Data-driven Approaches for Condition Monitoring and Predictive Analytics

By
Abdallah A. Chehade

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy
(Industrial Engineering)

at the
UNIVERSITY OF WISCONSIN-MADISON
2017

Date of final oral examination: 04/21/2017

The dissertation is approved by the following members of the Final Oral Committee:

Kaibo Liu, Assistant Professor, Industrial and Systems Engineering
Shiyu Zhou, Professor, Industrial and Systems Engineering
Dharmaraj Veeramani, Professor, Industrial and Systems Engineering
Xin Wang, Assistant Professor, Industrial and Systems Engineering
Linsey Steege, Assistant Professor, School of Nursing

© Copyright by Abdallah A. Chehade 2017
All Rights Reserved

To my beloved parents and siblings.

Acknowledgments

First, I would like to thank my advisor and mentor through my Ph.D. program, Professor Kaibo Liu, for devoting his invaluable time to guide me in the right direction, and his continuous support, help, and encouragement throughout the past 3 years. I have continuously kept learning from his experience and insightful suggestions, which have made a significant impact in my research and better shaped my personal life. His kindness and patience have made my Ph.D. journey a great and memorable experience, and for that I am thankful.

My gratitude also goes to my Master's academic advisor, Professor Neil Duffie, for sharing his invaluable experience and devoting his time mentoring me. His directed guidance, suggestions and comments have made me a better researcher and encouraged me to pursue my Ph.D. study.

I would also like to thank my thesis committee members, including Professor Shiyu Zhou, Professor Raj Veeramani, Professor Xin Wang and Professor Linsey Steege. Their valuable advice, productive discussions, and support have greatly aided me during my Ph.D. study.

My gratitude also goes for Professor Raj Veeramani, Professor Shiyu Zhou, Professor Jeffrey Linderoth and Professor Xi Zhang for their valuable recommendations throughout my job search.

My Ph.D. study would have been incomplete without my colleagues and friends: Ms. Areen Alsaid, Ms. Xiaochen Xian, Mr. Changyue Song, Mr. Scott Bonk and everyone who have supported me with productive and constructive comments.

Lastly, special thanks to my parents, Mr. Adnan Chehade and Ms. Raife Mansour, for their hard work, continuous encouragement, and unstoppable support. Also, special thanks to my siblings, Wafaa Chehade, Hassan Chehade and Ahmad Chehade, for being by my side for my entire life.

The author is greatly grateful to the supports in parts by the National Science Foundation Grant CMMI-1435809 and the Office of Naval Research grant N00014-17-1-2261

Table of Contents

List of Tables

List of Figures

Abstract

The rapid development of sensing and communication technologies has enabled an unprecedented opportunity for condition monitoring, making multiple data streams a commonplace to simultaneously monitor the health status of an operating unit. Such a big data environment poses essential challenges in determining (i) which data streams to use; and (ii) how to fuse/combine those multiple and relevant data streams for better failure diagnosis and prognostics as these multiple data streams are often correlated and each data stream may only contain partial information about the degraded unit. However, it is often hard to physically interpret the dependencies and relations between these data streams due to the complexity of the system. Given the massive amount of data have become available, nowadays many research companies are looking for effective tools to improve failure monitoring and predictive capabilities. As a consequence, my research focuses on developing effective data-driven methodologies to better monitor and infer the condition of an operating unit in real time. Such inference would be very useful for profitable managerial decision-making such as condition-based maintenance scheduling, work in progress distribution, shipment scheduling, and customer satisfaction.

This thesis contributes to the field of System Informatics and Data Analytics (SIDA) by developing systematic data-driven methodologies for better condition monitoring and prognostic analysis in complex systems. These developed methodologies enable (i) real time modeling and characterization of the health status of a system, (ii) predicting future measurements, trends and behaviors of the system, and (iii) further diagnosing the reasons for degradation and failure of the system. This research combines advanced statistical methods, data analytics tools, engineering knowledge, and decision science and operations research. The research is highly applicable in many applications such as health care, manufacturing, after sales and services.

In the third Chapter of this thesis, we first investigate a fundamental question that is how to measure the signal quality of a degradation signal. If such a question can be addressed, then the data fusion approach can be simplified as a mission-specific task: to construct a composite health index with the goal of optimizing its signal quality. In particular, a new signal-to-noise ratio (SNR) metric that is tailored to the needs of degradation signals is proposed. Then, based on the new quality metric, we develop a data-level fusion model to construct a health index via fusion of multiple degradation-based sensor data. Our goal is that the developed health index provides a much better characterization of the health condition of the unit and thus leads to a better prediction of the remaining lifetime. A case study that involves the degradation dataset of aircraft gas turbine engines is conducted to numerically evaluate the performance of the developed health index regarding prognostics and further compare the result with existing literature.

In the fourth Chapter of this thesis, we investigate one challenging question that is how to leverage the multiple degradation-based sensor data for better online degradation modeling and prognostic analysis during condition monitoring when there are multiple failure modes. In particular, we propose a data-level fusion methodology to construct a composite failure mode index, which is named as FM-INDEX via the fusion of multiple sensor information. Our goal is to utilize the FM-INDEX to better characterize the degradation status and failure mode of an operating unit in real time, thus leading to better degradation modeling and prognostic analysis. A case study that involves a degradation dataset of an aircraft gas turbine engine with two potential failure modes is implemented to numerically evaluate the performance of our proposed method and compare it with the existing literature.

In the fifth Chapter of this thesis, we develop a convex quadratic formulation that combines the information from the degradation profiles of historical units and the *in-situ* sensory data from an

operating unit to online estimate the failure threshold distribution of this particular unit in the field. With a more accurate estimation of the failure threshold of the operating unit in real time, then a better remaining useful life prediction is expected to be achieved. Simulation as well as a case study involving a degradation dataset of aircraft turbine engines were used to numerically evaluate and compare the performance of the proposed methodology with the existing literature in the context of failure threshold estimation and remaining useful life prediction.

In the sixth Chapter of this thesis, we propose a Structural Degradation Modeling (SDM) framework for sparse datasets as a recommender system taking into consideration of (i) the available data from the unit of interest; (ii) the relationship between the recommenders; (iii) the accuracy of the leveraged recommenders; and (iv) the population characteristics. The developed framework is tested and validated by simulation studies as well as the ADNI dataset and the results showed that the SDM framework outperforms the benchmark methods for degradation modeling.

In summary, this thesis contributes to data-driven predictive analytics for effective condition monitoring, diagnosis, and prognostics. The developed methods can be applied to various applications, which will lead to an improved maintenance scheduling, resource planning, work in progress allocation, and logistics.

Chapter 1

Introduction

1.1 Motivation and Overview

> "Prognostics and Health Management is a system discipline focusing on detection, prediction, and management of the health and status of complex engineered systems." – the first International Conference on PHM, 2008.

Nowadays, engineering and service systems (e.g., machines, products, companies, hospitals, and medical treatments) are becoming more and more complex in general. Such complexity poses many interesting and challenging questions including but limited to predicting the remaining useful life, conducting fault diagnosis in real time, scheduling maintenance, and designing efficient production planning strategies. Those questions have attracted much attention recently as solving these questions may provide direct insights for better managerial decision-making (e.g., condition-based maintenance, personalized medicine and treatment), leading to improved efficiency, quality and cost for the overall system. With the rapid development of sensing and communication technology, engineering and service systems are built with many embedded sensors to better understand the health status of the system. This availability of data-rich environments provides a unique and unprecedented opportunity to better monitor, diagnose and forecast the behavior of an

operating unit in real time. This thesis document aims to achieve this goal by developing advanced data fusion models and at the same time addressing the potential big data challenges resulted from such data-rich environments.

1.2 Condition Monitoring

Unexpected failures in systems often result in production downtime and delayed schedules, which may lead to severe economic losses, customer dissatisfaction, and safety issues. As a result, it is critically important to accurately measure the health status and prevent the unexpected failure of a unit (e.g., machine, tool, equipment) in real time. To achieve this goal, condition monitoring techniques have been widely used, which aim to fully understand and track the degradation status of a unit along its entire service lifecycle. The collected sensor measurements during condition monitoring are known as degradation signals, which characterize the degradation evolution of a unit. With such information available, we can then produce prognostic inferences [1], [2], such as remaining useful life (RUL) prediction. In this way, a condition-based maintenance strategy [3], [4] can be implemented, which will significantly improve production efficiency, reduce inventory and operational costs, and enhance customer loyalty.

While the rapid development of sensing and communication technologies has enabled an unprecedented opportunity for condition monitoring, making multiple sensors a commonplace to simultaneously monitor the health status of an operating unit, such a big data environment also poses essential challenges in determining (i) which sensors to use; and (ii) how to fuse/combine those multiple and relevant sensor data for better degradation modeling and prognostics. The existing research, however, mainly focuses on analyzing single sensor signals for modeling the

degradation process and performing prognostic analysis. These approaches rely on the assumption that the single sensor is able to fully characterize the underlying physical transition of a degraded unit. Unfortunately, such a simplified assumption may not be valid in many real-world applications, thus often leading to inaccurate or unreliable prognostic results [5], [6].

1.3 Data Fusion

Since multiple sensor data often contain correlated and partial information about the same unit, data fusion methodologies have been widely used to provide more accurate and robust inferences about degradation status and remaining lifetime of a unit. Generally speaking, data fusion methods can be fitted into two main categories [7], depending on which level the fusion technique is implemented: data-level fusion and decision-level fusion. An overview of multiple sensor data fusion approaches to condition monitoring, fault diagnosis, and prognostics can be found in [8]. Specifically, the data-level fusion combines multiple sensor data or extracted features into a one-dimensional health index that can be further used for decision making (e.g., degradation modeling and prognostics), whereas the decision-level fusion involves the integration of prediction results from separate analysis of each individual sensor data [9]. While the decision-level fusion is straightforward to implement, it ignores the dependence of multiple sensor data and requires repeated computations based on each individual sensor data. Thus, such an approach often leads to biased results, and its application to big data is quite limited.

Unlike the decision-level fusion, data-level fusion has shown promise as an effective solution and attracted more attention recently [10], [11]. However, one limitation of the existing data-level fusion methods for degradation modeling and prognostics is that they often assume there is only

one potential failure mode or ignore the effect of multiple failure modes on the degradation profile of a unit. Unfortunately, as shown in many real-world applications, different failure modes do have distinct influences on the service lifecycle path of a unit. Thus, an essential challenge here is how to accurately identify the failure mode and predict the remaining lifetime along the lifecycle of a unit in real time. To achieve this goal, several techniques, such as Hidden Markov Models [12], Artificial Intelligence [13], and Physical-based modeling approaches [14], have been developed in the literature to online diagnose the failure mode during condition monitoring; however, they mainly focus on analyzing a single sensor data. While some studies have considered the failure mode diagnosis based on multiple sensor data, they either employ a simple voting scheme that combines the results from separate analysis of each individual sensor data [15]–[17], or fail to address the specific needs of degradation modeling and prognostics during condition monitoring [18]–[24]. In particular, there are two fundamental requirements that need be satisfied: (i) the failure mode diagnosis needs to be conducted and updated continuously along the lifecycle of a unit, and (ii) the fault diagnostic result is desired to become more accurate as the unit approaches to the end of its life to ensure good maintenance planning and avoid unexpected failure.

1.4 Remaining Useful Life and Degradation Modeling

In the literature of prognostics, it is often assumed that a unit fails once its degradation signal hits a predefined failure threshold. As a result, to achieve an appropriate estimation on the RUL distribution, there are two essential requirements that must be satisfied: (i) an accurate estimation of the failure threshold; and (ii) a reliable degradation model that characterizes the degradation

profile of a unit, such that the status of the unit can be accurately predicted at future observation times.

Many research efforts have been made focusing on requirement (ii) [25]–[28]. For example, Lu and Meeker introduced the mixed-effects model to characterize both the common effect shared by the population and the unique stochastic nature of each degraded unit [25]. The model was further utilized by others (e.g., [29]) such that once the *in-situ* sensory data is collected from an operating unit, then we can leverage the Bayesian updating approaches to online calibrate the distribution of the random effects for this particular unit in real time.

1.5 Failure Threshold Estimation

While extensive work has been done in developing the degradation models, the current literature still falls short in studying the requirement (i), i.e., how to accurately estimate the failure threshold of a unit in real time. In most cases, such a fixed threshold is estimated based on the physical knowledge of the system/process. However, many recent studies have shown that using a fixed failure threshold is problematic [10], [11], [30], [31]. For instance, as explicitly discussed in Wang and Coit [31], "a probabilistic, rather than a deterministic threshold value is more appropriate." To address this issue, Liu *et al.* [10] considered the statistics (mean and variance) of the last sensor observations (right before failure) of multiple historical units as an estimation of the failure threshold. While the existing data-driven approaches consider the uncertainty in the failure threshold, such estimators are derived based on the population information of historical units, which may not fully capture the unique characteristics of each unit. As shown in many applications [32], the last observations of units are quite different. This is mainly because the

degradation is inherently a stochastic process that is dependent on many unknown and variable factors. In addition, units may be vulnerable to multiple potential failure modes [33] and each failure mode results in a different failure threshold. In these cases, the offline estimation of the failure threshold distribution using population-wide characteristics [10] may not be able to fully capture the unique properties of individual units. On the other hand, while the existing literature has considered leveraging the *in-situ* sensory data to online update the degradation model, such an approach is not easily extendable to estimate the failure threshold.

1.6 Thesis Layout

The rest of this thesis document is structured as follows. In Chapter 2, we review and summarize the existing literature on degradation modeling, RUL estimations, failure threshold estimations and data fusion approaches in presence of multiple sensors for prognostic analysis.

In Chapter 3, we propose a metric to quantify the degradation signal quality and then develop a data-level fusion model that combines multiple sensor signals to construct a composite health index by maximizing the proposed quality metric. A case study is introduced to evaluate the performance of the proposed method.

In Chapter 4, we introduce a data-level fusion approach for degradation modeling and prognostic analysis in presence of multiple failure modes. A case study with multiple failure modes is conducted to evaluate the performance of the proposed approach.

