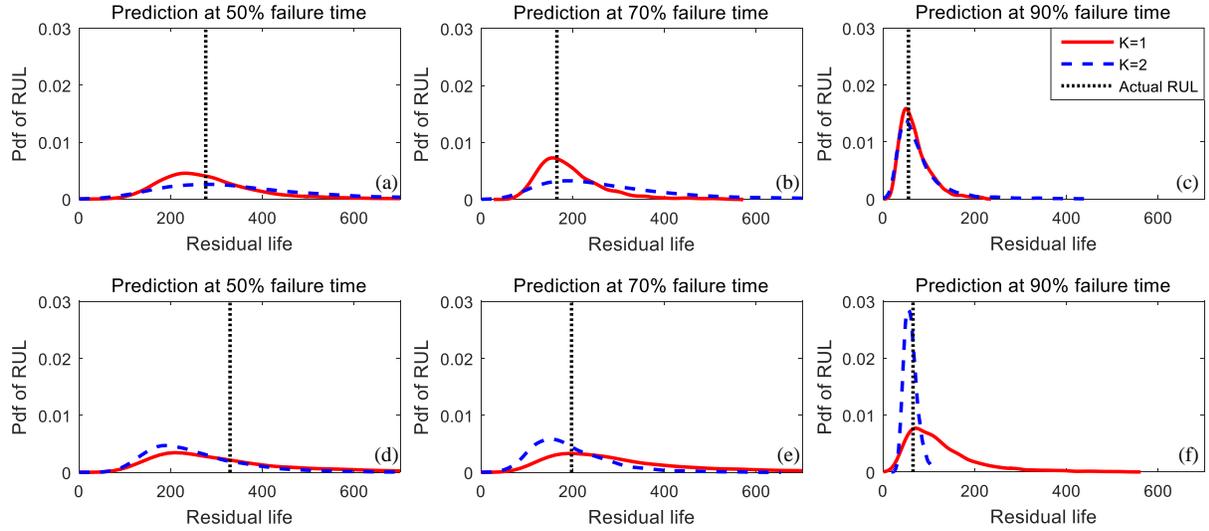
Figure 4-4: Comparison of the RUL prediction between $K = 1$ and $K = 2$. Top panel (a-c): $K_{true} = 1$; bottom panel (d-f): $K_{true} = 2$.

To further evaluate and analyze the overall prediction accuracy, 200 signals in each category are used for testing. We use the root-mean-square deviation (RMSD) as a performance metric for 200 testing signals in each category, which is defined as

$$\text{RMSD} = \sqrt{\frac{1}{I}\sum_{i=1}^{I} E(\hat{R}_i - R_{i,\text{true}})^2} \qquad (4.27)$$

where I is the total number of units, $\hat{R}_i$ and $R_{i,\text{true}}$ are the predicted and true RUL of unit i, respectively. Figure 4-5 shows the RMSD at six prediction times for both two types of signals. For comparison purpose, three models $K = 0, 1, 2$ are provided here. Clearly, the model with appropriate K outperforms all other models at all prediction times. From Figure 4-5(a) we can see that, the RMSD goes down gradually for both $K = 1$ and $K = 2$ models, indicating that with more observations collected, the prediction becomes more accurate. However, the model with $K = 1$ is more accurate in prediction than $K = 2$. The reason is that adding excessive

97

change points will introduce unnecessary uncertainty in RUL prediction, e.g., uncertainty regarding future change points, and Wiener process parameters of the last phase.

For $K = 0$, the RMSD first increases and then decreases. Similarly phenomenon can also be observed in Figure 4-5(b). If the number of change points is selected insufficiently, the RMSD first increases and then decreases as we increase the prediction times. The reason is that the early phases often have smaller degradation rates than the later phases, as shown in Figure 4-2. If insufficient change points are assumed, the early phases may be mistakenly detected as the last phase. The more observations are collected, the less influence the prior will have on the posterior updating and thus the lower the updated degradation rate will be. Consequently it will make the predicted RUL much higher than the actual value. When the unit approaches to its failure, the prediction accuracy for all of these models increases. Therefore, the parameter K is critical for the prognostic model to produce an accurate prediction.
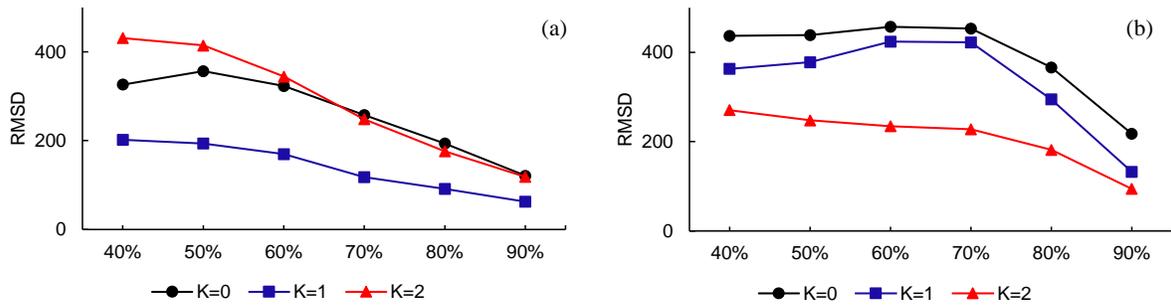
Figure 4-5: Comparison of the RMSD at six prediction times: (a) $K_{true} = 1$; (b) $K_{true} = 2$.

4.5.2 Application to Bearing Signals

In this subsection, the proposed approach is applied to the degradation signal of rotational bearings. These degradation signals are log-transformed vibrational signals obtained through an accelerated testing on a set of identical thrust ball bearings [12, 63]. In total, there are 25

degradation signals. The time interval for inspection is 2 minutes, i.e., $\Delta t = 2$ for all units.

Figure 4-6 shows three degradation paths of log-transformed vibrational signals for illustration. The failure threshold is specified as $\Gamma = \log(0.03)$, which is based on the published industrial standards [12]. There are obviously two phases for all signals. Moreover, the locations of change points vary from unit to unit.
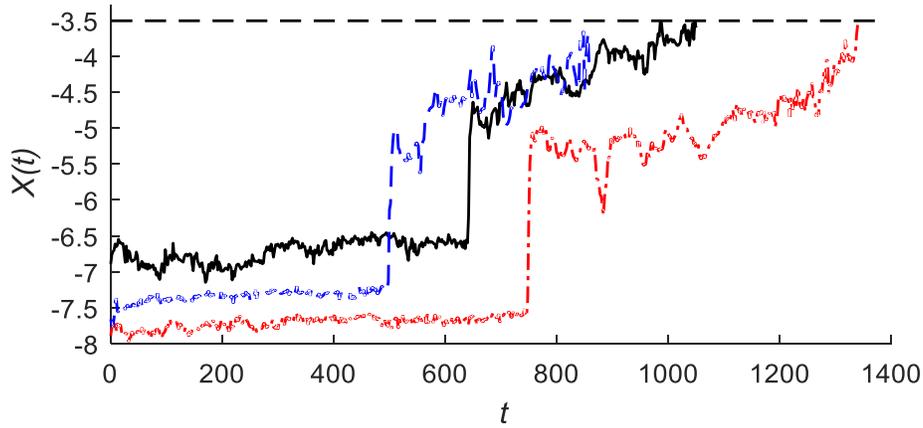
Figure 4-6: Degradation paths of three representative bearings.

Table 4-4: Estimated hyperparameters for three models

Variables	$K = 0$	$K = 1$	$K = 2$
$d^{(s)}$	$\mu_d^{(1)} = 676$ $\sigma_d^{2(1)} = 255^2$	$\mu_d^{(1)} = 360, \sigma_d^{2(1)} = 241^2$ $\mu_d^{(2)} = 317, \sigma_d^{2(2)} = 134^2$	$\mu_d^{(1)} = 310, \sigma_d^{2(1)} = 190^2$ $\mu_d^{(2)} = 62, \sigma_d^{2(2)} = 119^2$ $\mu_d^{(3)} = 306, \sigma_d^{2(3)} = 142^2$
$\beta^{(s)}$	$\mu_0^{(1)} = 0.0056$ $\kappa_0^{2(1)} = 8.23 \times 10^{-4}$	$\mu_0^{(1)} = 8.317 \times 10^{-4}$ $\kappa_0^{2(1)} = 0.0038$ $\mu_0^{(2)} = 0.0083$ $\kappa_0^{2(2)} = 0.0012$	$\mu_0^{(1)} = -3.12 \times 10^{-4}, \kappa_0^{2(1)} = 0.006$ $\mu_0^{(2)} = 0.0215, \kappa_0^{2(2)} = 0.4425$ $\mu_0^{(3)} = 0.0049, \kappa_0^{2(3)} = 0.0024$
$\sigma^{2(s)}$	$v_0^{(1)} = 10.07$ $\gamma_0^{(1)} = 0.067$	$v_0^{(1)} = 1.39, \gamma_0^{(1)} = 0.003$ $v_0^{(2)} = 2.29, \gamma_0^{(2)} = 0.017$	$v_0^{(1)} = 2.1, \gamma_0^{(1)} = 0.004$ $v_0^{(2)} = 0.42, \gamma_0^{(2)} = 0.004$ $v_0^{(3)} = 5.1, \gamma_0^{(3)} = 0.03$