In Chapter 5, we introduce a data-driven approach to online estimate the failure threshold for an operating unit. The method first measures the similarity between the operating unit and the historical units, and then infers the failure threshold of the operating unit from the historical units

based on the derived similarity measure. A simulation study as well as a case study are introduced to evaluate the performance of the proposed method.

In Chapter 6, we propose a Structural Degradation Modeling (SDM) framework for sparse datasets as a recommender system taking into consideration of (i) the available data from the unit of interest; (ii) the relationship between the recommenders; (iii) the accuracy of the leveraged recommenders; and (iv) the population characteristics. The developed framework is tested and validated by simulation studies as well as the ADNI dataset and the results showed that the SDM framework outperforms the benchmark methods for degradation modeling.

In Chapter 7, a summary of the original contributions of this thesis are provided and potential follow-up works to the proposed methods in Chapter 3, 4, 5 and 6 are also discussed.

Chapter 2

Literature Review

Many research efforts have been devoted to condition monitoring, prognostics. and degradation modeling. The main objectives behind those efforts are to extend the equipment lifetime, reduce down time, keep throughput and due dates on track, and minimize maintenance costs. With the new developments in sensor technology and computation capabilities, there is a pressing need to develop advanced data-driven methodologies to better infer and predict the future about the status of an operating unit in real time. This section presents a review on some of the commonly used degradation modeling techniques, RUL estimation techniques, failure threshold estimation techniques, and data fusion techniques for and degradation modeling prognostic analysis in presence of multiple sensor data.

Table 2.1. Scenario for units with partial degradation history

Time	**Unit 1**	**Unit 2**	**Unit 3**	**Unit 4**	**Unit 5**
t_1	$s_{1,1}$	$s_{2,1}$	$s_{3,1}$	$y_{4,1}$	$s_{5,1}$
t_2	$s_{1,2}$	$s_{2,2}$	$s_{3,2}$	$y_{4,2}$	$s_{5,2}$
t_3	N/A	$s_{2,3}$	$s_{3,3}$	N/A	$s_{5,3}$
t_4	$s_{1,4}$	N/A	$s_{3,4}$	N/A	$s_{5,4}$
t_5	$s_{1,5}$	$s_{2,5}$	$s_{3,5}$	N/A	$s_{5,5}$
t_k	N/A	N/A	N/A	N/A	$s_{5,k}$
t_{end}	$s_{1,end}$	$s_{2,end}$	N/A	N/A	$s_{5,end}$

For better illustration, we show one challenging task for degradation modeling in Table 2.1. The table summarizes the degradation statuses of multiple units at different times, e.g., the degradation status of unit i at time t_k is denoted by s_{i,t_k}. Here, we define t_{end} to be the last time

point where there are degradation records about any of the units. One challenging task is to accurately estimate the missing values and to precisely extrapolate the degradation status for each unit.

2.1 Degradation Modeling

To achieve this goal, degradation modeling is one of the critical tasks. Many research efforts have been devoted to accurate degradation modeling. Specifically, in this thesis, we focus on three main streams of degradation models: Weiner process models, Mixed-effects models and Sparse modeling approaches. There exists many other approaches to degradation modeling as well, such as autoregressive models [34], [35], physics-based models [36], [37], and data-driven models (neural networks, support vector machines, etc.) [10], [29]; however, they will not be discussed into details in the following.

2.1.1 Weiner Process Model

In a wide set of applications, the degradation is accumulative; in other words, the more damage the system receives, the higher the rate of degradation is expected. A commonly used degradation model that captures this phenomenon is the Wiener process with positive drifts. For example, the Wiener process model was used by (i) Le Son et al. [38] to model the 2008 PHM conference data; (ii) Whitmore and Schenkelberg [39] to model the resistance increase in a self-regulating heating cable; and (iii) Hu et al. [40] to model an LED dataset. A popular representation of the Wiener process in degradation analysis is:

$$y(t) = \nu G(t) + \sigma B\big(G(t)\big) + \varepsilon(t), \tag{2.1}$$

where ν is the rate of degradation and often called drift parameter; σ represents the volatile parameter; $B(.)$ follows a Brownian motion; $G(.)$ is a parametric function that reflects the expected evolution in the degradation path; and $\varepsilon(t)$ is the measurement error or noise and often assumed to follow a normal distribution with mean 0 and standard deviation σ_ε.

Specifically, the main assumption in the Weiner process models is that it has independent and normally distributed increments such that:

$$\Delta y(t) = y(t+\Delta t) - y(t) \sim N(\nu G(t+\Delta t) - \nu G(t), \sigma^2 G(t+\Delta t) - \sigma^2 G(t) + 2\sigma_\varepsilon^2).$$

There exists other types of the Weiner processes tailored to degradation modeling [41]–[45], and these models also inspired the utilization of other stochastic processes such as the Gaussian Process and Inverse Gaussian Process [46]–[53].

2.1.2 Mixed-effects Model

For degradation modeling, the mixed-effects model (MEM) is one of the most commonly used technique in the literature, which considers both the population and individual characteristics [10], [29], [54]–[56]. It was first introduced by Lu and Meeker [25] and can be written as the following:

$$s_t = \eta(\boldsymbol{\varphi}, \boldsymbol{\nu}, t) + \epsilon_t, \tag{2.2}$$

where $\eta(\cdot)$ is the parametric form of the degradation model; s_t is the measurement for describing the underlying degradation status at time t for the unit of interest; $\boldsymbol{\varphi}$ is a vector of fixed-effect parameters that represents common characteristics of the population; $\boldsymbol{\nu}$ is a vector of random-effect parameters that characterizes the unit-to-unit variability; and ϵ_t is an error term that represents the measurement noises. Depending on the parametric form of $\eta(\cdot)$, this degradation model can be used to describe a variety of functional forms according to the evolution of the

degradation signal. Based on the MEM, extensive studies have been developed in the literature. For example, Gebraeel *et al.* (2005) proposed a Bayesian version of the MEM (BMEM) that first fitted a MEM given a set of historical units, and then utilized the Bayesian approaches for updating the random-effect parameters given the degradation observations of the unit of interest. This approach accounts for (i) the population characteristics via the MEM; and (ii) the individual characteristics via the Bayesian update. Without loss of generality, below we focus on the *p*th order polynomial degradation model as considered in Liu *et al.* (2013) for a demonstration:

$$s_t = \sum_{\alpha=0}^{p} (\nu_\alpha) t^\alpha + \epsilon_t = \boldsymbol{\Gamma}_t \boldsymbol{\nu} + \epsilon_t, \tag{2.3}$$

where p is the order of the polynomial model; $\boldsymbol{\nu} = [\nu_0, \cdots, \nu_p]'$ is the random-effect parameters and often assumed to follow a multivariate normal distribution, $\boldsymbol{\nu} \sim N_{p+1}(\boldsymbol{u}^0, \boldsymbol{\Sigma}^0)$; ϵ_t is the random noise and follows $N(0, \sigma^2)$; and $\boldsymbol{\Gamma}_t = [1, t, \ldots, t^p]$.

Then, the estimated measurement for the degradation status at time t can be calculated as:

$$\hat{s}_t | \boldsymbol{s}_{\cdot} = \boldsymbol{\Gamma}_t * (\boldsymbol{\nu} | \boldsymbol{s}_{\cdot}), \tag{2.4}$$

where $\boldsymbol{s}_{\cdot} = \left[s_{t(1)}, \ldots, s_{t(n)}\right]'$ is the vector of the observed measurements for the unit of interest up to current time $t(n)$; and $\boldsymbol{\nu} | \boldsymbol{s}_{\cdot}$ is the posterior of the random-effect parameters and follows $N_{p+1}(\boldsymbol{u}^1, \boldsymbol{\Sigma}^1)$, such that $\boldsymbol{u}^1 = \left(\frac{\boldsymbol{\Psi}^T \boldsymbol{\Psi}}{\sigma^2} + (\boldsymbol{\Sigma}^0)^{-1}\right)^{-1} \left(\frac{\boldsymbol{\Psi}^T \boldsymbol{s}_{\cdot}}{\sigma^2} + (\boldsymbol{\Sigma}^0)^{-1} \boldsymbol{u}^0\right)$, $\boldsymbol{\Sigma}^1 = \left(\frac{\boldsymbol{\Psi}^T \boldsymbol{\Psi}}{\sigma^2} + (\boldsymbol{\Sigma}^0)^{-1}\right)^{-1}$,

$\boldsymbol{\Psi} \in R^{n \times (p+1)} = \begin{bmatrix} 1 & \ldots & t(1)^p \\ \ldots & \ldots & \ldots \\ 1 & \ldots & t(k)^p \\ \ldots & \ldots & \ldots \\ 1 & \ldots & t(n)^p \end{bmatrix}$, and n is the number of available observations up to current time $t(n)$.

Accordingly, the prediction of the degradation status at time t is just a realization of the distribution $\hat{s}_t|\boldsymbol{s}_{\cdot} \sim N(\boldsymbol{\Gamma}_t \boldsymbol{u}^1, \boldsymbol{\Gamma}_t \boldsymbol{\Sigma}^1 \boldsymbol{\Gamma}_t^T)$. While the BMEM shows promising results in rich datasets, the Bayesian update procedure may not be effective for units of interest with a limited degradation history. This is because, for units with a limited number of observations, the updating procedure tends to focus more on the population characteristics which may not be effective for future predictions of the degradation status for these units. In addition to that, this method does not assist in fault diagnosis. In this thesis, (i) we integrate, in Chapter 4, this method with machine learning techniques to simultaneously conduct fault diagnosis and prognostics analysis, which is expected to provide better overall prognostics performance; and (ii) we introduce, in Chapter 6, a framework for degradation modeling in sparse datasets to better predict the degradation status of units with limited observations.

2.1.3 Sparse Modeling Approaches

Recently, there are a few approaches for degradation modeling that attempt to address the challenges of sparse data environments. For example, Lin *et al*. (2015) assumed that each individual degradation model can be written as a weighted combination of K canonical models, where the weights and the K canonical models are learned simultaneously by an iterative algorithm. However, there are several limitations of the proposed method. First, this iterative algorithm only leads to a stationary point, which may not be globally optimal. Second, how to choose the initial starting point to reach a solution near the optimal parameters is not discussed either, which makes the entire proposed approach difficult to be implemented in practice. Third, the approach cannot be easily employed for monitoring of the units of interest in real time due to the high computational

cost of the iterative algorithm involved in learning the weights as well as the K canonical models. There also exists other approaches that utilize stochastic processes such as the Gaussian Process [47] and the Inverse Gaussian Process [46] to analyze the degradation processes with sparse data (e.g., large inter-arrival times between the degradation observations). In particular, Peng *et al.* (2017) proposed a Bayesian framework with Inverse Gaussian process models to analyze the degradation of heavy-duty machine tool's spindle systems where the position accuracy is measured at intermittent discrete time points. Although such methods showed good performance to continuously characterize and interpolate the degradation status of a unit over the observed time domain, they are ineffective for extrapolation and predictive analytics as mentioned in [57] because they (i) mainly focus on modeling the relationship between the available observations within each individual unit; and (ii) fail to capture the functional form of the model, which is critical for extrapolation.

2.2 Recommender System

The idea of the recommender system stems from application such movie recommendations, advertisement recommendations, grocery recommendations for users to increase the customer satisfaction, quality and speed of delivery, and to increase the revenues of the implementer. Although the problem seems different from degradation modeling, they are similar in the way that the recommender systems need to estimate what the quality of an item for the user in recommender system is, while the degradation modeling needs to predict what the quality for the operating unit at time t is. One main difference between the two setups, however, is that in degradation modeling, we often assume a monotonic trend but that is not necessary for recommender systems. Another

difference is that many recommender systems focus on discrete analysis because there usually exists a finite number of items, but degradation modeling focuses on continuous time analysis. Later, in this thesis, we try to utilize the collaborative capability of recommender systems to better model the degradation status of a unit in environments that contain huge amounts of units but limited information per unit.

There exists different frameworks for the recommender systems, including Matrix Completion [58], Collaborative Filtering [59], Unit-oriented Neighborhood [60], etc. However, (i) most of those frameworks are oriented to address the relationship between the user and a finite set of discrete items, which cannot be leveraged for continuous time analysis. Also, (ii) the existing frameworks are often application specific and assume prior correlated characteristics between different items such as movies that belong to the same genre, and advertisements that belong to the category. In this thesis, we focus on the Unit-oriented Neighborhood and its extension for continuous time analysis.

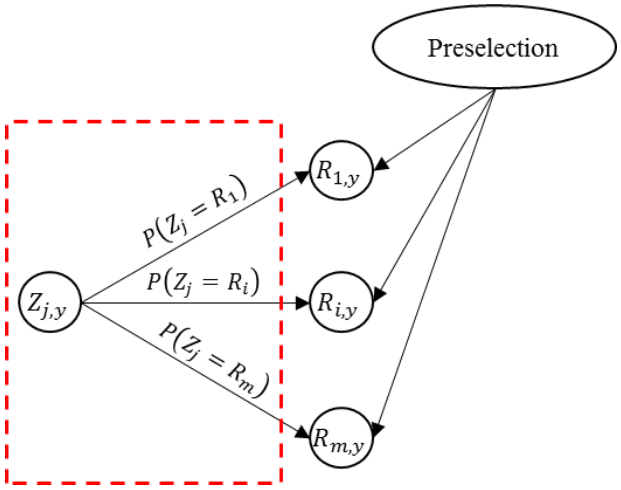

Fig. 2.1. The Conventional Recommender System Framework

Fig. 2.1 summarizes the conventional recommender system framework, which consists of learning the attitude of user j for item y from other recommenders who previously rated item y. For example, the authors in [60] implemented the recommender system to provide better movie recommendations to the users. The dashed box contains the following unknown values that are calculated upon request: (i) the estimated rating for item y by user j; and (ii) the probability that recommender i shares the same attitude as user j which is denoted by $P(Z_j = R_i)$.