For model selection and comparison, we mainly consider three cases, $K = 0, 1,$ and 2. Table 4-4 summarizes the estimated hyperparameters for each case. From Table 4-4 we can see that, if $K=1$ is specified, the second phase has a larger drift rate than the first one, which indicates a faster degradation. If $K = 2$ is specified, the degradation rate of the second phase is much larger than the third phase. Not surprisingly, almost all bearings have a sudden jump after the first phase. Although extremely short in duration, the sudden jump is so significant that it was identified as a single phase at the offline model fitting for all signals. Even for online model updating, as shown in Figure 4-7(b), the sudden jump is detected as a single phase.
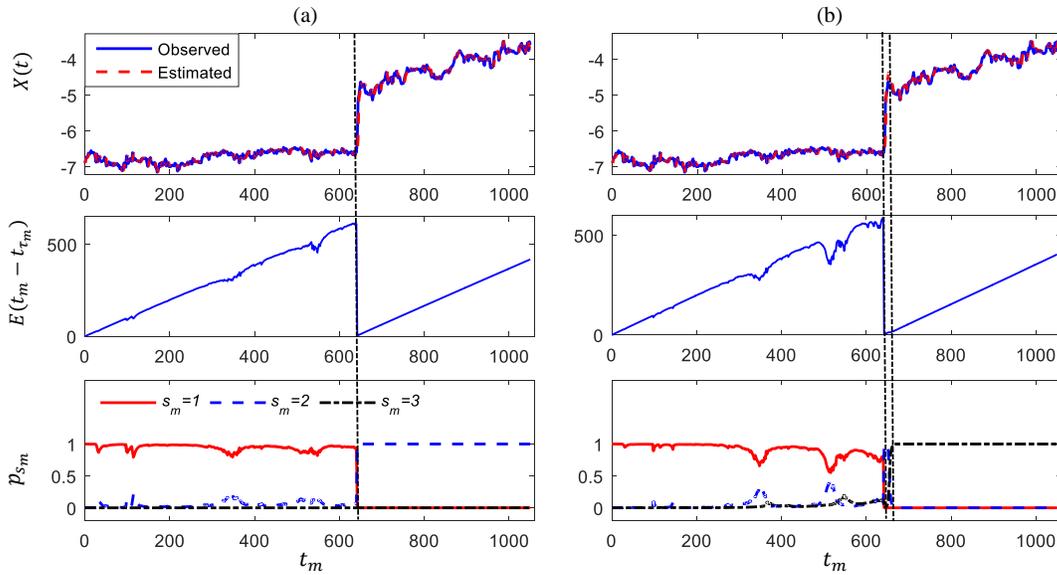
Figure 4-7: Illustration of online model updating process. (a): $K = 1$; (b):$K = 2$; top panel: observed and filtered signals; middle panel: the expected duration of the current phase; bottom panel: the posterior PMF of the index of current segment. The vertical dashed lines are change-point locations identified at the offline stage.

Figure 4-8 shows the RMSD of the 25 bearing signals at three prediction times for three models. Figure 4-9 shows the overall prediction accuracy at six prediction times. For comparison purpose, we add the case of $K = 3$ as well. Clearly, at all the six prediction times, the model

with only one change point performs best in terms of the overall RMSD. As observed in our study, specifying two change points only improves the prediction within the sudden-jump phase. It often reduce the prediction accuracy at the early stage and even at the late stage. Besides, assuming no change points is surprisingly better than using two and three change points. The reason is that for some signals, there is a sharp increase right before failure, as can be seen from the red curve in Figure 4-6. For those signals, adding more change points may improve individual model fitting at the offline stage. However, the addition of extra change points will also bring uncertainties on the prediction of the location and the degradation amplitude of the final change point, which may significantly reduce the prognostic accuracy.
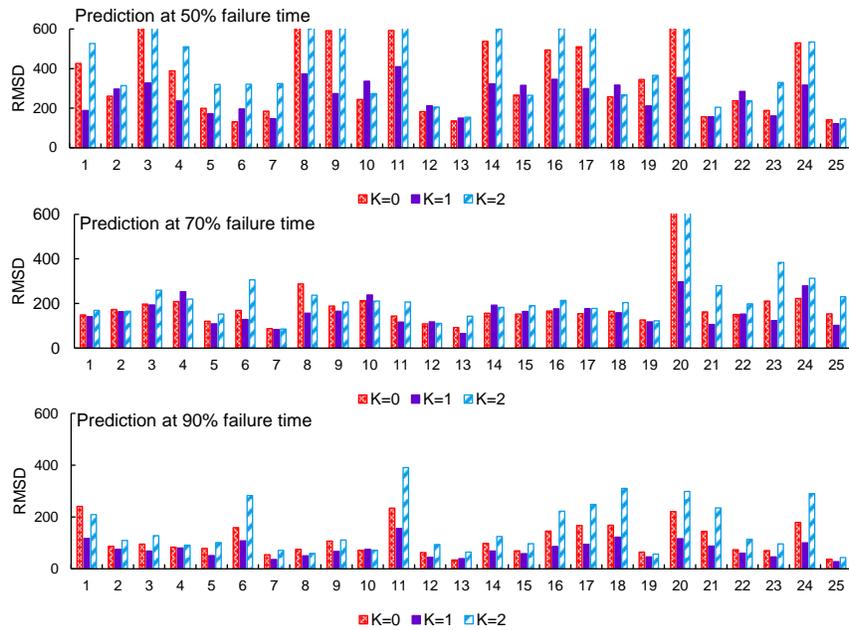
Figure 4-8: Comparison of RMSD for three models.
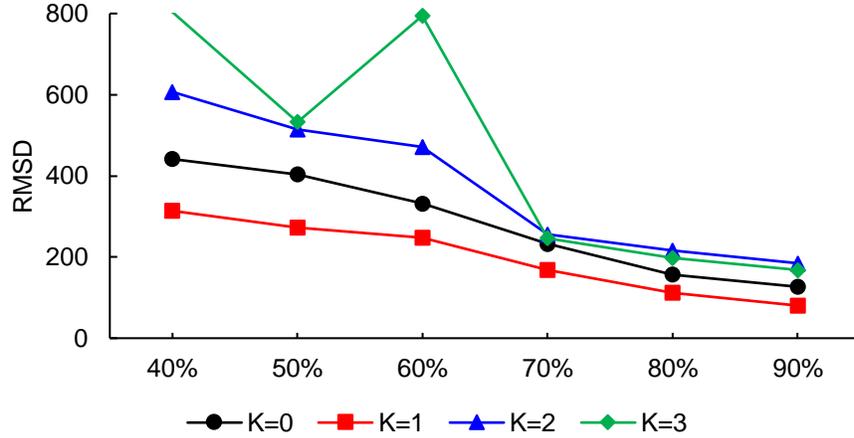
Figure 4-9: Prediction accuracy at different times for $K = 0, 1, 2$ and 3.

4.6 CONCLUSION

In this chapter, a Bayesian multiple change-point Wiener process is proposed for degradation modeling and online RUL prediction. To take into account the unit heterogeneity, all the model parameters except the number of change points are modeled with random distributions. At the offline stage, an empirical two-stage process is proposed for model estimation. Besides, a cross-validation approach is proposed for optimal change-point number selection. At the online monitoring stage, an exact recursive updating method is developed to sequentially calculate the joint posterior distribution of key parameters, including the latest change point, phase index, and model parameters of the current phase, which is essential for RUL prediction. To control the computational cost, a fixed-support-size strategy is proposed, which can effectively control and balance the computational load of each time step yet without influencing the estimation accuracy. In RUL prediction, an effective Monte Carlo simulation algorithm is proposed. Simulation and real case studies demonstrate that the proposed prognostic framework can effectively improve the prediction accuracy.

There are still some issues that are worthy of further investigation. First, measurement errors may occur in the data collection processes. Taking the measurement noise into account may improve the modeling and prognostic accuracy. In addition, in the current method, a linear drift function is assumed for all phases. In practice, however, nonlinear drift functions or a mixture of linear and nonlinear drift functions may be more preferable. It may significantly reduce the model complexity and consequently the model uncertainty. Furthermore, other stochastic processes, such as inverse Gaussian process, and Gamma process, may model the degradation signal better. How to incorporate change points into these stochastic processes needs to be investigated[*].

Chapter 5: Conclusions and Recommendations for Future Work

Prognostics has played an increasingly important role in modern engineering systems due to its capability of reducing maintenance costs, improving operational efficiency, and facilitating decision making. However, Due to the inherent uncertainty of predicting the future, accurate prognostics is challenging. The fast development of information and sensing technology has enabled the collection of in situ degradation signals during operations and provide the capability of real time individualized prognostics. To obtain accurate RUL prediction, selecting an accurate and effective prognostic model for the degradation signals is the key. In many practical applications, the degradation signals consist of multiple distinct phases, traditional models are inadequate to characterize the degradation processes. To overcome this critical issue, this research focuses on establishing a series of new data analytics methods for effective and accurate data-driven degradation modeling and real time individualized prognostics. The contributions of this dissertation can by summarized into the following categories:

First, the proposed multiple change-point modeling method, can address the multiple phase characteristic of the degradation data and provide an accurate model for RUL prediction. Second, we use the Bayesian framework to efficiently integrate the degradation information from historical data with in situ observations of each new unit in operation. Third, a novel non-standard state-space model is reformulated from multiple change-point modeling and a stratified particle filtering algorithm is developed to track the current status and predict the future trend of degradation signals in a real time manner. Moreover, to reduce the computational cost, an exact Bayesian inference approach is proposed so that the distribution of RUL can be calculated in a closed form. We also develop a multiple change-point Wiener process degradation model to capture the temporal uncertainties that are inherent in the degradation process. The advantages of

104

our methods have been demonstrated using extensive numerical studies from both simulation and real dataset experiments. This research establishes a new direction by proposing a multiple change-point modeling technique. The methodologies developed in this dissertation can lead to a great impact on data-driven degradation modeling and prognostic application.

Nevertheless, there are still some improvement that can be further investigated for the optimal decision making. First, in all developed models, failure threshold is set to be constant and deterministic for all units. However, given the stochastic nature of the underlying degradation mechanism or multiple potential failure modes, such a simplified assumption may cause under or over-estimate the actual RUL. Incorporating time-to-event data concerning the failure or censoring information and degradation signals could be an interesting topic to investigate. Second, the rapid advancements of internet of things technology have resulted in data explosion, which provides unprecedented opportunities for performance improvement in various complex systems, Meanwhile, it also raises new research challenges on data analysis and decision making, such as heterogeneous data formats, high-dimensional data structures, fast-flowing data streams. It would be of interest to investigate advanced data analytic methods for Big data stream. These will be studied and reported in future articles.

Appendix

Appendix A: Proof of Lemma 2.1 in chapter 2

For the sake of simplicity, we ignore all the superscripts (k and s) in the proof.

(1) If $\boldsymbol{\beta}_t \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, then

$$p(\boldsymbol{\beta}_t | y_{1:t}, \sigma_t^2, \tau_t) \propto p(\boldsymbol{\beta}_t) p(y_{\tau_t+1:t} | \tau_t, \sigma_t^2, \boldsymbol{\beta}_t)$$

$$\propto \exp\left[-\frac{(\boldsymbol{\beta}_t - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta}_t - \boldsymbol{\mu}_0)}{2}\right] \cdot \exp\left(-\frac{\left\|y_{\tau_t+1:t}^T - X_{\tau_t+1:t}\boldsymbol{\beta}_t\right\|^2}{2\sigma_t^2}\right)$$

$$\propto \exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}_t^T\left(\frac{X_{\tau_t+1:t}^T X_{\tau_t+1:t}}{\sigma_t^2} + \boldsymbol{\Sigma}_0^{-1}\right)\boldsymbol{\beta}_t - 2\left(\boldsymbol{\mu}_0^T\boldsymbol{\Sigma}_0^{-1} + \frac{y_{\tau_t+1:t}X_{\tau_t+1:t}}{\sigma_t^2}\right)\boldsymbol{\beta}_t\right]\right\}$$

$$\propto \exp\left[-\frac{1}{2}(\boldsymbol{\beta}_t - \boldsymbol{\mu}_t)^T \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\beta}_t - \boldsymbol{\mu}_t)\right]$$

where

$$\boldsymbol{\mu}_t = \left[\frac{X_{\tau_t+1:t}^T X_{\tau_t+1:t}}{\sigma_t^2} + \boldsymbol{\Sigma}_0^{-1}\right]^{-1}\left[\frac{X_{\tau_t+1:t}^T y_{\tau_t+1:t}^T}{\sigma_t^2} + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0\right]$$

$$\boldsymbol{\Sigma}_t = \left[\frac{X_{\tau_t+1:t}^T X_{\tau_t+1:t}}{\sigma_t^2} + \boldsymbol{\Sigma}_0^{-1}\right]^{-1}$$

Therefore

$$(\boldsymbol{\beta}_t | y_{1:t}, \sigma_t^2, \tau_t) \sim N(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$$

If $\boldsymbol{\beta}_t \sim TN(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0 | b_t > l)$, then similarly,

$$p(\boldsymbol{\beta}_t | y_{1:t}, \sigma_t^2, \tau_t) \propto p(\boldsymbol{\beta}_t) p(y_{\tau_t+1:t} | \tau_t, \sigma_t^2, \boldsymbol{\beta}_t)$$

$$\propto \exp\left[-\frac{(\boldsymbol{\beta}_t - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta}_t - \boldsymbol{\mu}_0)}{2}\right] \cdot I_{(l,+\infty)}(b_t) \cdot \exp\left(-\frac{\left\|y_{\tau_t+1:t}^T - X_{\tau_t+1:t}\boldsymbol{\beta}_t\right\|^2}{2\sigma_t^2}\right)$$

$$\propto \exp\left[-\frac{1}{2}(\boldsymbol{\beta}_t - \boldsymbol{\mu}_t)^T \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{\beta}_t - \boldsymbol{\mu}_t)\right] \cdot I_{(l,+\infty)}(b_t)$$

where $I_{(l,+\infty)}(\cdot)$ is an indicator function, i.e., $I_{(l,+\infty)}(x) = 0$ if $x \le l$ and $I_{(l,+\infty)}(x) = 1$ if

$x > l$

Therefore

$(\boldsymbol{\beta}_t | y_{1:t}, \sigma_t^2, \tau_t) \sim TN(\boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t | b_t > l)$

(2) $p(\sigma_t^2 | \boldsymbol{\beta}_t, y_{1:t}, \tau_t) \propto p(\sigma_t^2) p(y_{\tau_t+1:t} | \boldsymbol{\beta}_t, \tau_t, \sigma_t^2)$

$\propto (\sigma_t^2)^{-(\alpha_1+1)} \exp\left(-\frac{\alpha_2}{\sigma_t^2}\right) (\sigma_t^2)^{-\frac{t-\tau_t}{2}} \exp\left(-\frac{\left\|y_{\tau_t+1:t}^T - X_{\tau_t+1:t}\boldsymbol{\beta}_t\right\|^2}{2\sigma_t^2}\right)$

$\propto (\sigma_t^2)^{-\left(\alpha_1 + \frac{t-\tau_t}{2} - 1\right)} \exp\left(-\frac{\alpha_2 + \left\|y_{\tau_t+1:t}^T - X_{\tau_t+1:t}\boldsymbol{\beta}_t\right\|^2/2}{\sigma_t^2}\right)$

Therefore

$$(\sigma_t^2 | \boldsymbol{\beta}_t, y_{1:t}, \tau_t) \sim IG\left(\alpha_1 + \frac{t-\tau_t}{2}, \alpha_2 + \frac{\left\|y_{\tau_t+1:t}^T - X_{\tau_t+1:t}\boldsymbol{\beta}_t\right\|^2}{2}\right)$$

Appendix B: Derivation of Equation (3.9) in chapter 3

For notational convenience, we ignore the superscript k and s here. Suppose $\{\boldsymbol{\beta}_i, \sigma_i^2\}_{i=1}^{i=n}$ are the observed samples from distribution

$$\pi(\boldsymbol{\beta}, \sigma^2) = IG(\sigma^2 | \alpha_1, \alpha_2) N(\boldsymbol{\mu_0}, \sigma^2 \boldsymbol{\Sigma_0})$$

The likelihood function can be written as

$L\left(\alpha_1, \alpha_2, \boldsymbol{\mu_0}, \boldsymbol{\Sigma_0} \big| \{\boldsymbol{\beta}_i, \sigma_i^2\}_{i=1}^{i=n}\right)$

$\propto \prod_{i=1}^{n}\left[\frac{\alpha_2^{\alpha_1}}{\Gamma(\alpha_1)}(\sigma_i^2)^{-\alpha_1-1}\exp\left(-\frac{\alpha_2}{\sigma_i^2}\right)\right]\left|\sigma_i^2\boldsymbol{\Sigma_0}\right|^{-\frac{1}{2}}\exp\left[-\frac{(\boldsymbol{\beta}_i - \boldsymbol{\mu_0})'\boldsymbol{\Sigma_0}^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\mu_0})}{2\sigma_i^2}\right]$

The log-likelihood function l is

$$l\left(\alpha_1, \alpha_2, \boldsymbol{\mu_0}, \boldsymbol{\Sigma_0} \middle| \{\boldsymbol{\beta}_i, \sigma_i^2\}_{i=1}^{i=n}\right)$$

$$= \sum_{i=1}^{n} \log\left[\frac{\alpha_2^{\alpha_1}}{\Gamma(\alpha_1)} (\sigma_i^2)^{-\alpha_1-1} \exp\left(-\frac{\alpha_2}{\sigma_i^2}\right)\right] - \frac{n}{2}\log|\boldsymbol{\Sigma_0}|$$

$$- \sum_{i=1}^{n} \frac{(\boldsymbol{\beta}_i - \boldsymbol{\mu_0})'\boldsymbol{\Sigma_0}^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\mu_0})}{2\sigma_i^2} + C$$