Specifically, the attitude of user j for item y, $Z_{j,y}$, can be estimated as the following:

$$Z_{j,y} = \sum_{i=1}^{m} P(Z_j = R_i) * R_{i,y} + \epsilon_{j,y}, \tag{2.5}$$

where $R_{i,y}$ is the rating of recommender i for item y; $\epsilon_{j,y}$ represents the error term and it is usually assumed to follow $N(0, \sigma_j^2)$, where σ_j^2 can be estimated based on the historical recommendations from user j; and m is the number of available recommenders for item y.

Different deterministic approaches have been proposed in the literature to estimate $P(Z_j = R_i)$ such as the constrained least squares estimator that models the user as a weighted average of the recommenders and aims to minimize $\sum_{y \in \boldsymbol{Y}_{hist}^j} \left(\sum_{i=1}^{m} P(Z_j = R_i) * R_{i,y} - Z_{j,y}\right)^2$ given that $P(Z_j = R_i) \geq 0$, $\sum_{i=1}^{m} P(Z_j = R_i) = 1$, and $\boldsymbol{Y}_{hist}^j$ is the list of items that were previously rated by user j. However, these approaches often require that all the suggested recommenders for item y have rated enough common items in the past with respect to user j, in order to achieve an accurate estimation of $P(Z_j = R_i)$. Furthermore, as mentioned in the review by Adomavicius and Tuzhilin (2005), many approaches approximate $P(Z_j = R_i)$ based on the similarity between user j and recommender i.

$$P(Z_j = R_i) \propto sim(i,j), \tag{2.6}$$

where $sim(i,j)$ measures the similarity between user j and recommender i. For more details about the similarity metric, refer to [61].

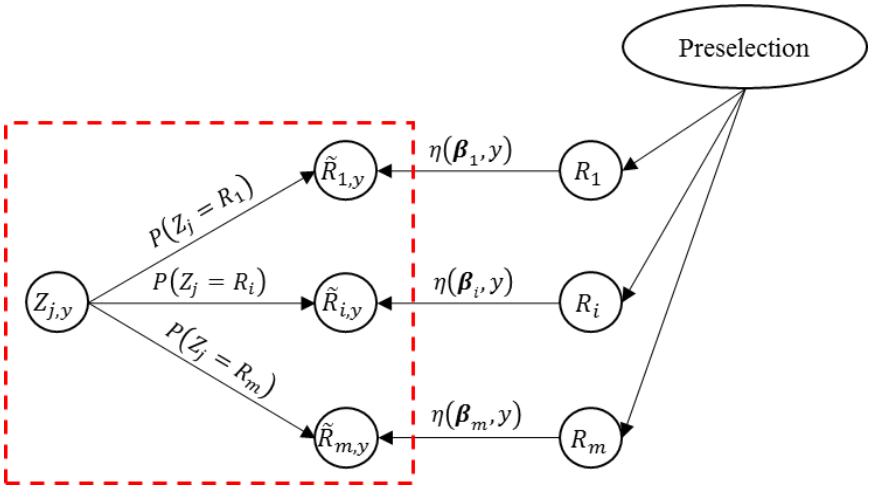

Fig. 2.2. The Extended Recommender System Framework

The conventional recommender system framework requires that all the utilized recommenders should have rated item y before; otherwise, these recommenders will not be informative for learning the attitude of user j for item y. However, it is common in practice that only a few recommenders have previously rated an item y (e.g., a new movie that has been recently released and not yet rated by many recommenders). In such a case, the estimated attitude for user j for item y based on equation (2.1) is dependent on a limited number of recommenders, which may not be reliable. To address this issue, further extensions have been proposed in the literature as shown in Fig 2.2 [59]. The dashed box contains the following unknown values that are calculated upon request: (i) the probability measure $P(Z_j = R_i)$, (ii) estimated ratings from the recommenders, and (iii) the estimated rating for the user. Unlike the conventional approach that only focuses on the

actual ratings, the extended framework utilizes the estimated rating from the recommenders based on their historical preference. In other words, the attitude of user j for item y, $Z_{j,y}$, is estimated as the following:

$$Z_{j,y} = \sum_{i=1}^{m} P\left(Z_j = R_i\right) * \tilde{R}_{i,y} + \epsilon_{j,y}. \quad (2.7)$$

Here, $\tilde{R}_{i,y}$ is the estimated rating for recommender i and it is assumed to follow the parametric form $\eta(\boldsymbol{\beta}_i, y)$, where $\boldsymbol{\beta}_i$ is estimated based on the recommenders' historical preference.

Although this approach tackles the challenge of missing values from each recommender (i.e., items that a recommender has never rated), it is very critical to accurately calculate $\boldsymbol{\beta}_i$ to obtain a reliable estimate $\tilde{R}_{i,y}$. In addition to that, $P\left(Z_j = R_i\right)$ may not be accurately estimated, if there is only a limited rating history for user j. For many implementers of recommender systems, they do not focus on users with limited history. This is because those users already expect poor recommendations due to their limited history.

2.3 Multi-output Gaussian Process (MOGP)

Gaussian process is a particular kind of statistical model where observations occur in a continuous domain, e.g. time or space. In a Gaussian process, every point in some continuous input space is associated with a normally distributed random variable and the main task of the Gaussian process is to model the covariance between any two input points. Gaussian processes are mostly designed to analyze single outputs and their main advantages are: (i) they allow for flexible choice of the parametric nature of a signal; and (ii) they can be utilized for continuous time analysis. The MOGP further fosters the correlation between the different outputs to better predict a set of multiple outputs. Here, we focus on a recent method, the collaborative MOGP, introduced by

Ngyuen and Bonilla (2014). Specifically, the authors assumed all the outputs (e.g., the degradation signals from the units) share Q latent functions, where each latent function follows an independent Gaussian process (GP). They also assumed that each output function has a GP prior. Finally, they mixed the stochastic processes from the latent functions and the output functions to produce a high-quality MOGP. The authors also provided the MATLAB code with the necessary libraries that have been used to produce the results in their paper. We use the provided code as a comparison study in Chapter 6. For more details about the formulation of the collaborative MOGP, please refer to [62].

While the MOGP shows reliable performance in rich datasets, they do have major limitations in presence of sparse datasets:

i. Assuming the parametric forms of the output functions are not known, GPs are suitable for interpolation but not extrapolation [57] which also extends to MOGPs. For the scenario in Table 2.1, MOGPs may not be reliable to approximate the value of y after the time point t_{end}. Note that it is possible to assume some basis functions that the outputs converge to; however, this removes the non-parametric nature of the MOGP and diminishes one of its major advantages.

ii. The wide majority of the existing MOGPs are based on maximizing the likelihood of the available output observations. Therefore, it is expected that units with a high number of observations to be well-fitted but units with limited observations may not be fitted accurately. For example in Table 2.1, unit 5 be fitted accurately because it is rich with observations; however, unit 4 may not be fitted properly because it has a limited set of observations.

iii. Similarly, time points with enough available observations may be fitted accurately, but time points with limited available observations may not be fitted accurately. For the scenario in Table 2.1, time points t_1 and t_2 are expected to be fitted accurately; however, predictions for time point t_k may not be reliable due to (i) the limited available observations at that time point and (ii) the unknown parametric form of the degradation model. Note this limitation will be severe for predictions at a time point near t_{end} because even the available observations may not help to interpolate and recover the missing values at such a time point.

iv. As a consequence of the third limitation, if there are a small subset of shared time points between different units, then MOGPs may not be effective because the correlation between the units is expected to be miss-calculated. In the scenario in Table 2.1, this is equivalent to having many missing values in a row. In such scenarios, the designed MOGP may be ill-conditioned and it is expected to produce inaccurate predictions.

Finally, MOGPs are computationally expensive which limits them from real time analysis in comparison to other parametric methods such as the mixed-effect models.

2.4 Remaining Useful Life Distribution Estimation

In the literature of prognostics, it is often assumed that a unit fails once its degradation signal exceeds a predefined failure threshold [8], [10], [29]. Thus, a common approach is to first model the degradation signal by an appropriate degradation model, e.g., the mixed-effects model mentioned in Section 2.1.2. Then, we can estimate the future evolution of the degradation signal based on the fitted degradation model, and predict the time at which the projected degradation

signal exceeds the predefined failure threshold. In particular, the cumulative distribution function of the RUL, can be written as:

$$P\left(\tilde{T}_j \le t \middle| \boldsymbol{s}_{j,.}\right) = P\left(s_{j,n_j+t} \ge D_j \middle| \boldsymbol{\phi}, \left(\boldsymbol{\theta}_j | \boldsymbol{s}_{j,.}\right)\right), \tag{2.8}$$

where $\tilde{T}_j$ is the estimated RUL of unit j; $\boldsymbol{s}_{j,.} = \left[s_{j,1}, s_{j,1}, \dots, s_{j,n_j}\right]^T$ is the set of observed degradation statuses for unit j up to the time point n_j; n_j is the time of the last available observation of unit j; s_{j,n_j+t} is the predicted degradation status at the time point $n_j + t$ and it can be estimated using the mixed-effects model, stochastic processes or recommender systems; and D_j is the predefined failure threshold for unit j.

2.5 Failure Threshold Estimation

Most of the existing work simply assumes that the failure threshold distribution is known *a priori*, i.e., D_j is a constant and deterministic value for all units. In most cases, such a threshold is estimated based on the physical knowledge of the system/process. When such knowledge is unavailable, a data-driven approach can be used to estimate the failure threshold [10], [31]. For example, Liu *et al.* [10] considered the statistics (mean and variance) of the last sensor observations (right before failure) from multiple historical units as an estimation of the failure threshold. In particular, in the absence of physical knowledge of the system/process, the mean and variance of the failure threshold can be calculated as the following:

$$\begin{cases} E\left(D_j\right) = \frac{\sum_{i=1}^{m} s_{i,n_i}}{n} \\ var\left(D_j\right) = \frac{\sum_{i=1}^{m}\left(s_{i,n_i} - E\left(D_j\right)\right)^2}{n-1} \end{cases}, \tag{2.9}$$

where s_{i,n_i} is the last observed degradation status just before failure in historical unit i; and m is the number of available historical units.

2.6 Data Fusion Approaches to Prognostics

The techniques mentioned in Section 2.1 are suitable for modeling one degradation signal. However, in many applications, there are multiple embedded sensors in a unit. As a result, it is highly desired to develop effective data fusion models that can combine the multiple-sensor data to produce better inferences regarding the future status of a unit. Generally speaking, data fusion methods for prognostics can be classified into three categories [7], [24] depending on the level of implementation of the fusion model: data-level fusion [15]–[17], [63], feature-level fusion [64], and decision-level fusion [1], [65]. A review on multiple sensor data fusion approaches to condition monitoring, fault diagnosis, and prognostics can be found in [8]. As the feature development can be regarded as a transformation of the original sensor data, data-level and feature-level fusion techniques essentially focus on the same scheme that develops a composite health index $h_{i,t}$ for each unit i at each observation epoch t via combining multiple sensor signals $\boldsymbol{s}_{i,.,t} = [s_{i,1,t} \dots, s_{i,s,t}] \in R^{1\times S}$, i.e., $h_{i,t} = f(\boldsymbol{s}_{i,.,t})$. Here, S denotes the number of available sensors and $f(\cdot)$ is the fusion model. On the basis of the developed health index, a degradation model can then be built to describe the evolution of the degraded unit over time, and further be used to compute its remaining lifetime in real time. For example, using the technique mentioned in Section 2.1, the health index can be characterized as [25]:

$$h_{i,t} = f(\boldsymbol{s}_{i,.,t}) = \eta(\boldsymbol{\phi}, \boldsymbol{\theta}_i, t) + \varepsilon_{i,t}, \tag{2.10}$$

2.7 Data-level Fusion Models for Prognostics

For the data-level fusion model development, one of such research efforts was done recently by Liu *et al.* [10]. The authors considered two unique properties that the developed health index should possess for successful prognostics:

Property 1: Once an initial fault occurs, the trend of the health index should be monotonic.

Property 2: Given the same environmental condition and failure mode, the variance in the failure threshold of the developed health index should be minimal.

In other words, Property 1 tries to minimize the total amount of violation in the monotonicity of the developed health index: $\sum_{i=1}^{m}\sum_{t=1}^{n_i-1}\Delta_{i,t}$, where $\Delta_{i,t}= max\left(h_{i,t}-h_{i,t+1},0\right)$ represents the violation in the monotonicity for unit i at time t if we assume the developed health index has an increasing trend and each unit starts to degrade at the initial observation. Here, n_i is the total number of observations for unit i; and m is the total number of units in the training stage. On the other hand, Property 2 tries to minimize the variance in the failure threshold of the developed health index: $\sum_{i=1}^{m}\left(h_{i,n_i}-\bar{h}_{.,n_i}\right)^2/(m-1)$, where h_{i,n_i} is the constructed health index of unit i at time n_i (i.e., the time right before the unit fails) and $\bar{h}_{.,n_i}=\sum_{i=1}^{m}h_{i,n_i}/m$ is the average of the last health index value before failure in all units. The authors further used a linear link function to construct the health index, i.e., $h_{i,t}=\boldsymbol{y}_{i,.,t}\cdot\boldsymbol{w}$ and derived the optimal weight $\boldsymbol{w}^*$ by minimizing both properties 1 and 2. Although such method provides a promising approach, there is no guarantee that the developed health index is suitable to the selected degradation model when performing prognostics. This is because this data-level fusion model separates the fusion procedure and the degradation modeling into two disjoint tasks.

To address this limitation, Liu and Huang [11] further proposed a data-level fusion model that considered both the degradation modeling and fusion procedure in a unified manner to improve prognostic performance. Specifically, the authors tried to minimize both the model fitting errors, $\sum_{i=1}^{m}\left(\boldsymbol{e}_{i,\cdot}\right)'\boldsymbol{e}_{i,\cdot}/m$ and the variance in the failure threshold of the developed health index: $\sum_{i=1}^{m}\left(h_{i,n_i}-\bar{h}_{\cdot,n_i}\right)^2/(m-1)$. Here, $\boldsymbol{e}_{i,\cdot}=\left[e_{i,1},\ldots,e_{i,n_i}\right]'$ represents the residual errors (i.e., $e_{i,t}=h_{i,t}-\eta\left(\boldsymbol{\phi},\tilde{\boldsymbol{\theta}}_i,t\right)$) in the fitted degradation model for the constructed health index of unit i, and $\left(\boldsymbol{e}_{i,\cdot}\right)'\boldsymbol{e}_{i,\cdot}$ is the residual sum of squares of unit i. In other words, the authors put forward another important property that the health index should possess for successful prognostics:

Property 3: Given the selected degradation model, the model fitting errors for the developed health index should be minimal.