Therefore the MLE of α_1, α_2 can be obtained by maximizing the log-likelihood function $l(\alpha_1, \alpha_2 | \{\sigma_i^2\}_{i=1}^{i=n})$. For $\boldsymbol{\mu_0}, \boldsymbol{\Sigma_0}$, the log-likelihood function is

$$l\left(\boldsymbol{\mu_0}, \boldsymbol{\Sigma_0} \middle| \{\boldsymbol{\beta}_i, \sigma_i^2\}_{i=1}^{i=n}\right) = -\frac{n}{2}\log|\boldsymbol{\Sigma_0}| - \sum_{i=1}^{n} \frac{(\boldsymbol{\beta}_i - \boldsymbol{\mu_0})'\boldsymbol{\Sigma_0}^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\mu_0})}{2\sigma_i^2} + C$$

$$\frac{\partial l}{\partial \boldsymbol{\mu_0}} = -\sum_{i=1}^{n} \frac{\boldsymbol{\Sigma_0}^{-1}(\boldsymbol{\mu_0} - \boldsymbol{\beta}_i)}{\sigma_i^2} = 0$$

Therefore

$$\widehat{\boldsymbol{\mu}}_0 = \frac{\sum_{i=1}^{n} \frac{\boldsymbol{\beta}_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$$

Plug in $\widehat{\boldsymbol{\mu}}_0$ we can get

$$l\left(\widehat{\boldsymbol{\mu}}_0, \boldsymbol{\Sigma_0} \middle| \{\boldsymbol{\beta}_i, \sigma_i^2\}_{i=1}^{i=n}\right)$$

$$= -\frac{n}{2}\log|\boldsymbol{\Sigma_0}| - \sum_{i=1}^{n} \frac{(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\mu}}_0)'\boldsymbol{\Sigma_0}^{-1}(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\mu}}_0)}{2\sigma_i^2} + C$$

$$= -\frac{n}{2}\log|\boldsymbol{\Sigma_0}| - \frac{1}{2}\text{tr}\left(\sum_{i=1}^{n} \frac{\boldsymbol{\Sigma_0}^{-1}(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\mu}}_0)(\boldsymbol{\beta}_i - \widehat{\boldsymbol{\mu}}_0)'}{\sigma_i^2}\right)$$

$$= -\frac{n}{2}\log|\boldsymbol{\Sigma_0}| - \frac{n}{2}\text{tr}\left(\boldsymbol{\Sigma_0}^{-1}S\right)$$

where

$$S = \frac{1}{n} \sum_{i=1}^{n} \frac{(\boldsymbol{\beta}_i - \hat{\boldsymbol{\mu}}_0)(\boldsymbol{\beta}_i - \hat{\boldsymbol{\mu}}_0)'}{\sigma_i^2}$$

It can be shown that $\hat{\boldsymbol{\Sigma}}_0 = S$.

Let $S = EE^T$, $\Psi = E^T \boldsymbol{\Sigma}_0^{-1} E$, then $|\boldsymbol{\Sigma}_0| = \frac{|S|}{|\Psi|}$, $\text{tr}(\boldsymbol{\Sigma}_0^{-1} S) = \text{tr}(\boldsymbol{\Sigma}_0^{-1} E E^T) = \text{tr}(E^T \boldsymbol{\Sigma}_0^{-1} E) = \text{tr}(\Psi)$.

So $l\left(\boldsymbol{\Sigma}_0, \hat{\boldsymbol{\mu}}_0 \big| \{\boldsymbol{\beta}_i, \sigma_i^2\}_{i=1}^{i=n}\right) = -\frac{n}{2}\left(\log|S| - \log|\Psi| + \text{tr}(\Psi)\right)$

Let $\Psi = LL^T$ where L is a lower triangular matrix

$$l\left(\boldsymbol{\Sigma}_0, \hat{\boldsymbol{\mu}}_0 \big| \{\boldsymbol{\beta}_i, \sigma_i^2\}_{i=1}^{i=n}\right)$$

$$= -\frac{n}{2}\left(\log|S| - \sum_{i=1}^{2} \log l_{ii}^2 + \sum_{i=1}^{2} l_{ii}^2 + \sum_{i>j} l_{ij}^2\right)$$

It is easy to show that when $l_{ij} = 0$ for $i \neq j$ and $l_{ii}^2 = 1$, $l\left(\boldsymbol{\Sigma}_0, \hat{\boldsymbol{\mu}}_0 \big| \{\boldsymbol{\beta}_i, \sigma_i^2\}_{i=1}^{i=n}\right)$ reaches the maximum

Therefore $\hat{\Psi} = I$, $\hat{\boldsymbol{\Sigma}}_0 = S$

Appendix C: Proof of Theorem 3.1 in chapter 3

For notational convenience, we ignore the superscripts k and s here. We also ignore the subscript t for $\boldsymbol{\beta}_t$ and σ_t^2. Suppose $\pi(\boldsymbol{\beta}, \sigma^2) = IG(\sigma^2|\alpha_1, \alpha_2)N(\boldsymbol{\mu}_0, \sigma^2 \boldsymbol{\Sigma}_0)$ and $\boldsymbol{\beta}$ is of dimension q.

$$p(\boldsymbol{\beta}, \sigma^2 | y_{j+1:t}) \propto p(\boldsymbol{\beta}, \sigma^2) p(y_{j+1:t} | \boldsymbol{\beta}, \sigma^2)$$

$$\propto \left[\frac{1}{2\pi|\sigma^2 \boldsymbol{\Sigma}_0|^{\frac{1}{2}}} e^{-\frac{(\boldsymbol{\beta} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_0)}{2\sigma^2}}\right]\left[\frac{\alpha_2^{\alpha_1}}{\Gamma(\alpha_1)}(\sigma^2)^{-\alpha_1 - 1} e^{-\frac{\alpha_2}{\sigma^2}}\right]\left[(2\pi)^{-\frac{t-j}{2}}(\sigma^2)^{-\frac{t-j}{2}} e^{\frac{-\|y_{j+1:t} - X_{1,t-j}\boldsymbol{\beta}\|^2}{2\sigma^2}}\right]$$

$$\propto (\sigma^2)^{-\alpha_1 - 1 - \frac{t-j}{2}} \frac{1}{2\pi(\sigma^2)^{\frac{q}{2}}} \exp\left[-\frac{(\boldsymbol{\beta} - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_0) + 2\alpha_2 + \|y_{j+1:t} - \boldsymbol{X}_{1:t-j}\boldsymbol{\beta}\|^2}{2\sigma^2}\right]$$

$$\propto (\sigma^2)^{-\alpha_1 - \frac{(t-j)}{2} - 1}$$

$$\exp\left[-\frac{y_{j+1:t}^T y_{j+1:t} + 2\alpha_2 + \boldsymbol{\mu}_0^T \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 - \boldsymbol{\mu}_{j+1:t}^T(\boldsymbol{X}_{1:t-j}^T\boldsymbol{X}_{1:t-j} + \boldsymbol{\Sigma}_0^{-1})\boldsymbol{\mu}_{j+1:t}}{2\sigma^2}\right]$$

$$\times \frac{1}{(\sigma^2)^{q/2}} \exp\left[-\frac{(\boldsymbol{\beta} - \boldsymbol{\mu}_{j+1:t})^T(\boldsymbol{X}_{1:t-j}^T\boldsymbol{X}_{1:t-j} + \boldsymbol{\Sigma}_0^{-1})(\boldsymbol{\beta} - \boldsymbol{\mu}_{j+1:t})}{2\sigma^2}\right]$$

$$\propto IG\left(\alpha_1 + \frac{t-j}{2}, \alpha_2 + \frac{H_{j+1,t}}{2}\right) \cdot N(\boldsymbol{\mu}_{j+1,t}, \sigma^2\boldsymbol{\Sigma}_{j+1,t})$$

where

$$\boldsymbol{\Sigma}_{j+1,t} = \left(\boldsymbol{X}_{1,t-j}^T\boldsymbol{X}_{1,t-j} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}$$

$$\boldsymbol{N}_{j+1,t} = \left(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0^{(k,s)} + \boldsymbol{X}_{1,t-j}^T y_{j+1:t}\right)$$

$$\boldsymbol{\mu}_{j+1,t} = \boldsymbol{\Sigma}_{j+1,t}\boldsymbol{N}_{j+1,t}$$

$$H_{j+1,t} = y_{j+1:t}^T y_{j+1:t} + \boldsymbol{\mu}_0^T\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 - \boldsymbol{N}_{j+1,t}^T\boldsymbol{\Sigma}_{j+1,t}\boldsymbol{N}_{j+1,t}$$

Appendix D: Proof of Theorem 3.2 in chapter 3

If $\tau_t = j < t - 1$, based on Theorem 3.1 we can get

$$\left(\boldsymbol{\beta}_{t-1} | \sigma_{t-1}^2, \tau_{t-1} = j, s_{t-1} = s, k, y_{j+1:t-1}\right) \sim N(\boldsymbol{\mu}_{j+1,t-1}, \sigma_{t-1}^2\boldsymbol{\Sigma}_{j+1,t-1})$$

$$\left(\boldsymbol{X}_{t-j}\boldsymbol{\beta}_{t-1} | \sigma_{t-1}^2, \tau_{t-1} = j, s_{t-1} = s, k, y_{j+1:t-1}\right) \sim N(\boldsymbol{X}_{t-j}\boldsymbol{\mu}_{j+1,t-1}, \sigma_{t-1}^2\boldsymbol{X}_{t-j}\boldsymbol{\Sigma}_{j+1,t-1}\boldsymbol{X}_{t-j}^T)$$

$$\left(\sigma_{t-1}^2 | \tau_{t-1} = j, s_{t-1} = s, k, y_{j+1:t-1}\right) \sim IG\left(\alpha_1^{(k,s)} + \frac{t-1-j}{2}, \alpha_2^{(k,s)} + \frac{H_{j+1,t-1}}{2}\right)$$

Since

$$y_t = \boldsymbol{X}_{t-j}\boldsymbol{\beta}_{t-1} + \sigma_{t-1}\varepsilon_t$$

for $\tau_t = \tau_{t-1} = j < t - 1$,

then

$$\left(y_t \middle| \sigma_{t-1}^2, \tau_{t-1} = j, s_{t-1} = k, y_{j+1:t-1}\right) \sim N\left(\boldsymbol{X}_{t-j}\boldsymbol{\mu}_{j+1,t-1}, \sigma_{t-1}^2(1 + \boldsymbol{X}_{t-j}\boldsymbol{\Sigma}_{j+1,t-1}\boldsymbol{X}_{t-j}^{\mathrm{T}})\right),$$

Let $\mu_* = \boldsymbol{X}_{t-j}\boldsymbol{\mu}_{j+1,t-1}$, $\sigma_*^2 = 1 + \boldsymbol{X}_{t-j}\boldsymbol{\Sigma}_{j+1,t-1}\boldsymbol{X}_{t-j}^{\mathrm{T}}$

$$p(y_t | \tau_t = j, s_t = s, k, y_{1:t-1}) = \int p(y_t | \sigma_{t-1}^2, \tau_{t-1} = j, s_{t-1} = s, k, y_{1:t-1})$$
$$p(\sigma_{t-1}^2 | \tau_{t-1} = j, s_{t-1} = s, k, y_{1:t-1}) d\sigma_{t-1}^2$$

$$\propto \int (\sigma_{t-1}^2 \sigma_*^2)^{-\frac{1}{2}} \exp\left[-\frac{(y_t - \mu_*)^2}{2\sigma_{t-1}^2 \sigma_*^2}\right] (\sigma_{t-1}^2)^{-\alpha_1^{(k,s)} - \frac{(t-j-1)}{2} - 1}$$
$$\exp\left[-\frac{2\alpha_2^{(k,s)} + H_{j+1:t-1}}{2\sigma_{t-1}^2}\right] d\sigma_{t-1}^2$$

$$\propto \int (\sigma_{t-1}^2)^{-\alpha_1^{(k,s)} - \frac{(t-j)}{2} - 1} \exp\left(-\frac{(y_t - \mu_*)^2 + (2\alpha_2^{(k,s)} + H_{j+1:t-1})\sigma_*^2}{2\sigma_{t-1}^2 \sigma_*^2}\right) d\sigma_{t-1}^2$$

$$\propto \frac{\Gamma\left(\alpha_1^{(k,s)} + \frac{(t-j)}{2}\right)}{\left[\frac{(y_t - \mu_*)^2 + (2\alpha_2^{(k,s)} + H_{j+1:t-1})\sigma_*^2}{2\sigma_*^2}\right]^{\alpha_1^{(k,s)} + \frac{(t-j)}{2}}}$$

$$\propto \left[1 + \frac{1}{v}\frac{(y_t - \mu_*)^2 v}{\sigma_*^2(2\alpha_2^{(k,s)} + H_{j+1:t-1})}\right]^{-\frac{1+v}{2}}$$

where $v = 2\alpha_1^{(k,s)} + t - j - 1$. Therefore

$$\left(y_t|\tau_t = j, s_t = s, k, y_{j+1:t-1}\right) \sim t_1\left(2\alpha_1^{(k,s)} + t - j\right.$$

$$\left. - 1, \mathbf{X}_{t-j}\boldsymbol{\mu}_{j+1,t-1}, \frac{(2\alpha_2^{(k,s)} + H_{j+1:t-1})(1 + \mathbf{X}_{t-j}\boldsymbol{\Sigma}_{j+1,t-1}\mathbf{X}_{t-j}^{\mathrm{T}})}{2\alpha_1^{(k,s)} + t - j - 1}\right)$$

The proof for $j = t - 1$ is similar to the above derivation process and is neglected here.

Appendix E: Proof of Theorem 4.1 in chapter 4

For notational convenience, we ignore the subscript m for β_m and σ_m^2, and ignore the subscript s for phase index. Suppose $\pi(\beta, \sigma^2) = IG(v_0, \gamma_0)N(\mu_0, \sigma^2\kappa_0^2)$

$$p(\beta, \sigma^2|\boldsymbol{\delta}_{j+1:m}, \tau_m = j) \propto p(\beta, \sigma^2)p(\boldsymbol{\delta}_{j+1:m}|\beta, \sigma^2)$$

$$\propto \left[\frac{1}{(2\pi\sigma^2\kappa_0^2)^{\frac{1}{2}}}e^{-\frac{(\beta-\mu_0)^2}{2\sigma^2\kappa_0^2}}\right]\left[\frac{\gamma_0^{v_0}}{\Gamma(v_0)}(\sigma^2)^{-v_0-1}e^{-\frac{\gamma_0}{\sigma^2}}\right]\left[(2\pi)^{-\frac{m-j}{2}}(\sigma^2)^{-\frac{m-j}{2}}e^{\frac{-\sum_{i=j+1}^{m}(\delta_i-\beta\lambda_i)^2/\lambda_i}{2\sigma^2}}\right]$$

$$\propto (\sigma^2)^{-v_0-1-\frac{m-j}{2}}\frac{1}{2\pi(\sigma^2)^{1/2}}\exp\left[-\frac{(\beta-\mu_0)^2/\kappa_0^2 + 2\gamma_0 + \sum_{i=j+1}^{m}(\delta_i-\beta\lambda_i)^2/\lambda_i}{2\sigma^2}\right]$$

$$\propto (\sigma^2)^{-v_0-1-\frac{m-j}{2}}\frac{1}{2\pi(\sigma^2)^{\frac{1}{2}}}\exp\left(-\frac{\beta^2\left(\sum_{i=j+1}^{m}\lambda_i + \frac{1}{\kappa_0^2}\right) - 2\beta\left(\frac{\mu_0}{\kappa_0^2} + \sum_{i=j+1}^{m}\delta_i\right) + \frac{\mu_0^2}{\kappa_0^2} + 2\gamma_0 + \sum_{i=j+1}^{m}\frac{\delta_i^2}{\lambda_i}}{2\sigma^2}\right)$$

$$\propto (\sigma^2)^{-v_0-1-\frac{m-j}{2}}\frac{1}{2\pi(\sigma^2)^{\frac{1}{2}}}\exp\left(-\frac{(\beta-\mu_m)^2}{2\sigma^2\left(\sum_{i=1}^{m}\lambda_{j+i} + \frac{1}{\kappa_0^2}\right)^{-1}}\right)$$

$$\exp\left(-\frac{\frac{\mu_0^2}{\kappa_0^2} + 2\gamma_0 + \sum_{i=j+1}^{m}\frac{\delta_i^2}{\lambda_i} - \left(\frac{\mu_0}{\kappa_0^2} + \sum_{i=j+1}^{m}\delta_i\right)^2\left(\sum_{i=j+1}^{m}\lambda_i + \frac{1}{\kappa_0^2}\right)^{-1}}{2\sigma^2}\right)$$

$$\propto IG\left(v_{j+1,m}, \gamma_{j+1,m}\right)N\left(\mu_{j+1,m}, \sigma^2\kappa_{j+1,m}^2\right)$$

where

$$v_{j+1,m} = v_0 + \frac{m-j}{2}$$

$$\gamma_{j+1,m} = \gamma_0 + \frac{\left[\frac{\mu_0^2}{\kappa_0^2} + \sum_{i=j+1}^{m} \frac{\delta_i^2}{\lambda_i} - \left(\frac{\mu_0}{\kappa_0^2} + \sum_{i=j+1}^{m} \delta_i\right)^2 \left(\sum_{i=j+1}^{m} \lambda_i + \frac{1}{\kappa_0^2}\right)^{-1}\right]}{2}$$

$$\mu_{j+1,m} = \left(\frac{\mu_0}{\kappa_0^2} + \sum_{i=j+1}^{m} \delta_i\right)\left(\sum_{i=j+1}^{m} \lambda_i + \frac{1}{\kappa_0^2}\right)^{-1}$$

$$\kappa_{j+1,m}^2 = \left(\sum_{i=j+1}^{m} \lambda_i + \frac{1}{\kappa_0^2}\right)^{-1}$$

Appendix F: Proof of Theorem 4.2 in chapter 4

For notational convenience, we ignore the subscript s for phase index. Based on Wiener process increment property

$$X_m = X_{m-1} + \beta_m \lambda_m + \sigma_m B(\lambda_m)$$

Note that here β_m and σ_m are the drift and diffusion parameters from t_{τ_m} to t_m.