Chapter 3

Optimize the Signal Quality of the Composite Health Index via Data Fusion for Degradation Modeling and Prognostic Analysis

This Chapter develops a data-level fusion methodology to construct a composite health index by combining multiple degradation-based sensor data. As mentioned previously, one challenging question needs to be resolved first is to identify the desired properties of the degradation signal and then provide a quantitative metric for characterizing the signal quality of the degradation signal by combining these identified properties. In science and engineering, SNR is a commonly used quality measure that compares the level of a desired signal to the level of background noise. Specifically, the SNR is defined as $SNR = u/\sigma$, where u is the signal mean or expected value and σ is the standard deviation of the noise. However, it is straightforward to see that this definition cannot be effectively used here as the degradation signal evolves across the entire lifecycle of the unit and exhibits a unique characteristic pattern. To address this issue, we aim to develop a new SNR metric tailored for the needs of degradation signals. We denote this new quality metric as SNR^d, and based on this, we further propose a novel data fusion methodology that optimizes this metric SNR^d when constructing the health index.

3.1 Desired Properties and SNR of the Degradation Signal

To begin with, we would like to first provide an illustrative example to show that the properties 2 and 3 used in [11] are not sufficient to quantify the signal quality of a degradation signal. Fig.

3.1 shows the degradation signals of two different sensors, in which the model fitting errors (by using the same type of degradation model) and the variances in the failure threshold by using both sensors are the same. In addition, since both sensors are used to monitor the same unit, the two signals have the same number of observations as well. However, it clearly shows that the degradation signal from sensor 1 is more useful for prognostic application as it exhibits a clearer degradation trend (i.e., a larger slope of the data). Prognostics typically requires extrapolating the partially degraded sensor data to the failure threshold based on the selected degradation model. As the range information of the sensor 1 is much larger than sensor 2, sensor 1 shows a shorter period of observations within the variation zone of the failure threshold. In such a case, the predicted failure time by using sensor 1 will be within a smaller range, leading to a more accurate prognostic estimation than using sensor 2. In other words, the range information of the degradation signal, starting from the initial to the final observation that is across the entire lifecycle of a unit is also another important characteristic for measuring the quality of the degradation signal.

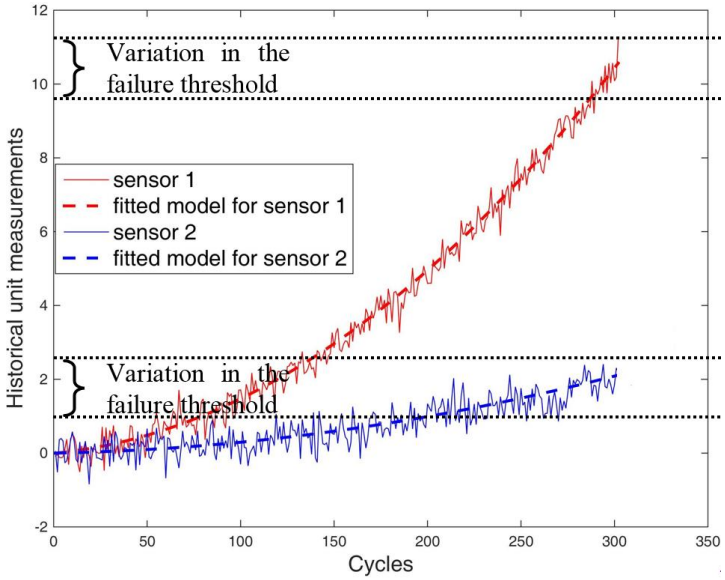

Fig. 3.1. An example of showing the range information of the degradation signal is also another important characteristic for successful prognostics.

Based on this observation, we define the SNR of a degradation signal to be the ratio of the range of the degradation signal during the entire course of the degradation process over the uncertainties occurred during the prognostic analysis. Such uncertainties have two folds: 1) the random errors introduced in the degradation model; and 2) the variation in the failure threshold. The first uncertainty arises when we use a selected degradation model to fit the degradation signal, while the second uncertainty stems from the stochastic nature of the degradation process such that different units fail at different time points. Mathematically, the SNR of a degradation signal, SNR^d is defined as:

$$SNR^d = \frac{R^2}{\sigma^2 + v}, \tag{3.1}$$

where R is the range information of the degradation signal, σ^2 is the variance of the model fitting errors, and v is the variance of the failure threshold. Specifically, for the constructed health index, R can be estimated by $\sum_{i=1}^{m}(h_{i,n_i} - h_{i,1})/m$, v can be estimated by $\sum_{i=1}^{m}\left(h_{i,n_i} - \bar{h}_{\cdot,n_i}\right)^2/(m-1)$, and σ^2 can be estimated by $\sum_{i=1}^{m}\left(\boldsymbol{c}_{i,\cdot}\boldsymbol{e}_{i,\cdot}\right)'\boldsymbol{c}_{i,\cdot}\boldsymbol{e}_{i,\cdot}/m$. The detailed variable meanings can be referred to Section 3.3. Here, we consider R^2 instead of R in the numerator as the degradation signal may exhibit either an increasing or decreasing trend. Another advantage is that this definition of SNR is scale-invariant with respect to the measurement units and the scales of degradation signals. This new SNR formula for the degradation signal in (3.1) resembles the conventional definition of SNR, which also measures the strength (i.e., the range) of the degradation signal over the background noises (i.e., the two types of uncertainty involved in the degradation signal), but tailored to the specific needs of degradation signals.

If we compare the formulation in (3.1) with the efforts in Liu *et al.* [10], we can see that Property 1 aims at maximizing the monotonicity in the degradation signal, which is similar to the maximization of the range of the degradation signal. However, monotonicity focuses on comparing the signal strength at each two consecutive observations, i.e., $h_{i,t} - h_{i,t+1}$, whereas the range focuses on maximizing the signal strength in the entire lifecycle of the unit, i.e., $h_{i,n_i} - h_{i,1}$. On the other hand, the approach in Liu and Huang [11] only focuses on minimizing the noise components (i.e., the denominator in (3.1)) of the degradation signal but ignores maximizing the signal strength.

3.2 Selected Degradation Models

Based on the new SNR definition in (2), our goal next is to develop a data fusion model that combines the information from multiple sensors to construct a health index with a maximal SNR measure. As the model fitting errors (i.e., σ^2 in (2)) are involved in the definition, we need to first specify the selected degradation model used in the fusion procedure.

The degradation model in (2.1) is generic to provide different types of functional forms for modeling degradation signals. To highlight the main ideas of our proposed data fusion method, here we focus on a specific parametric degradation model with the p-th order polynomial form:

$$L_{i,j,t} = \sum_{k=0}^{p} \theta_{i,j}^{(k)} t^k + \varepsilon_{i,j,t}, \tag{3.2}$$

where $L_{i,j,t}$ is the sensor measurement for unit i, sensor j and time t; $\boldsymbol{\Gamma}_{i,j} = \left[\theta_{i,j}^{(0)}, \ldots, \theta_{i,j}^{(p)}\right]'$ is the random-effect parameter and often assumed to follow a multivariate normal distribution, i.e., $\boldsymbol{\Gamma}_{i,j} \sim N_{p+1}(\boldsymbol{u}_j^0, \boldsymbol{\Sigma}_j^0)$; and $\varepsilon_{i,j,t}$ is the random noise. Many degradation models discussed in the

existing literature belongs to the form in (3.2) after transformations, such as the simple random coefficient growth model [66], [67] and the exponential functional form model[10], [28], [56], [68]. For example, Liu *et al.* [10]considered an exponential degradation model:

$$x_{i,j,t} = \phi_j + \alpha_{i,j} e^{\theta_{i,j}^{(1)} t + \theta_{i,j}^{(2)} t^2 + \varepsilon_{i,j,t} - \sigma_j^2/2}, \tag{3.3}$$

where ϕ_j is the constant deterministic parameter for sensor j; $\alpha_{i,j}$, $\theta_{i,j}^{(1)}$ and $\theta_{i,j}^{(2)}$ are random-effect parameters; and $\varepsilon_{i,j,t}$ follows $N(0, \sigma_j^2)$. Since $E\left(e^{\varepsilon_{i,j,t} - \sigma_j^2/2}\right) = 1$, $E\left(x_{i,j,t} \middle| \alpha_{i,j}, \theta_{i,j}^{(1)}, \theta_{i,j}^{(2)}\right) = \phi_j + \alpha_{i,j} e^{\theta_{i,j}^{(1)} t + \theta_{i,j}^{(2)} t^2}$. For convenience, the authors applied a log-transformation and focused on modeling the logged signal:

$$L_{i,j,t} = ln\left(x_{i,j,t} - \phi_j\right) = \theta_{i,j}^{(0)} + \theta_{i,j}^{(1)} t + \theta_{i,j}^{(2)} t^2 + \varepsilon_{i,j,t}, \tag{3.4}$$

where $\theta_{i,j}^{(0)} = ln\alpha_{i,j} - \sigma_j^2/2$. In addition, the authors assumed that the random-effect parameters $\theta_{i,j}^{(0)}$, $\theta_{i,j}^{(1)}$ and $\theta_{i,j}^{(2)}$ are jointly distributed and follow a multivariate normal distribution: $\boldsymbol{\Gamma}_{i,j} = \left[\theta_{i,j}^{(0)}, \theta_{i,j}^{(1)}, \theta_{i,j}^{(2)}\right]' \sim N_3(\boldsymbol{u}_j^0, \boldsymbol{\Sigma}_j^0)$.

3.3 Data Fusion Model Based on the New SNR Definition

Recall that s represents the number of available sensors, m is the total number of historical units in the training stage, n_i is the number of available observations for unit i, $\boldsymbol{L}_{i,.,t} = [L_{i,1,t} \dots, L_{i,s,t}] \in R^{1 \times s}$ represents the multiple sensor data for unit i and time t, and $f(\cdot)$ is the link function to construct the health index from multiple sensor data, i.e., $h_{i,t} = f\left(\boldsymbol{L}_{i,.,t}\right)$. Specifically, in this Chapter, we consider the linear link function as a demonstration of our main idea.

Consequently, $\boldsymbol{h}_{i,.} = \boldsymbol{L}_{i,.,.}\,\boldsymbol{w}$, where $\boldsymbol{w} \in R^{s\times 1}$ represents the vector of weight coefficients to combine the multiple sensor data, $\boldsymbol{L}_{i,.,.} = [\boldsymbol{L}_{i,.,1}; \dots; \boldsymbol{L}_{i,.,n_i}] \in R^{n_i\times s}$.

Denote $\boldsymbol{Y_0} = [\boldsymbol{L}_{1,.,n_i} - \boldsymbol{L}_{1,.,1}; \dots; \boldsymbol{L}_{m,.,n_i} - \boldsymbol{L}_{m,.,1}] \in R^{m\times s}$ as the range information matrix that records the difference between the last and the initial observations with the rows representing each training unit and the columns representing each sensor. The numerator in (3.1) (i.e., R^2) for the constructed health index can then be estimated by $\frac{1}{m^2}\boldsymbol{w'Y_0}'\boldsymbol{11'Y_0w}$, where $\mathbf{1}$ is the column vector with all elements equal to 1.

Next, we consider the denominator part of SNR in (3.1). Denote $\boldsymbol{\Psi}_i \in R^{n_i\times(p+1)} = \begin{bmatrix} 1 & \dots & 1 \\ \dots & \dots & \dots \\ 1 & \dots & t^p \\ \dots & \dots & \dots \\ 1 & \dots & n_i^p \end{bmatrix}$ as the design matrix, $\boldsymbol{\Gamma}_i \in R^{(p+1)\times 1}$ as the random-effect parameter of the degradation model for the health index of unit i, and $e_{i,t}$ as the residual term for unit i and time t when fitting the health index by using the polynomial degradation model in (3.2). In other words, $e_{i,t}$ can be expressed by $e_{i,t} = h_{i,t} - [1, \dots, t^p]\boldsymbol{\Gamma}_i$. Here, we consider a weighted least square regression model, and let $c_{i,t}$ be the weight coefficient for the residual term $e_{i,t}$ and $\boldsymbol{c}_{i,.} = diag(c_{i,1}, \dots, c_{i,n_i}) \in R^{n_i\times n_i}$ be the weight coefficient matrix for the residual errors $\boldsymbol{e}_{i,.} = \left[e_{i,1}, \dots, e_{i,n_i}\right]' \in R^{n_i\times 1}$. Given the information from multiple sensor data $\boldsymbol{L}_{i,.,.}$, the random-effect parameter can be estimated by $\boldsymbol{\Gamma}_i = \left(\boldsymbol{\Psi}_i'\boldsymbol{c}_{i,.}^2\boldsymbol{\Psi}_i\right)^{-1}\left(\boldsymbol{\Psi}_i'\boldsymbol{c}_{i,.}^2\boldsymbol{L}_{i,.,.}\boldsymbol{w}\right)$ using the weighted least square approach. As a result, the weighted residual terms for unit i at different times can be expressed as $\boldsymbol{c}_{i,.}\boldsymbol{e}_{i,.} = \boldsymbol{c}_{i,.}\boldsymbol{L}_{i,.,.}\boldsymbol{w} - \boldsymbol{c}_{i,.}\boldsymbol{\Psi}_i\boldsymbol{\Gamma}_i = (\boldsymbol{I}-\boldsymbol{H})\boldsymbol{c}_{i,.}\boldsymbol{L}_{i,.,.}\boldsymbol{w} \in R^{n_i\times 1}$, in which $\boldsymbol{H} =$

$\boldsymbol{c}_{i,\cdot}\boldsymbol{\Psi}_i\left(\boldsymbol{\Psi}_i'\boldsymbol{c}_{i,\cdot}^2\boldsymbol{\Psi}_i\right)^{-1}\boldsymbol{\Psi}_i'\boldsymbol{c}_{i,\cdot} \in R^{n_i \times n_i}$ is known as the projection matrix. In other words, the model fitting uncertainty in the degradation model (i.e., σ^2 in (3.1)) can be estimated by $\sum_{i=1}^{m}\left(\boldsymbol{c}_{i,\cdot}\boldsymbol{e}_{i,\cdot}\right)'\boldsymbol{c}_{i,\cdot}\boldsymbol{e}_{i,\cdot}/m = \sum_{i=1}^{m}\boldsymbol{w}'\boldsymbol{L}_{i,\cdot,\cdot}'\boldsymbol{c}_{i,\cdot}(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{c}_{i,\cdot}\boldsymbol{L}_{i,\cdot,\cdot}\boldsymbol{w}/m$. On the other hand, since $h_{i,n_i} = \boldsymbol{L}_{i,\cdot,n_i}\boldsymbol{w}$ and $\bar{h}_{\cdot,n_i} = \sum_{i=1}^{m}h_{i,n_i}/m$, the variance in the failure threshold (i.e., v in (3.1)) can be estimated by $\sum_{i=1}^{m}\left(h_{i,n_i}-\bar{h}_{\cdot,n_i}\right)^2/(m-1) = \boldsymbol{w}'\boldsymbol{L}_{\cdot,\cdot,n_i}'\boldsymbol{D}\boldsymbol{L}_{\cdot,\cdot,n_i}\boldsymbol{w}$ (see Appendix for details), where $\boldsymbol{L}_{\cdot,\cdot,n_i} \in R^{m\times s}$ is the matrix recording the last observations before failure with the rows representing each historical unit and the columns representing each sensor, and $\boldsymbol{D} \in R^{m\times m} = \frac{\boldsymbol{I}-\boldsymbol{O}/m}{m-1}$ is a symmetric matrix, in which $\boldsymbol{O}$ is a matrix of all ones and $\boldsymbol{I}$ is an identity matrix.