If $\tau_m = j < m - 1$, then $\tau_{m-1} = j, \beta_m = \beta_{m-1}, \sigma_m = \sigma_{m-1}$. Based on Theorem 1 we can get

$$(\beta_{m-1}|\sigma_{m-1}^2, \tau_{m-1} = j, s_{m-1} = s, X_{0:m-1}) \sim N\left(\mu_{j+1,m-1}, \sigma_{m-1}^2 \kappa_{j+1,m-1}^2\right)$$

$$(\sigma_{m-1}^2|\tau_{m-1} = j, s_{m-1} = s, X_{0:m-1}) \sim IG\left(v_{j+1,m-1}, \gamma_{j+1,m-1}\right)$$

Then

$$(X_m|\sigma_{m-1}^2, \tau_m = \tau_{m-1} = j, s_{m-1} = s, X_{0:m-1}) \sim N\left(\mu_{j+1,m-1}\lambda_m\right.$$

$$\left. + X_{m-1}, \sigma_{m-1}^2\left(\lambda_m + \kappa_{j+1,m-1}^2\lambda_m^2\right)\right)$$

Let $\mu' = \mu_{j+1,m-1}\lambda_m + X_{m-1}$, $\eta' = \lambda_m + \kappa_{j+1,m-1}^2\lambda_m^2$, then

$$p(X_m|\tau_m = j, s_m = s, X_{0:m-1})$$

$$\propto \int_0^\infty (\sigma_{m-1}^2 \eta')^{-\frac{1}{2}} e^{\left[\frac{-(X_m-\mu')^2}{2\sigma_{m-1}^2 \eta'}\right]} (\sigma_{m-1}^2)^{-v_0-\frac{m-1-j}{2}-1} e^{-\frac{\gamma_{j+1,m-1}}{\sigma_{m-1}^2}} d\sigma_{m-1}^2$$

$$\propto (\sigma_{m-1}^2)^{-v_0-\frac{m-j}{2}-1} e^{\left[-\frac{(X_m-\mu')^2+2\gamma_{j+1,m-1}\eta'}{2\sigma_{m-1}^2 \eta'}\right]} d\sigma_{m-1}^2$$

$$\propto \frac{\Gamma(v_0 + \frac{m-j}{2})}{\left[\frac{(X_m-\mu')^2}{2\eta'} + \gamma_{j+1,m-1}\right]^{v_0+\frac{m-j}{2}}}$$

$$\propto \left[1 + \frac{1}{2v_0 + m - j - 1}\frac{(X_m - \mu')^2(2v_0 + m - j - 1)}{2\gamma_{j+1,m-1}\eta'}\right]^{-\frac{(2v_0+m-j)}{2}}$$

$$(X_m|X_{0:m-1}, \tau_m = j) \sim t_1\left(2v_0 + m - j - 1, \mu', \frac{2\gamma_{j+1,m-1}\eta'}{2v_0 + m - j - 1}\right)$$

If $j = m - 1$, similarly, we can get

$$(X_m|X_{0:m-1}, \tau_m = m - 1) \sim t_1\left(2v_0, \mu', \frac{\gamma_0\eta'}{v_0}\right)$$

where $\mu' = \mu_0\lambda_m + X_{m-1}$, $\eta' = \lambda_m + \kappa_0^2\lambda_m^2$

References

[1] N. Gorjian, L. Ma, M. Mittinty, P. Yarlagadda, and Y. Sun, "A review on degradation models in reliability analysis," 2009.

[2] H. M. Elattar, H. K. Elminir, and A. Riad, "Prognostics: a literature review," *Complex & Intelligent Systems,* vol. 2, no. 2, pp. 125-154, 2016.

[3] K. L. Tsui, N. Chen, Q. Zhou, Y. Hai, and W. Wang, "Prognostics and health management: A review on data driven approaches," *Mathematical Problems in Engineering,* vol. 2015, 2015.

[4] K. Yang and J. Xue, "Continuous state reliability analysis," in *Reliability and Maintainability Symposium, 1996 Proceedings. International Symposium on Product Quality and Integrity., Annual*, 1996: IEEE, pp. 251-257.

[5] J. W. McPherson, *Reliability physics and engineering: time-to-failure modeling*. Springer Science & Business Media, 2013.

[6] M. J. Crowder, A. Kimber, T. Sweeting, and R. Smith, *Statistical analysis of reliability data*. CRC Press, 1994.

[7] Z. Zhang, X. Si, C. Hu, and X. Kong, "Degradation modeling–based remaining useful life estimation: A review on approaches for systems with heterogeneity," *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability,* vol. 229, no. 4, pp. 343-355, 2015.

[8] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing,* vol. 20, no. 7, pp. 1483-1510, 2006.

[9] A. H. Elwany and N. Z. Gebraeel, "Sensor-driven prognostic models for equipment replacement and spare parts inventory," *IIE Transactions,* vol. 40, no. 7, pp. 629-639, 2008.

[10] A. Heng, A. C. Tan, J. Mathew, N. Montgomery, D. Banjevic, and A. K. Jardine, "Intelligent condition-based prediction of machinery reliability," *Mechanical Systems and Signal Processing,* vol. 23, no. 5, pp. 1600-1614, 2009.

[11] Z. S. Ye and M. Xie, "Stochastic modelling and analysis of degradation for highly reliable products," *Applied Stochastic Models in Business and Industry,* vol. 31, no. 1, pp. 16-32, 2015.

[12] N. Z. Gebraeel, M. A. Lawley, R. Li, and J. K. Ryan, "Residual-life distributions from component degradation signals: A Bayesian approach," *IIE Transactions,* vol. 37, no. 6, pp. 543-557, 2005.

[13] J. Sikorska, M. Hodkiewicz, and L. Ma, "Prognostic modelling options for remaining useful life estimation by industry," *Mechanical Systems and Signal Processing,* vol. 25, no. 5, pp. 1803-1836, 2011.

[14] A. Heng, S. Zhang, A. C. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges and opportunities," *Mechanical Systems and Signal Processing,* vol. 23, no. 3, pp. 724-739, 2009.

[15] G. J. Vachtsevanos, F. Lewis, A. Hess, and B. Wu, *Intelligent fault diagnosis and prognosis for engineering systems*. Wiley Online Library, 2006.

[16] J. Luo, M. Namburu, K. Pattipati, L. Qiao, M. Kawamoto, and S. Chigusa, "Model-based prognostic techniques [maintenance applications]," in *AUTOTESTCON 2003. IEEE Systems Readiness Technology Conference. Proceedings*, 2003: IEEE, pp. 330-340.

[17]   C. J. Li and H. Lee, "Gear fatigue crack prognosis using embedded model, gear dynamic model and fracture mechanics," *Mechanical Systems and Signal Processing,* vol. 19, no. 4, pp. 836-846, 2005.

[18]   G. Kacprzynski, A. Sarlashkar, M. Roemer, A. Hess, and B. Hardman, "Predicting remaining life by fusing the physics of failure modeling with diagnostics," *JOM Journal of the Minerals, Metals and Materials Society,* vol. 56, no. 3, pp. 29-35, 2004.

[19]   J. Luo, K. R. Pattipati, L. Qiao, and S. Chigusa, "Model-based prognostic techniques applied to a suspension system," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans,* vol. 38, no. 5, pp. 1156-1168, 2008.

[20]   C. H. Oppenheimer and K. A. Loparo, "Physically based diagnosis and prognosis of cracked rotor shafts," in *AeroSense 2002*, 2002: International Society for Optics and Photonics, pp. 122-132.

[21]   Y. Peng, M. Dong, and M. J. Zuo, "Current status of machine prognostics in condition-based maintenance: a review," *The International Journal of Advanced Manufacturing Technology,* vol. 50, no. 1, pp. 297-313, 2010.

[22]   F. Camci, K. Medjaher, N. Zerhouni, and P. Nectoux, "Feature evaluation for effective bearing prognostics," *Quality and Reliability Engineering International,* vol. 29, no. 4, pp. 477-486, 2013.

[23]   E. Zio and F. Di Maio, "A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system," *Reliability Engineering & System Safety,* vol. 95, no. 1, pp. 49-57, 2010.

[24]   O. E. Dragomir, R. GOURTVEAU, N. Zerhouni, and F. Dragomir, "Framework for a distributed and hybrid prognostic system," *IFAC Proceedings Volumes,* vol. 40, no. 18, pp. 431-436, 2007.

[25]   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,* vol. 213, no. 1, pp. 1-14, 2011.