As a result, the SNR of the constructed health index in (3.1) can be mathematically expressed as follows:

$$SNR^d = \frac{\boldsymbol{w}'\boldsymbol{Y_0}'\boldsymbol{1}\boldsymbol{1}'\boldsymbol{Y_0}\boldsymbol{w}/m^2}{\boldsymbol{w}'\boldsymbol{L}_{\cdot,\cdot,n_i}'\boldsymbol{D}\boldsymbol{L}_{\cdot,\cdot,n_i}\boldsymbol{w}+\sum_{i=1}^{m}\boldsymbol{w}'\boldsymbol{L}_{i,\cdot,\cdot}'\boldsymbol{c}_{i,\cdot}(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{c}_{i,\cdot}\boldsymbol{L}_{i,\cdot,\cdot}\boldsymbol{w}/m}. \quad (3.5)$$

It is straightforward to see that the function value of (3.5) does not depend on the scale of $\boldsymbol{w}$. In other words, the SNR of the health index will remain constant (i.e., *scale-invariant*) for a solution $\boldsymbol{w}$ multiplied by any non-zero constant factor.

Our goal here is to find the optimal $\boldsymbol{w}^*$ that maximizes SNR^d in (3.5). As what follows, we will show that the optimal solution can be obtained by the eigendecomposition approach. Let matrix $\boldsymbol{B} \in R^{s\times s} = \boldsymbol{L}_{\cdot,\cdot,n_i}'\boldsymbol{D}\boldsymbol{L}_{\cdot,\cdot,n_i} + \sum_{i=1}^{m}\boldsymbol{L}_{i,\cdot,\cdot}'\boldsymbol{c}_{i,\cdot}(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{c}_{i,\cdot}\boldsymbol{L}_{i,\cdot,\cdot}$ and matrix $\boldsymbol{A} \in R^{s\times s} = \boldsymbol{Y_0}'\boldsymbol{1}\boldsymbol{1}'\boldsymbol{Y_0}/m^2$. Then, the optimal solution $\boldsymbol{w}^*$ is the eigenvector corresponding to the largest eigenvalue λ_{max} by solving the generalized eigenvalue problem $\boldsymbol{A}\boldsymbol{w} = \lambda_{max}\boldsymbol{B}\boldsymbol{w}$ (see Appendix for details). Here, we can show that both $\boldsymbol{B}$ and $\boldsymbol{A}$ are symmetric positive semidefinite (P.S.D.) matrices (see Appendix

for details). In the case that all the eigenvalues of matrix $\boldsymbol{B}$ are positive (i.e., $\boldsymbol{B}$ is a positive definite matrix), its inverse matrix exists and thus $\boldsymbol{w}^*$ is the eigenvector corresponding to the largest eigenvalue of $\boldsymbol{B}^{-1}\boldsymbol{A}$. Otherwise, we may introduce any small positive number r to make all the eigenvalues of $\boldsymbol{B}+r\boldsymbol{I}$ positive [69]. In such a case, the matrix $\boldsymbol{B}+r\boldsymbol{I}$ becomes invertible and the optimal solution $\boldsymbol{w}^*$ is the eigenvector corresponding to the largest eigenvalue of $(\boldsymbol{B}+r\boldsymbol{I})^{-1}\boldsymbol{A}$. Here, it is worth mentioning that as long as the largest eigenvalue of $\boldsymbol{B}^{-1}\boldsymbol{A}$ is unique, we will achieve a unique solution $\boldsymbol{w}^*$ with unit length. Due to the randomness in the multiple sensor data, the chance for the largest eigenvalue of $\boldsymbol{B}^{-1}\boldsymbol{A}$ being non-unique is extremely small and thus is ignored in the current Chapter.

Recall that in (3.5), $c_{i,t}$ is the weight coefficient corresponding to the residual term $e_{i,t}$ in the fitted degradation model. Below we provide some guidelines in setting $\{c_{i,t}\}$. Since the accuracy of the remaining life prediction becomes increasingly sensitive to the model fitting result as the unit approaches to failure, Liu and Huang [11] proposed to assign $\{c_{i,t}\}$ as an increasing series: $c_{i,t+1} \geq c_{i,t} \geq 0, i = 1, \dots, m, t = 1, \dots, n_i - 1$. If we assume each training unit is equally important, the following constraints need to be satisfied as well: $\sum_{t=1}^{n_i} c_{i,t} = 1, i = 1, \dots, m$. Under these two assumptions, the authors further provided two options for setting the weight coefficients by assuming $\{c_{i,t}\}$ either follows (i) an arithmetic series (i.e., $2c_{i,t} = c_{i,t+1} + c_{i,t-1}$) or (ii) a geometric series (i.e., $c_{i,t}^2 = c_{i,t+1}c_{i,t-1}$) depending on the emphasis placed on the model fitting:

(i) For the arithmetic series:

$$c_{i,t} = c_{i,1} + (t-1)\frac{2-2c_{i,1}n_i}{(n_i-1)n_i}, t = 1, \dots, n_i.$$

(ii) For the geometric series:

$$c_{i,t} = c_{i,1} \cdot q^{t-1},\ t = 1, \dots, n_i,$$

where q statisfies $c_{i,1}q^{n_i} - c_{i,1} - q + 1 = 0$.

3.4 Case Study

In this Section, we will implement and evaluate our proposed (we call it SNR-based) data fusion model when the constructed health index is used for prognostics. Specially, we will consider three benchmark approaches that include using (1) each original individual sensor data, (2) the health index constructed based on the model in [10] (we call it non-parametric data fusion model), and (3) the health index constructed based on the in [11] (we call it semi-parametric data fusion model).

We focus on the degradation-based dataset of turbofan engines under a single failure mode and a single operation condition. The dataset is provided in [32] and consists of 100 training units (i.e., $m = 100$) that include a total of 20631 observations (i.e., $\sum_{i=1}^{m} n_i = 20631$), 100 testing units that include a total of 13096 observations, and a file that records the actual remaining lifetime of the 100 testing units. Measurements of 21 sensors that include comprehensive information (e.g., temperature, pressure, speed) at different locations of the engine are collected after each observation epoch for each unit. The detailed descriptions of these 21 sensors are given in Table 3.1.

TABLE 3.1. DETAILED DESCRIPTION OF THE 21 SENSORS [27].

Symbol	Description	Units
T2	Total temperature at fan inlet	°R
T24	Total temperature at LPC outlet	°R
T30	Total temperature at HPC outlet	°R
T50	Total temperature at LPT outlet	°R
P2	Pressure at fan inlet	psia
P15	Total pressure in bypass-duct	psia
P30	Total pressure at HPC outlet	psia

Nf	Physical fan speed	rpm
Nc	Physical core speed	rpm
epr	Engine pressure ratio (P50/P2)	--
Ps30	Static pressure at HPC outlet	psia
phi	Ratio of fuel flow to Ps30	pps/psi
NRf	Corrected fan speed	rpm
NRc	Corrected core speed	rpm
BPR	Bypass Ratio	--
farB	Burner fuel-air ratio	--
htBleed	Bleed Enthalpy	--
Nf_dmd	Demanded fan speed	rpm
PCNfR_dmd	Demanded corrected fan speed	rpm
W31	HPT coolant bleed	lbm/s
W32	LPT coolant bleed	lbm/s

The dataset is generated by C-MAPSS, a model for simulating realistic large commercial turbofan engines [23], [27]. C-MAPSS has been widely used to characterize the degradation performance in engines due to wear and tear according to the usage pattern of the engines. However, this simulation model is not explicitly accessible to the users [27]. Thus, the underlying assumption here is that users need to rely on the available datasets from these 21 sensors, which include the historical degradation data of the 100 training units and the *in-situ* sensor measurements collected during the testing stage to infer the remaining lifetime of each testing unit and further compare the results with the actual ones.

3.4.1 Data Preprocessing

Before implementing our proposed SNR-based data fusion model, it is required to first determine which sensors to use. In order to achieve a fair comparison, here we use the exactly same amount of sensor information as mentioned in [10] and [11]. Specifically, the authors [10], [11] selected the sensors as an input to the data fusion model if their degradation signals exhibit

an increasing or decreasing trend. In other words, a sensor will be selected if its last observation is consistently larger (increasing trend) or smaller (decreasing trend) than the initial observation in all the training units. This is a meaningful criterion as the engine fails under a single failure mode and a single operation condition, and thus a reliable sensor signal should present a similar degradation trajectory in all the units. By using this criterion, 11 (i.e., $s = 11$) out of 21 sensors are selected, which include T24, T50, P30, Nf, Ps30, phi, NRf, BPR, htBleed, W31 and W32. As these sensors measure different characteristics of the engine with distinct units of measurement (e.g., psia, rpm, lbm/s), we further standardize the dataset before the data fusion as mentioned in [10] and [11]. Moreover, an arithmetic series for $\{c_{i,t}\}$ is used by assuming the weight coefficient $c_{i,t}$ is linearly increasing, which is consistent with the ones in [10] and [11].

3.4.2 SNR-Based Data Fusion Model

Since, the authors [10], [11] considered the exponential degradation model in (3.3) and showed that it provided a good model fitting for the degradation dataset. Here, for a fair comparison, we consider the same degradation model. By using the proposed SNR-based data fusion model, the optimal weight $\boldsymbol{w}^*$ that combines multiple sensor data to construct the health index is presented in Table 3.2. Note that in order to be consistent with [10] and [11], the optimal weight $\boldsymbol{w}^*$ has been normalized from the obtained eigenvector in (3.5) such that the sum of the absolute values of all weight coefficients equals one. This normalization can be done as the equation (3.5) does not depend on the scale of $\boldsymbol{w}$. In this way, the constructed health index is a weighted average of the selected sensor data, $h_{i,t} = \boldsymbol{L}_{i,.,t}\, \boldsymbol{w}^*$, in which $\boldsymbol{L}_{i,.,t}$ is the sensor measurements collected from the selected 11 sensors for unit i and time t.

Fig. 3.2 shows an example of the model fitting result by using the exponential model in (3.3) to each selected sensor data and the constructed health index for a training unit. It clearly shows that the developed health index by using the SNR-based data fusion model provides a better model fitting result than each original sensor data.

TABLE 3.2. OPTIMAL WEIGHT $\boldsymbol{w}^*$ FOR THE SELECTED SENSOR DATA.

Name	T24	T50	P30	Nf	Ps30	phi
Value	0.0793	0.1263	-0.0911	0.0200	0.1664	-0.1298
Name	NRf	BPR	htBleed	W31	W32	
Value	0.0346	0.0983	0.0815	-0.0871	-0.0856	

TABLE 3.3. SNR METRIC VALUES OF ALL SELECTED SENSORS AND THE HEALTH INDICES BASED ON DIFFERENT DATA FUSION MODELS.

Name	T24	T50	P30	Nf	Ps30	phi	NRf
Value	9.912	25.985	18.580	10.147	32.835	23.154	11.225
Name	BPR	htBleed	W31	W32	HI-non	HI-semi	HI-SNR
Value	13.392	10.165	14.581	12.878	108.840	141.142	143.379

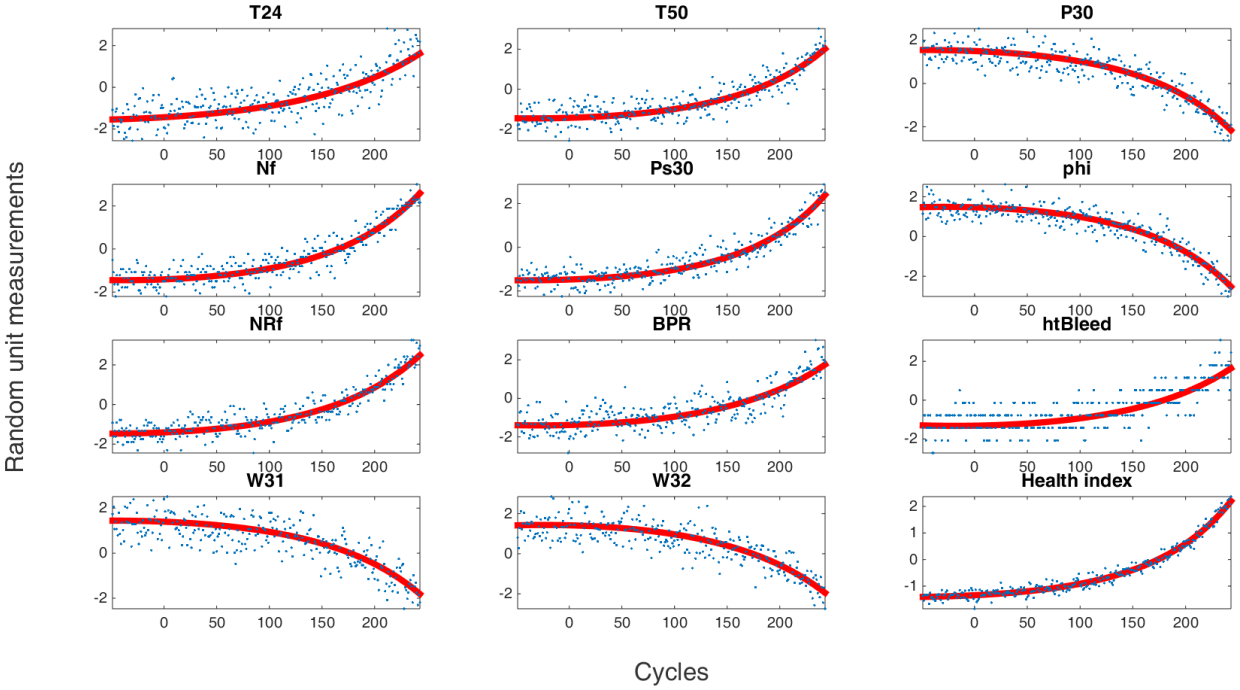

Fig. 3.2. An illustration plot of the model fittings results in all selected sensor data and the constructed health index for a training unit.

Table 3.3 further calculates and compares the *SNR metric* defined in (3.1) for all the selected sensors, the constructed health indices based on the non-parametric data fusion model (HI-non), the semi-parametric data fusion model (HI-semi), and the proposed SNR-based data fusion model (HI-SNR). The result shows that the SNR-based data fusion model is able to construct a health index with a maximal SNR measure. In Section 3.4.4, we will further demonstrate that the constructed health index with a higher SNR measure will provide a more accurate remaining life prediction result.

Another thing we would like to emphasize is that similar to the semi-parametric data fusion model in [11], the proposed SNR-based data fusion model only requires solving for $s = 11$ variables, whereas the non-parametric data fusion model in [10] needs to solve for $s + \sum_{i=1}^{m}(n_i - 1) = 20542$ variables via a large quadratic programming. Thus, the computational cost of the proposed SNR-based data fusion model is much smaller than the non-parametric data fusion model in [10], but similar to the semi-parametric data fusion model in [11].

3.4.3 Robustness Study

In the followings, we would like to further provide a robustness analysis for the derived optimal weight, $\boldsymbol{w}^*$ by using the proposed SNR-based data fusion algorithm when there are different degrees of missing data. Specifically, we randomly hide different percentage of observations from the 100 training units and then implement our proposed SNR-based data fusion model based on the incomplete dataset. As the initial and last observations in all training units are needed to estimate the range information as well as the variance in the failure threshold, we require all initial and last observations are available when constructing the missing dataset.

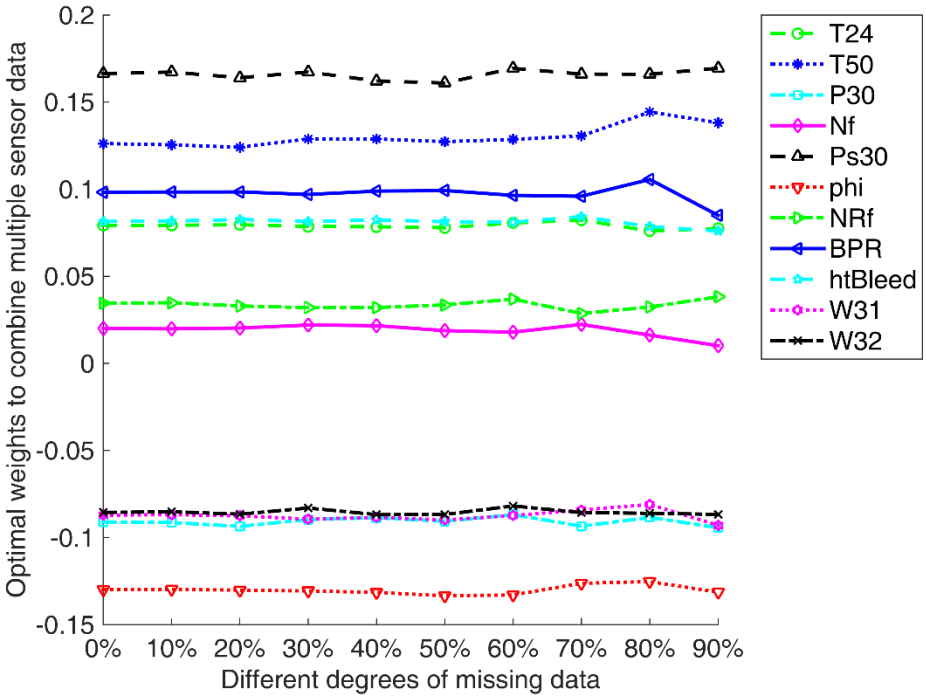

Fig. 3.3. Optimal weights derived at different degrees of missing data.

Fig. 3.3 shows the changes in the optimal weight $\boldsymbol{w}^*$ based on different degrees of missing data. For example, the label "10%" refers to the case that random $(20631 - 200) * 10\% \approx 2043$ observations are missing when learning the optimal weights. Recall that "20631" is the total number of observations in all 100 training units and "200" stands for the initial and last observations in these units.

According to Fig. 3.3, it clearly shows that the SNR-based data fusion model is relatively robust especially when there is less than or equal to 20% of missing data. This result is similar to the one given in [11] which is based on minimizing only the model fitting errors and the variance in the failure threshold. This study shows the advantage of the SNR-based data fusion model over the non-parametric data fusion model in [10] that requires all observations must be available and

sampled at a regular interval, especially when the measurements are costly and some missing data are inevitable.

3.4.4 Estimation of the RUL Distribution

In this Section, we would like to further evaluate the performance of our constructed health index when it is used for real time prognostics. To answer this question, one standard approach is the one proposed by Gebraeel [56] who introduced a Bayesian updating approach by incorporating real time degradation signals for calculating the predicted remaining lifetime of each individual unit. Specifically, the algorithm starts from learning the prior joint distributions of the random-effect parameter $\boldsymbol{\Gamma}_{i,j}$ by fitting the degradation signals of each historical unit with the selected degradation model (e.g., see Fig. 3.2). Next, the prior joint distribution of the random-effect parameter $\boldsymbol{\Gamma}_{i,j}$ is calibrated based on the *in-situ* sensor data collected from each new operating unit in the testing stage. Then, based on the updated degradation model, the RUL of this particular unit can be calculated. It has been shown that this Bayesian updating approach outperformed either a similar model based on [25] that did not incorporate the *in-situ* degradation information, or a model that was based entirely on calculating the conditional RUL without considering historical data. Consequently, this Chapter adopts this modeling framework as an illustration for remaining life prediction.

As shown in Fig. 3.2, the degradation model in (3.3) can provide a good model fitting result. The authors [10] further assumed that the random-effect parameters $\theta_{i,j}^{(0)}$, $\theta_{i,j}^{(1)}$ and $\theta_{i,j}^{(2)}$ were jointly distributed and followed a multivariate normal distribution: $\boldsymbol{\Gamma}_{i,j} = [\theta_{i,j}^{(0)}, \theta_{i,j}^{(1)}, \theta_{i,j}^{(2)}]' \sim N_3(\boldsymbol{u}_j^0, \boldsymbol{\Sigma}_j^0)$. Here, we adopt the same approach in order to provide a fair

comparison. Let $\boldsymbol{L}_{i,j,.} = \left[L_{i,j,1}, \dots, L_{i,j,n_i}\right]' \in R^{n_i \times 1}$ be the sequence of the sensor measurements collected up to the current observation epoch n_i for unit i and sensor j. Then, the posterior distribution of $\boldsymbol{\Gamma}_{i,j}$ still follows a multivariate normal distribution:

$$\boldsymbol{\Gamma}_{i,j} | \boldsymbol{L}_{i,j,.} \sim N_3(\boldsymbol{u}_j^1, \boldsymbol{\Sigma}_j^1), \tag{3.6}$$

where $\boldsymbol{u}_j^1 = \left(\frac{\boldsymbol{\Psi}_i{}'\boldsymbol{\Psi}_i}{\sigma_j^2} + (\boldsymbol{\Sigma}_j^0)^{-1}\right)^{-1} \left(\frac{\boldsymbol{\Psi}_i{}'\boldsymbol{L}_{i,j,.}}{\sigma_j^2} + (\boldsymbol{\Sigma}_j^0)^{-1}\boldsymbol{u}_j^0\right)$, $\boldsymbol{\Sigma}_j^1 = \left(\frac{\boldsymbol{\Psi}_i{}'\boldsymbol{\Psi}_i}{\sigma_j^2} + (\boldsymbol{\Sigma}_j^0)^{-1}\right)^{-1}$ and $\boldsymbol{\Psi}_i \in R^{n_i \times 3} = \begin{bmatrix} 1 & \dots & 1 \\ \dots & \dots & \dots \\ 1 & \dots & t^2 \\ \dots & \dots & \dots \\ 1 & \dots & n_i^2 \end{bmatrix}$ (see Appendix for details).

Based on the updated parameter $\boldsymbol{\Gamma}_{i,j} | \boldsymbol{L}_{i,j}(\cdot)$, below we derive the RUL, the distribution of the time at which the sensor signal crosses the failure threshold. Since different units may actually fail at different threshold values, we consider the failure threshold as also a random variable. Denote l_j as the failure threshold for sensor j and $\tilde{T}_{i,j}$ as the estimated remaining lifetime of the new unit i based on the degradation signal of sensor j. Let u_j^d and v_j^d be the mean and the variance of the failure threshold l_j, whose values can be estimated by the sample mean and sample variance based on the last observations before failure in all historical units for sensor j, respectively. Since $L_{i,j,n_i+t} = \theta_{i,j}^{(0)} + \cdots + \theta_{i,j}^{(p)} * (n_i + t)^p + \varepsilon_{i,j,n_i+t}$, L_{i,j,n_i+t} is normally distributed with mean $\tilde{u}_{i,j,n_i+t} = [1, \dots, (n_i + t)^p]\boldsymbol{u}_j^1$ and variance $\tilde{\sigma}_{i,j,n_i+t}^2 = [1, \dots, (n_i + t)^p]\boldsymbol{\Sigma}_j^1[1, \dots, (n_i + t)^p]' + \sigma_j^2$. Thus, the conditional cumulative distribution function (cdf) of the estimated remaining lifetime $\tilde{T}_{i,j}$ given the collected *in-situ* sensor information $\boldsymbol{L}_{i,j,.}$ can be calculated as:

$$F_{\tilde{T}_{i,j}|\boldsymbol{L}_{i,j,\cdot}}(t) = P\left(\tilde{T}_{i,j} \le t \middle| \boldsymbol{L}_{i,j,\cdot}\right) = P\left(L_{i,j,n_i+t} \ge l_j \middle| \boldsymbol{L}_{i,j,\cdot}\right) = \Phi\left(\frac{\tilde{u}_{i,j,n_i+t} - u_j^d}{\sqrt{\tilde{\sigma}_{i,j,n_i+t}^2 + v_j^d}}\right) = \Phi\big(g(t)\big). \quad (3.7)$$

Here, $\Phi(\cdot)$ represents the cdf of the standard normal distribution. Given that $\tilde{T}_{i,j} \ge 0$, the truncated cdf for $\tilde{T}_{i,k}$ conditioning on $\tilde{T}_{i,k} \ge 0$ is:

$$P\left(\tilde{T}_{i,j} \le t \middle| \boldsymbol{L}_{i,j,\cdot}, \tilde{T}_{i,j} \ge 0\right) = \frac{P\left(0 \le \tilde{T}_{i,j} \le t \middle| \boldsymbol{L}_{i,j,\cdot}\right)}{P\left(\tilde{T}_{i,j} \ge 0 \middle| \boldsymbol{L}_{i,j,\cdot}\right)} =$$

$$\frac{\Phi(g(t)) - \Phi(g(0))}{1 - \Phi(g(0))}. \quad (3.8)$$

Since the truncated cdf is skewed, we need to use the median as the point estimator for remaining life prediction. Numerically, this can be done by finding the time point t when $P\left(\tilde{T}_{i,j} \le t \middle| \boldsymbol{L}_{i,j,\cdot}, \tilde{T}_{i,j} \ge 0\right) = 0.5$.

Define the percentage error, $err_{i,j}$ as the relative difference between the predicted and the actual failure time for unit i and sensor j:

$$err_{i,j} = \frac{(n_i + \tilde{T}_{i,j}) - (n_i + T_i)}{n_i + T_i} = \frac{\tilde{T}_{i,j} - T_i}{n_i + T_i}, \quad (3.9)$$

where n_i is the number of available observations of testing unit i when it stops further usage at some point prior to failure; T_i is the actual remaining lifetime for testing unit i; and $\tilde{T}_{i,j}$ is the estimated remaining lifetime for unit i by using sensor j. Fig. 5 shows the comparison results by using the best single sensor signal from all 11 selected sensors, HI-non, HI-semi and HI-SNR at different levels of actual remaining lifetime. For example, the "all" label refers to the percentage errors based on all 100 testing units, while the "100" label refers to the percentage errors based on only the testing units with equal to or less than 100 actual remaining lifetime.