[26]   Y. Wen *et al.*, "Performance evaluation of probabilistic methods based on bootstrap and quantile regression to quantify PV power point forecast uncertainty," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 31, no. 4, pp. 1134-1144, 2019.

[27]   J. Lee, "Measurement of machine performance degradation using a neural network model," *Computers in Industry,* vol. 30, no. 3, pp. 193-209, 1996.

[28]   N. Gebraeel, M. Lawley, R. Liu, and V. Parmeshwaran, "Residual life predictions from vibration-based degradation signals: a neural network approach," *IEEE Transactions on industrial electronics,* vol. 51, no. 3, pp. 694-700, 2004.

[29]   N. Z. Gebraeel and M. A. Lawley, "A neural network degradation model for computing and updating residual life distributions," *IEEE Transactions on Automation Science and Engineering,* vol. 5, no. 1, pp. 154-163, 2008.

[30]   Y.-L. Dong, Y.-J. Gu, K. Yang, and W.-K. Zhang, "A combining condition prediction model and its application in power plant," in *Machine Learning and Cybernetics, 2004. Proceedings of 2004 International Conference on*, 2004, vol. 6: IEEE, pp. 3474-3478.

[31]   P. Wang and G. Vachtsevanos, "Fault prognostics using dynamic wavelet neural networks," *AI EDAM,* vol. 15, no. 4, pp. 349-365, 2001.

[32]   L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomputing,* vol. 240, pp. 98-109, 2017.

[33]   K. L. Butler, "An expert system based framework for an incipient failure detection and predictive maintenance system," in *Intelligent Systems Applications to Power Systems, 1996. Proceedings, ISAP'96., International Conference on*, 1996: IEEE, pp. 321-326.

[34]   W. Q. Wang, M. F. Golnaraghi, and F. Ismail, "Prognosis of machine health condition using neuro-fuzzy systems," *Mechanical Systems and Signal Processing,* vol. 18, no. 4, pp. 813-831, 2004.

[35]   X. Guoxin, X. LiChuan, B. Meihua, and L. Siwei, "Fault prediction of boilers with fuzzy mathematics and RBF neural network," in *Communications, Circuits and Systems, 2005. Proceedings. 2005 International Conference on*, 2005, vol. 2: IEEE.

[36]   P. Baraldi, F. Di Maio, S. Al-Dahidi, E. Zio, and F. Mangili, "Prediction of industrial equipment Remaining Useful Life by fuzzy similarity and belief function theory," *Expert Systems with Applications,* vol. 83, pp. 226-241, 2017.

[37]   S.-T. Tseng, M. Hamada, and C.-H. Chiao, "Using degradation data to improve fluorescent lamp reliability," *Journal of Quality Technology,* vol. 27, no. 4, pp. 363-369, 1995.

[38]   C. J. Lu and W. O. Meeker, "Using degradation measures to estimate a time-to-failure distribution," *Technometrics,* vol. 35, no. 2, pp. 161-174, 1993.

[39]   R. B. Chinnam, "On‐line reliability estimation for individual components using statistical degradation signal models," *Quality and Reliability Engineering International,* vol. 18, no. 1, pp. 53-73, 2002.

[40]   S. Chakraborty, N. Gebraeel, M. Lawley, and H. Wan, "Residual-life estimation for components with non-symmetric priors," *IIE Transactions,* vol. 41, no. 4, pp. 372-387, 2009.

[41]   G. Whitmore, "Estimating degradation by a Wiener diffusion process subject to measurement error," *Lifetime data analysis,* vol. 1, no. 3, pp. 307-319, 1995.

[42]   S. Tang, C. Yu, X. Wang, X. Guo, and X. Si, "Remaining useful life prediction of lithium-ion batteries based on the wiener process with measurement error," *Energies,* vol. 7, no. 2, pp. 520-547, 2014.

[43]   Z.-S. Ye, Y. Wang, K.-L. Tsui, and M. Pecht, "Degradation data analysis using Wiener processes with measurement errors," *IEEE Transactions on Reliability,* vol. 62, no. 4, pp. 772-780, 2013.

[44]   C.-Y. Peng and S.-T. Tseng, "Mis-specification analysis of linear degradation models," *IEEE Transactions on Reliability,* vol. 58, no. 3, pp. 444-455, 2009.

[45]   X.-S. Si, W. Wang, C.-H. Hu, D.-H. Zhou, and M. G. Pecht, "Remaining useful life estimation based on a nonlinear diffusion degradation process," *IEEE Transactions on Reliability,* vol. 61, no. 1, pp. 50-67, 2012.

[46]   X. Wang, "Wiener processes with random effects for degradation data," *Journal of Multivariate Analysis,* vol. 101, no. 2, pp. 340-351, 2010.

[47]   X. Wang, P. Jiang, B. Guo, and Z. Cheng, "Real‐time reliability evaluation with a general Wiener process‐based degradation model," *Quality and Reliability Engineering International,* vol. 30, no. 2, pp. 205-220, 2014.

[48]   X.-S. Si, W. Wang, C.-H. Hu, M.-Y. Chen, and D.-H. Zhou, "A Wiener-process-based degradation model with a recursive filter algorithm for remaining useful life estimation," *Mechanical Systems and Signal Processing,* vol. 35, no. 1, pp. 219-237, 2013.

[49] Z.-S. Ye, Y. Shen, and M. Xie, "Degradation-based burn-in with preventive maintenance," *European journal of operational research,* vol. 221, no. 2, pp. 360-367, 2012.

[50] Z.-S. Ye and N. Chen, "The inverse Gaussian process as a degradation model," *Technometrics,* vol. 56, no. 3, pp. 302-311, 2014.

[51] J. Van Noortwijk, "A survey of the application of gamma processes in maintenance," *Reliability Engineering & System Safety,* vol. 94, no. 1, pp. 2-21, 2009.

[52] X. Wang and D. Xu, "An inverse Gaussian process model for degradation data," *Technometrics,* vol. 52, no. 2, pp. 188-197, 2010.

[53] E. Zio and A. Zoia, "Parameter identification in degradation modeling by reversible-jump Markov Chain Monte Carlo," *IEEE Transactions on Reliability,* vol. 58, no. 1, pp. 123-131, 2009.

[54] J. P. Kharoufeh, C. J. Solo, and M. Y. Ulukus, "Semi-Markov models for degradation-based reliability," *IIE Transactions,* vol. 42, no. 8, pp. 599-612, 2010.

[55] M. E. Cholette and D. Djurdjanovic, "Degradation modeling and monitoring of machines using operation-specific hidden Markov models," *IIE Transactions,* vol. 46, no. 10, pp. 1107-1123, 2014.

[56] M. Dong and D. He, "Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis," *European journal of operational research,* vol. 178, no. 3, pp. 858-878, 2007.

[57] M. Dong and D. He, "A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology," *Mechanical Systems and Signal Processing,* vol. 21, no. 5, pp. 2248-2266, 2007.

[58] S. J. Bae and P. H. Kvam, "A nonlinear random-coefficients model for degradation testing," *Technometrics,* vol. 46, no. 4, pp. 460-469, 2004.

[59] J. Son, Y. Zhang, C. Sankavaram, and S. Zhou, "RUL prediction for individual units based on condition monitoring signals with a change point," *IEEE Transactions on Reliability,* vol. 64, no. 1, pp. 182-196, 2015.

[60] J. Feng, Q. Sun, and T. Jin, "Storage life prediction for a high-performance capacitor using multi-phase Wiener degradation model," *Communications in Statistics-Simulation and Computation,* vol. 41, no. 8, pp. 1317-1335, 2012.

[61] T. S. Ng, "An application of the EM algorithm to degradation modeling," *IEEE Transactions on Reliability,* vol. 57, no. 1, pp. 2-13, 2008.

[62] X. Wang, P. Jiang, B. Guo, and Z. Cheng, "Real‑time Reliability Evaluation for an Individual Product Based on Change‑point Gamma and Wiener Process," *Quality and Reliability Engineering International,* vol. 30, no. 4, pp. 513-525, 2014.

[63] N. Chen and K. L. Tsui, "Condition monitoring and remaining useful life prediction using degradation signals: Revisited," *IIE Transactions,* vol. 45, no. 9, pp. 939-952, 2013.

[64] W. Wang, "A two-stage prognosis model in condition based maintenance," *European journal of operational research,* vol. 182, no. 3, pp. 1177-1187, 2007.

[65] C. Chiao and M. Hamada, "Using degradation data from an experiment to achieve robust reliability for light emitting diodes," *Quality and Reliability Engineering International,* vol. 12, no. 2, pp. 89-94, 1996.

[66] D. M. Hawkins, "Fitting multiple change-point models to data," *Computational Statistics & Data Analysis,* vol. 37, no. 3, pp. 323-341, 2001.

[67] P. Fearnhead and Z. Liu, "On‑line inference for multiple changepoint problems," *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* vol. 69, no. 4, pp. 589-605, 2007.

[68] A. Tomé and P. Miranda, "Piecewise linear fitting and trend changing points of climate parameters," *Geophysical Research Letters,* vol. 31, no. 2, 2004.