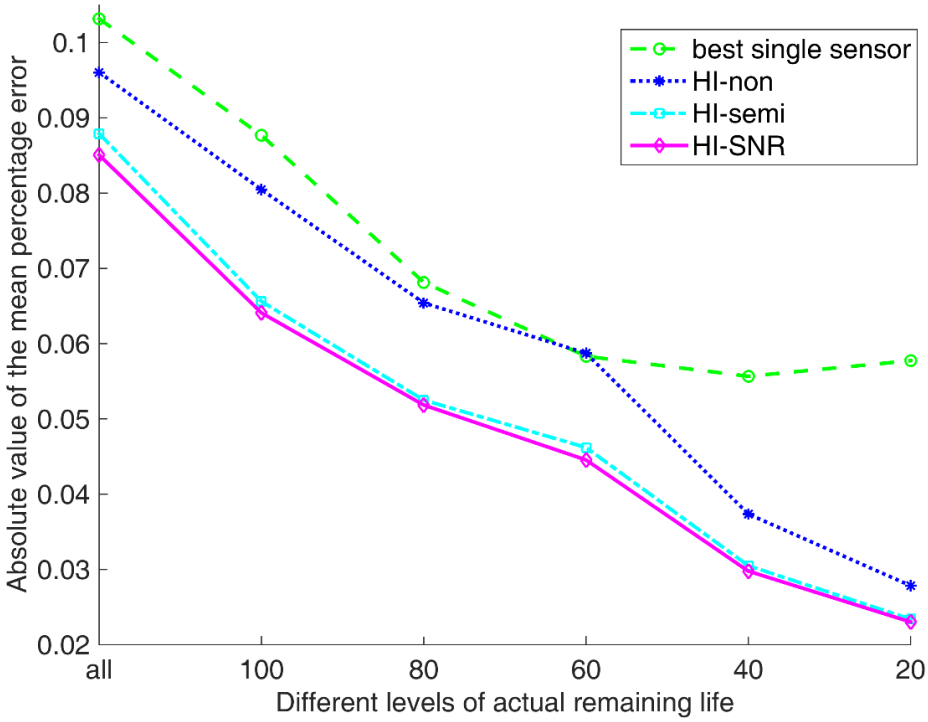

Fig. 3.4. Comparison results of the absolute value of the mean percentage error by using the best single sensor signal, HI-non, HI-semi, and HI-SNR at different levels of the actual remaining lifetime.

From Fig. 3.4, it shows that (i) the health index constructed by the SNR-based data fusion model provides the best prognostic result; and (ii) the remaining life prediction becomes more accurate by using any of the health indices constructed from the three data fusion models as the unit approaches the actual failure time. The first observation (i) stems from the model efforts in (6) that maximize the defined SNR metric to construct a health index with an optimized signal quality. In this particular case study, the SNR-based data fusion model shows a comparable result as the semi-parametric data fusion model. This is probably because the range information of the selected 11 sensor signals is similar to each other as shown in Fig. 3.2, and thus the proposed SNR-based data fusion model only provides a health index with a slightly better SNR value than the semi-parametric model as shown in Table 3.3. Such a result further validates that the defined SNR metric

is reliable to be used for characterizing the signal quality of the degradation-based sensor data. However, in general cases, when the range information of different sensors is quite distinct to each other, the proposed SNR-based data fusion model is expected to show more advantages than the semi-parametric model. We have also evaluated the 95% confidence interval (CI) of the predicted remaining lifetime in all testing units and achieved a similar result as in Fig. 3.4 that the narrowest CIs are given by the HI-SNR and HI-semi, and then followed by HI-non. However, due to the page limit, the result of the CIs is omitted here.

The second observation (ii) mainly results from two reasons. First, less actual remaining lifetime often indicates that more historical data have been acquired, and thus we are more confident about the updated degradation models. Meanwhile, as the actual remaining lifetime becomes smaller, predictions are only made for a shorter future time period, and thus less uncertainty is involved. Second, recall that we used an arithmetic series for $\{c_{i,t}\}$ by assuming the weight coefficient $c_{i,t}$ linearly increases with the residual term $e_{i,t}$. Thus, more penalties are given to the model fitting errors at the observation epochs that are closer to the actual failure time. As a result, the remaining life predictions by using the data fusion methods become more accurate as the unit approaches to failure.

TABLE 3.4. PERFORMANCE COMPARISONS BY USING THE EXISTING AND THE PROPOSED DATA FUSION MODELS ('1' FOR THE BEST AND '3' FOR THE WORST).

Method	Non-parametric	Semi-parametric	SNR-based
Prognostic Performance	3	2	1
Scale Invariant	No	No	Yes
Resistant to missing data	No	Yes	Yes
Computational Cost	3	1	1
Signal quality measure	No	No	Yes

Table 3.4 further summarizes the performance comparisons by using the existing and the proposed data fusion models.

3.5 Conclusions

The contributions of the Chapter is twofold: (i) it fills the literature gap by identifying a quantitative measure, the signal-to-noise ratio (SNR) metric tailored to the needs of degradation signals; and (ii) it develops a systematic data-level fusion model that combines the degradation-based signals from multiple sensors with the goal of maximizing the defined SNR metric. Such optimization efforts result in a composite health index that better characterizes the health condition of the degraded unit and thus leads to an improved remaining life prediction. The developed data fusion method was tested and validated by using the degradation dataset of aircraft gas turbine engines that were generated by C-MAPSS [32]. Our experimental studies showed that the developed health index outperformed each original sensor data and the health indices constructed through other existing data-level fusion methods. In addition, the robustness analysis showed that the proposed data fusion method can allow certain level of missing data without sacrificing too much on the accuracy of the derived optimal solutions. Furthermore, it is worth mentioning that the constructed health index can be regarded as an additional sensor data which can be integrated with other decision-level fusion models to further enhance prognostic performance.

3.6 Appendix

The variance in the failure threshold of the health index can be expressed as: $\boldsymbol{w}'\boldsymbol{L}'_{\cdot,\cdot,n_i}\boldsymbol{D}\boldsymbol{L}_{\cdot,\cdot,n_i}\boldsymbol{w}$. The proof has been given in [10] and here we provide the derivation again for convenience. Denote

$\mathbf{1}$ as the column vector with all ones, $\boldsymbol{I}$ as the identity matrix and $\boldsymbol{D} = \frac{1}{m-1} * (\boldsymbol{I} - \frac{\mathbf{11'}}{m})$. The unbiased estimator of the variance in the failure threshold is: $\sum_{i=1}^{m}\left(h_{i,n_i} - \bar{h}_{.,n_i}\right)^2/(m-1) = \frac{\left(\boldsymbol{L}_{.,.,n_i}\boldsymbol{w}\right)'\left(\boldsymbol{L}_{.,.,n_i}\boldsymbol{w}\right) - m\left(\frac{\mathbf{1'}\boldsymbol{L}_{.,.,n_i}\boldsymbol{w}}{m}\right)^2}{m-1} = \frac{\boldsymbol{w'L'}_{.,.,n_i}\boldsymbol{L}_{.,.,n_i}\boldsymbol{w} - \frac{\boldsymbol{w'L'}_{.,.,n_i}\mathbf{11'}\boldsymbol{L}_{.,.,n_i}\boldsymbol{w}}{m}}{m-1} = \boldsymbol{w'L'}_{.,.,n_i}\left(\frac{\boldsymbol{I} - \frac{\mathbf{11'}}{m}}{m-1}\right)\boldsymbol{L}_{.,.,n_i}\boldsymbol{w} =$

$\boldsymbol{w'L'}_{.,.,n_i}\boldsymbol{DL}_{.,.,n_i}\boldsymbol{w}$.

Here, we prove the optimal solution $\boldsymbol{w}^*$ that maximizes SNR^d in (3.5) is the eigenvector corresponding to the largest eigenvalue λ_{max} by solving the generalized eigenvalue problem $\boldsymbol{Aw} = \lambda_{max}\boldsymbol{Bw}$. The optimization problem in (3.5) is equivalent to

$$max_{\boldsymbol{w}}\, \boldsymbol{w'Aw}, \text{ s.t. } \boldsymbol{w'Bw} = K, \tag{3.10}$$

where K is a constant number. Define the Lagrangian to be $L = \boldsymbol{w'Aw} - \lambda(\boldsymbol{w'Bw} - K)$. By taking the derivative with respect to $\boldsymbol{w}$, we obtain the solution:

$$\boldsymbol{Aw} = \lambda\boldsymbol{Bw}. \tag{3.11}$$

This is a generalized eigenvalue problem and λ is the eigenvalue. To show that the optimal value of (3.10) is achieved when λ equals to the largest eigenvalue, we multiple $\boldsymbol{w'}$ on both sides of (3.11), $\boldsymbol{w'Aw} = \lambda\boldsymbol{w'Bw} = \lambda K$. Consequently, the optimal solution of (3.10) is the generalized eigenvector corresponding to the largest eigenvalue λ_{max} by solving the generalized eigenvalue problem $\boldsymbol{Aw} = \lambda_{max}\boldsymbol{Bw}$.

Here, we show that both $\boldsymbol{B}$ and $\boldsymbol{A}$ are symmetric positive semidefinite (P.S.D.) matrices. First, it is straightforward to see that $\boldsymbol{D} = \frac{\boldsymbol{I} - \boldsymbol{O}/m}{m-1}$ is a symmetric P.S.D matrix. Thus, according to the Cholesky decomposition, $\boldsymbol{D}$ can be written as $\boldsymbol{D} = \boldsymbol{GG'}$, in which $\boldsymbol{G}$ is a lower triangular matrix.

As a result, $\boldsymbol{L}'_{\cdot,\cdot,n_i}\boldsymbol{D}\boldsymbol{L}_{\cdot,\cdot,n_i} = \left(\boldsymbol{G}'\boldsymbol{L}_{\cdot,\cdot,n_i}\right)'\boldsymbol{G}'\boldsymbol{L}_{\cdot,\cdot,n_i}$ is a P.S.D. matrix. Similarly, $\boldsymbol{A} = \boldsymbol{Y_0}'\boldsymbol{1}\boldsymbol{1}'\boldsymbol{Y_0}$ is also a P.S.D. matrix. Next, as $\boldsymbol{H}$ is the projection matrix, $\left((\boldsymbol{I}-\boldsymbol{H})\boldsymbol{c}_{i,\cdot}\boldsymbol{L}_{i,\cdot,\cdot}\right)'(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{c}_{i,\cdot}\boldsymbol{L}_{i,\cdot,\cdot} = \boldsymbol{L}'_{i,\cdot,\cdot}\boldsymbol{c}_{i,\cdot}(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{c}_{i,\cdot}\boldsymbol{L}_{i,\cdot,\cdot}$ is also P.S.D. Since the sum of P.S.D. matrices is also P.S.D., $\boldsymbol{B} = \boldsymbol{L}'_{\cdot,\cdot,n_i}\boldsymbol{D}\boldsymbol{L}_{\cdot,\cdot,n_i} + \sum_{i=1}^{m}\boldsymbol{L}'_{i,\cdot,\cdot}\boldsymbol{c}_{i,\cdot}(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{c}_{i,\cdot}\boldsymbol{L}_{i,\cdot,\cdot}$ is a P.S.D. matrix. In addition, as $\boldsymbol{B} = \boldsymbol{B}'$ and $\boldsymbol{A} = \boldsymbol{A}'$, both matrices $\boldsymbol{B}$ and $\boldsymbol{A}$ are symmetric and P.S.D.

Here, we show that the posterior distribution of $\boldsymbol{\Gamma}_{i,j}$ still follows a multivariate normal distribution with mean $\boldsymbol{u}_j^1 = \left(\frac{\boldsymbol{\Psi}_i{}'\boldsymbol{\Psi}_i}{\sigma_j^2} + \left(\boldsymbol{\Sigma}_j^0\right)^{-1}\right)^{-1}\left(\frac{\boldsymbol{\Psi}_i{}'\boldsymbol{L}_{i,j,\cdot}}{\sigma_j^2} + \left(\boldsymbol{\Sigma}_j^0\right)^{-1}\boldsymbol{u}_j^0\right)$ and variance $\boldsymbol{\Sigma}_j^1 = \left(\frac{\boldsymbol{\Psi}_i{}'\boldsymbol{\Psi}_i}{\sigma_j^2} + (\boldsymbol{\Sigma}_j^0)^{-1}\right)^{-1}$. The proof has been given in [10] and here we provide the derivation again for convenience.

The probability density function of $\boldsymbol{\Gamma}_{i,j}$ can be expressed as $P\left(\boldsymbol{\Gamma}_{i,j}\right) = \frac{1}{(2\pi)^{3/2}|\boldsymbol{\Sigma}_j^0|^{1/2}}e^{-\frac{1}{2}(\boldsymbol{\Gamma}_{i,j}-\boldsymbol{u}_j^0)'(\boldsymbol{\Sigma}_j^0)^{-1}(\boldsymbol{\Gamma}_{i,j}-\boldsymbol{u}_j^0)}$. Thus,

$$P\left(\boldsymbol{\Gamma}_{i,j}\middle|\boldsymbol{L}_{i,j,\cdot}\right) \propto P\left(\boldsymbol{L}_{i,j,\cdot}\middle|\boldsymbol{\Gamma}_{i,j}\right)P\left(\boldsymbol{\Gamma}_{i,j}\right) \propto e^{-\frac{1}{2\sigma_j^2}\left(\boldsymbol{L}_{i,j,\cdot}-\boldsymbol{\Psi}_i\boldsymbol{\Gamma}_{i,j}\right)'\left(\boldsymbol{L}_{i,j,\cdot}-\boldsymbol{\Psi}_i\boldsymbol{\Gamma}_{i,j}\right)}e^{-\frac{1}{2}\left(\boldsymbol{\Gamma}_{i,j}-\boldsymbol{u}_j^0\right)'\left(\boldsymbol{\Sigma}_j^0\right)^{-1}\left(\boldsymbol{\Gamma}_{i,j}-\boldsymbol{u}_j^0\right)}$$

$$\propto e^{\boldsymbol{\Gamma}'_{i,j}\left(\frac{\boldsymbol{\Psi}_i{}'\boldsymbol{\Psi}_i}{\sigma_j^2}+\left(\boldsymbol{\Sigma}_j^0\right)^{-1}\right)\boldsymbol{\Gamma}_{i,j}-2\left(\frac{\boldsymbol{L}'_{i,j,\cdot}\boldsymbol{\Psi}_i}{\sigma_j^2}+\left(\boldsymbol{u}_j^0\right)'\left(\boldsymbol{\Sigma}_j^0\right)^{-1}\right)\boldsymbol{\Gamma}_{i,j}}.$$

It is known that the Gaussian family is the *conjugate prior* for the Gaussian likelihood function, and thus the posterior distribution also follows a multivariate normal distribution. As a result, $\boldsymbol{\Gamma}_{i,j}|\boldsymbol{L}_{i,j,\cdot}\sim N_3(\boldsymbol{u}_j^1, \boldsymbol{\Sigma}_j^1)$and $P\left(\boldsymbol{\Gamma}_{i,j}\middle|\boldsymbol{L}_{i,j,\cdot}\right) \propto e^{-\frac{1}{2}(\boldsymbol{\Gamma}_{i,j}-\boldsymbol{u}_j^1)'(\boldsymbol{\Sigma}_j^1)^{-1}(\boldsymbol{\Gamma}_{i,j}-\boldsymbol{u}_j^1)}$.