[69] C. Beaulieu, J. Chen, and J. L. Sarmiento, "Change-point analysis as a tool to detect abrupt climate variations," *Phil. Trans. R. Soc. A,* vol. 370, no. 1962, pp. 1228-1249, 2012.

[70] Y. Li and P. Nilkitsaranont, "Gas turbine performance prognostic for condition-based maintenance," *Applied energy,* vol. 86, no. 10, pp. 2152-2161, 2009.

[71] S. J. Bae and P. H. Kvam, "A change-point analysis for modeling incomplete burn-in for light displays," *IIE Transactions,* vol. 38, no. 6, pp. 489-498, 2006.

[72] S. J. Bae, T. Yuan, S. Ning, and W. Kuo, "A Bayesian approach to modeling two-phase degradation using change-point regression," *Reliability Engineering & System Safety,* vol. 134, pp. 66-74, 2015.

[73] D. Wang and K.-L. Tsui, "Two novel mixed effects models for prognostics of rolling element bearings," *Mechanical Systems and Signal Processing,* vol. 99, pp. 1-13, 2018.

[74] X.-S. Si, C.-H. Hu, Q. Zhang, and T. Li, "An Integrated Reliability Estimation Approach With Stochastic Filtering and Degradation Modeling for Phased-Mission Systems," *IEEE transactions on cybernetics,* vol. 47, no. 1, pp. 67-80, 2017.

[75] X.-S. Si, C.-H. Hu, X. Kong, and D.-H. Zhou, "A residual storage life prediction approach for systems with operation state switches," *IEEE Transactions on industrial electronics,* vol. 61, no. 11, pp. 6304-6315, 2014.

[76] N. Whiteley, C. Andrieu, and A. Doucet, "Particle Markov Chain Monte Carlo for multiple change-point problems," Technical Report, 2009.

[77] A. Hannart and P. Naveau, "An improved Bayesian information criterion for multiple change-point models," *Technometrics,* vol. 54, no. 3, pp. 256-268, 2012.

[78] J. Wu, Y. Chen, and S. Zhou, "Online detection of steady-state operation using a multiple-change-point model and exact Bayesian inference," *IIE Transactions,* vol. 48, no. 7, pp. 599-613, 2016.

[79] F. Caron, A. Doucet, and R. Gottardo, "On-line changepoint detection and parameter estimation with application to genomic data," *Statistics and computing,* vol. 22, no. 2, pp. 579-595, 2012.

[80] N. A. Heard and M. J. Turcotte, "Adaptive sequential Monte Carlo for multiple changepoint analysis," *Journal of Computational and Graphical Statistics,* vol. 26, no. 2, pp. 414-423, 2017.

[81] P. Fearnhead, "Exact and efficient Bayesian inference for multiple changepoint problems," *Statistics and computing,* vol. 16, no. 2, pp. 203-213, 2006.

[82] J. Wu, Y. Chen, S. Zhou, and X. Li, "Online steady-state detection for process control using multiple change-point models and particle filters," *IEEE Transactions on Automation Science and Engineering,* vol. 13, no. 2, pp. 688-700, 2016.

[83] Y. Hou, J. Wu, and Y. Chen, "Online Steady State Detection Based on Rao‑Blackwellized Sequential Monte Carlo," *Quality and Reliability Engineering International,* vol. 32, no. 8, pp. 2667-2683, 2016.

[84] G. Schwarz, "Estimating the dimension of a model," *The annals of statistics,* vol. 6, no. 2, pp. 461-464, 1978.

[85]   R. Killick, P. Fearnhead, and I. A. Eckley, "Optimal detection of changepoints with a linear computational cost," *Journal of the American Statistical Association,* vol. 107, no. 500, pp. 1590-1598, 2012.

[86]   B. E. Olivares, M. A. C. Munoz, M. E. Orchard, and J. F. Silva, "Particle-filtering-based prognosis framework for energy storage devices with a statistical characterization of state-of-health regeneration phenomena," *IEEE Transactions on Instrumentation and Measurement,* vol. 62, no. 2, pp. 364-376, 2013.

[87]   E. Zio and G. Peloni, "Particle filtering prognostic estimation of the remaining useful life of nonlinear components," *Reliability Engineering & System Safety,* vol. 96, no. 3, pp. 403-409, 2011.

[88]   Q. Miao, L. Xie, H. Cui, W. Liang, and M. Pecht, "Remaining useful life prediction of lithium-ion battery with unscented particle filter technique," *Microelectronics Reliability,* vol. 53, no. 6, pp. 805-810, 2013.

[89]   N. Li, Y. Lei, J. Lin, and S. X. Ding, "An improved exponential model for predicting remaining useful life of rolling element bearings," *IEEE Transactions on industrial electronics,* vol. 62, no. 12, pp. 7762-7773, 2015.

[90]   Y. Hu, P. Baraldi, F. Di Maio, and E. Zio, "A particle filtering and kernel smoothing-based approach for new design component prognostics," *Reliability Engineering & System Safety,* vol. 134, pp. 19-31, 2015.

[91]   A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and computing,* vol. 10, no. 3, pp. 197-208, 2000.

[92]   W. R. Gilks and C. Berzuini, "Following a moving target—Monte Carlo inference for dynamic Bayesian models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* vol. 63, no. 1, pp. 127-146, 2001.

[93]   N. Chopin, "Dynamic detection of change points in long time series," *Annals of the Institute of Statistical Mathematics,* vol. 59, no. 2, pp. 349-366, 2007.

[94]   N. Gebraeel, "Sensory-updated residual life distributions for components with exponential degradation patterns," *IEEE Transactions on Automation Science and Engineering,* vol. 3, no. 4, pp. 382-393, 2006.

[95]   J. Son, S. Zhou, C. Sankavaram, X. Du, and Y. Zhang, "Remaining useful life prediction based on noisy condition monitoring signals using constrained Kalman filter," *Reliability Engineering & System Safety,* vol. 152, pp. 38-50, 2016.

[96]   A. Saxena *et al.*, "Metrics for evaluating performance of prognostic techniques," presented at the 2008 International Conference on Prognostics and Health Management, Denver, CO, USA, 2008.

[97]   Y. Wen, J. Wu, Q. Zhou, and T.-L. Tseng, "Multiple-Change-Point Modeling and Exact Bayesian Inference of Degradation Signal for Prognostic Improvement," *IEEE Transactions on Automation Science and Engineering,* no. 99, pp. 1-16, 2018.

[98]   Z. Pan and N. Balakrishnan, "Reliability modeling of degradation of products with multiple performance characteristics based on gamma processes," *Reliability Engineering & System Safety,* vol. 96, no. 8, pp. 949-957, 2011.

[99]   Z. Ye, N. Chen, and K.-L. Tsui, "A Bayesian approach to condition monitoring with imperfect inspections," *Quality and Reliability Engineering International* vol. 31, no. 3, pp. 513-522, 2015.

[100] Y. Wen, J. Wu, and Y. Yuan, "Multiple-Phase Modeling of Degradation Signal for Condition Monitoring and Remaining Useful Life Prediction," *IEEE Transactions on Reliability,* vol. 66, no. 3, pp. 924-938, 2017.

[101] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in Selected Papers of Hirotugu Akaike: Springer, 1998, pp. 199-213.

Vita

Yuxin Wen was born in Nanchong, China. She received the B.S. degree in medical information engineering from Sichuan University, Chengdu, China, in 2011, and the M.S. degree in biomedical engineering from Zhejiang University, Hangzhou, China, in 2014. After that, she went to industry as a software Engineer for two years. She came to U.S. in January 2016 to pursue her PhD degree from UTEP Electrical and Computer Engineering program. During the Ph.D. study, Wen actively published five peer reviewed journal papers as listed below.

1. Y. Gao, **Y. Wen**, and J. Wu. "A Neural Network-Based Joint Prognostic Model for Data Fusion and Remaining Useful Life Prediction." IEEE Transactions on Neural Networks and Learning Systems (2020), Early Access.
2. **Y. Wen**, D. AlHakeem, P. Mandal, S. Chakraborty, Y-K Wu, T. Senjyu, S. Paudyal, and B. Tseng, "Performance Evaluation of Probabilistic Methods Based on Bootstrap and Quantile Regression to Quantify PV Power Point Forecast Uncertainty," IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 4, pp. 1134-1144, April 2020.
3. **Y. Wen**, J. Wu, Q. Zhou and B. Tseng, "Multiple-Change-Point Modeling and Exact Bayesian Inference of Degradation Signal for Prognostic Improvement," IEEE Transactions on Automation Science and Engineering, vol. 16, no. 2, pp. 613-628, April 2019.
4. **Y. Wen**, J. Wu, D. Das, and B. Tseng, "Degradation modeling and RUL prediction using Wiener process subject to multiple change points and unit heterogeneity," Reliability Engineering & System Safety, vol. 176, Aug. 2018.
5. **Y. Wen**, J. Wu and Y. Yuan, "Multiple-Phase Modeling of Degradation Signal for Condition Monitoring and Remaining Useful Life Prediction," IEEE Transactions on Reliability, vol. 66, no. 3, pp. 924-938, Sept. 2017.

Contact Information: ywen@miners.utep.edu