Comparing the two equations for $P(\boldsymbol{\Gamma}_{i,j}|\boldsymbol{L}_{i,j,.})$, we get the results $(\boldsymbol{\Sigma}_j^1)^{-1} = \left(\frac{\boldsymbol{\Psi}_i{}'\boldsymbol{\Psi}_i}{\sigma_j^2} + \left(\boldsymbol{\Sigma}_j^0\right)^{-1}\right)$ and $(\boldsymbol{u}_j^1)'(\boldsymbol{\Sigma}_j^1)^{-1} = \left(\frac{\boldsymbol{L}_{i,j,.}'\boldsymbol{\Psi}_i}{\sigma_j^2} + \left(\boldsymbol{u}_j^0\right)'\left(\boldsymbol{\Sigma}_j^0\right)^{-1}\right)$. This finishes the proof that $\boldsymbol{\Gamma}_{i,j}|\boldsymbol{L}_{i,j,.} \sim N_3(\boldsymbol{u}_j^1, \boldsymbol{\Sigma}_j^1)$, where $\boldsymbol{\Sigma}_j^1 = \left(\frac{\boldsymbol{\Psi}_i{}'\boldsymbol{\Psi}_i}{\sigma_j^2} + \left(\boldsymbol{\Sigma}_j^0\right)^{-1}\right)^{-1}$ and $\boldsymbol{u}_j^1 = \left(\frac{\boldsymbol{\Psi}_i{}'\boldsymbol{\Psi}_i}{\sigma_j^2} + \left(\boldsymbol{\Sigma}_j^0\right)^{-1}\right)^{-1} (\frac{\boldsymbol{\Psi}_i{}'\boldsymbol{L}_{i,j,.}}{\sigma_j^2} + \left(\boldsymbol{\Sigma}_j^0\right)^{-1}\boldsymbol{u}_j^0)$.

Chapter 4

A Data-level Fusion Approach for Degradation Modeling and Prognostic Analysis Under Multiple Failure Modes

4.1 Introduction

As mentioned in Chapters 1 and 2, different failure modes may have distinct influences on the degradation path of a unit. To illustrate this point, Fig. 4.1 displays the degradation profiles of two

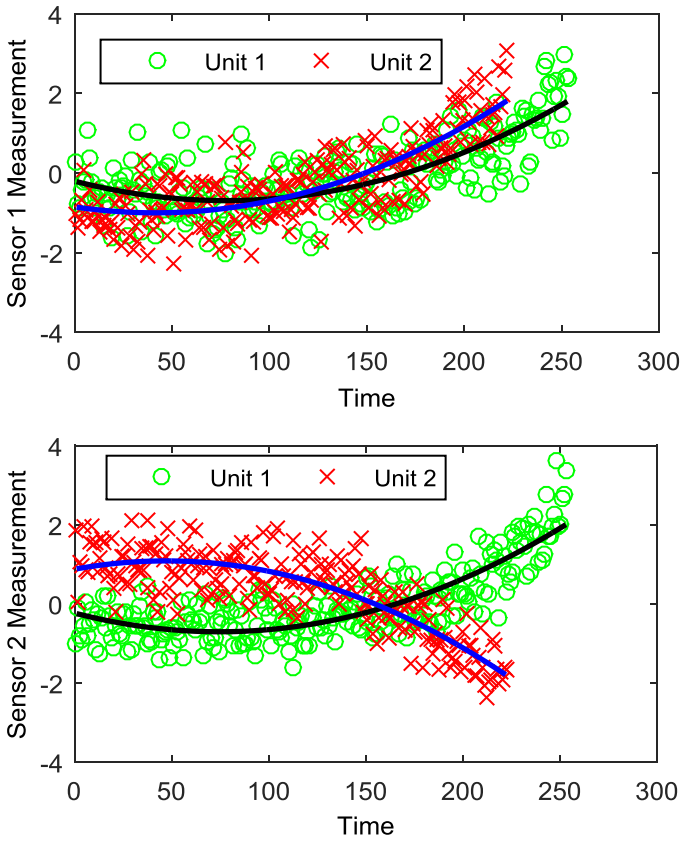

Fig. 4.1. An example of the degradation profiles of two sensors for two units. The circle marks correspond to the sensor measurements of unit 1 that fails due to failure mode 1, whereas the cross marks correspond to the sensor measurements of unit 2 that fails due to failure mode -1. The lines represent a second order polynomial fit for the sensor measurements.

different sensors for two units that fail due to two different failure modes. It clearly shows that the sensor measurements tend to be highly correlated, and some sensors are more sensitive to certain types of failure modes (e.g., sensor 2) while others (e.g., sensor 1) are not. This example thus motivates us to consider data-level fusion approaches that take advantage of the multiple and dependent sensor information to better infer the failure mode of a unit along its service lifecycle. If such failure mode information is known, then it can be readily integrated with the existing data-level fusion models for better degradation modeling and prognostics.

To the best of our knowledge, the existing literature still lacks a generic data fusion approach that can be effectively used for degradation modeling and prognostics when the challenges of multiple sensor data and multiple failure modes simultaneously occur. Thus, this Chapter seeks to fill this literature gap by developing a data-level fusion methodology via the combination of multiple sensor data under multiple failure modes. Specifically, our proposed method will first construct a failure mode index, named FM-INDEX, which allows for the online estimation of the probability that a unit belongs to a certain failure mode. Once developed, the FM-INDEX can be integrated with existing data-level fusion approaches (e.g., Liu *et al.* (2015), Liu and Huang (2016)) to address the challenges of degradation modeling and prognostics when there are multiple failure modes. Fig. 4.2 shows a schematic diagram of the proposed data fusion approach.

The rest of the Chapter is structured as follows: Section 4.2 first introduces the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) that was developed by NASA [32]. Then, reviews a degradation dataset in [27], which was generated by C-MAPSS to simulate the degradation process of aircraft gas turbine engines. Section 4.3 proposes a data-level fusion method for failure mode diagnosis during condition monitoring, which is tailored to the needs of

degradation modeling and prognostics. Section 4.4 discusses the construction of a health-index. Section 4.5 shows the procedure of estimating the remaining useful life under multiple failure modes. Section 4.6 demonstrates the effectiveness of the proposed method in terms of its failure mode diagnosis, degradation modeling, and remaining life prediction based on the dataset introduced in Section 4.2. Section 4.7 draws a conclusion and discusses future research directions.

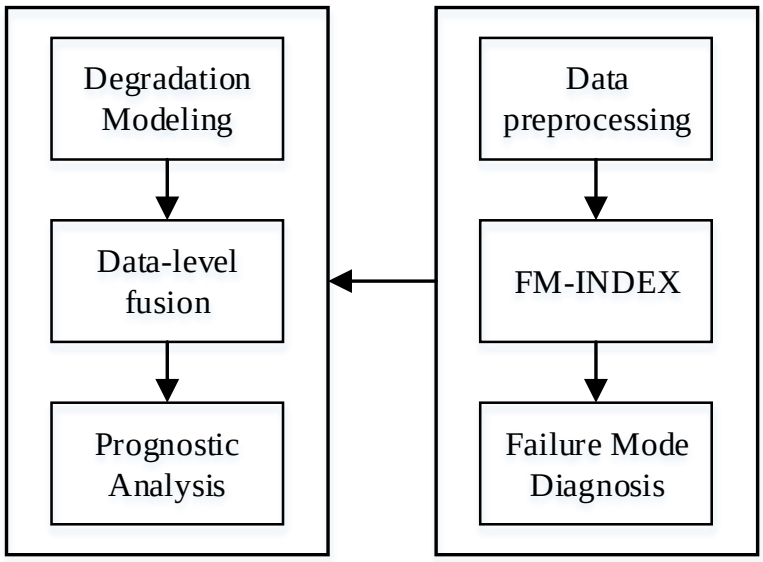

Fig. 4.2. A flow chart of the proposed data-level fusion approach for degradation modeling and prognostics when there are multiple failure modes.

4.2 Dataset Overview

C-MAPSS is a tool for simulating realistic large commercial turbofan engines (Frederick *et al.* (2007)), which has been widely used in the literature of condition monitoring and degradation analysis (Ramasso and Saxena (2014)). Fig. 4.3 shows a schematic diagram of the aircraft gas turbine engine that is simulated by using C-MAPSS.

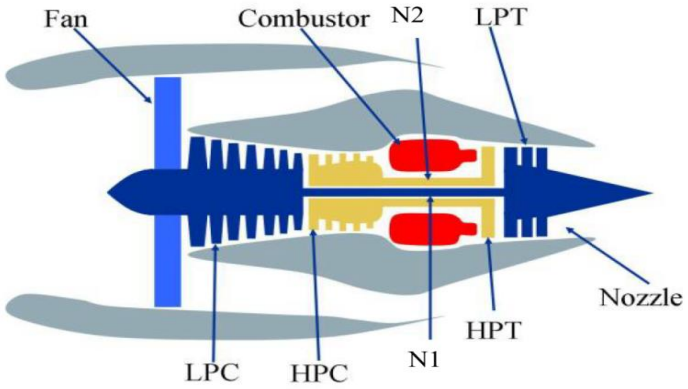

Fig. 4.3. Simplified Engine Diagram simulated by C-MAPSS (Saxena and Goebel *et al.* (2008)).

This simulator generates and records 21 sensor signals of aircraft engines in their service lifecycles. The description of the 21 sensors is given in Table 4.1. Furthermore, each operating unit is subject to one of two potential failure modes: the failure is either due to the high pressure compressor (HPC) or the fan of the engine. Specifically, each engine starts with an unknown initial wear condition, and proceeds to degrade until failure. Due to the complexity of the engine system and the dynamics of the degradation process, the damage profile of an engine cannot be directly inferred based on the flight duration or physical models. Also, the simulation model for generating the dataset is not available to the users. Thus, the underlying assumption here is that users have to solely rely on the multiple sensor data collected from each engine to estimate the current health status and make predictions about the future behavior.

TABLE 4.1. DETAILED SENSOR DESCRIPTIONS [27]

Symbol	Description	Units
T2	Total temperature at fan inlet	°R
T24	Total temperature at LPC outlet	°R
T30	Total temperature at HPC outlet	°R
T50	Total temperature at LPT outlet	°R
P2	Pressure at fan inlet	psia

P15	Total pressure in bypass-duct	psia
P30	Total pressure at HPC outlet	psia
Nf	Physical fan speed	rpm
Nc	Physical core speed	rpm
epr	Engine pressure ratio (P50/P2)	--
Ps30	Static pressure at HPC outlet	psia
phi	Ratio of fuel flow to Ps30	pps/psi
NRf	Corrected fan speed	rpm
NRc	Corrected core speed	rpm
BPR	Bypass Ratio	--
farB	Burner fuel-air ratio	--
htBleed	Bleed Enthalpy	--
Nf_dmd	Demanded fan speed	rpm
PCNfR_dmd	Demanded corrected fan speed	rpm
W31	HPT coolant bleed	lbm/s
W32	LPT coolant bleed	lbm/s

The collected degradation dataset contains 100 training units. To obtain the failure mode for the training units, we leveraged another dataset (called "control group") from the same source (Saxena and Goebel *et al.* (2008)), which contains units that are known to fail only due to the HPC failure. Specifically, we compare the trends (considered to be an increasing/decreasing trend if the initial observation is less/larger than the last observation before failure) of multiple sensor data (e.g., P30, phi, BPR, W31, W32) from these two datasets, and find that there are 44 training units from this dataset consistently showing different trend information compared to the control group. As a result, our preliminary analysis concluded that 56 training units failed due to the fault at the HPC (we denote this failure mode type with label "-1"), and the remaining 44 training units failed due to the fault at the fan (we denote this failure mode type with label "1"). The dataset also contains another 100 testing units with unknown failure modes. Each training unit runs until it fails, while each testing unit stops operation at some points before failure. There is also a file recording the

actual remaining lifetime of the 100 testing units. Thus, the key challenge here is how to accurately estimate the remaining lifetime of these testing units based on the available observations. A more detailed description of the dataset can be found in [27].

4.3 Data-level Fusion for Failure Mode Diagnosis

This sub-section will develop a data-level fusion model for failure mode diagnosis by constructing a composite index, named as FM-INDEX, via the combination of multiple sensor data.

4.3.1 Model Formulation

Let $\boldsymbol{L}_{i,.,t} = [L_{i,1,t}, \dots, L_{i,s,t}] \in \mathbb{R}^{1\times s}$ be the vector of sensor measurements of s sensors for unit i at time t for $t = 1, \cdots, n_i$. Here, $\mathbb{R}$ is the set of real numbers, $L_{i,j,t}$ is the measurement of sensor j for unit i at time t, and n_i is the number of available observations for unit i. In addition, we denote z_i to be the true failure mode that training unit i belongs to, m to be the total number of training units, and $V(f(\cdot), z)$ to be the loss function that quantifies the deviation between the predicted label given the fusion model $f(\cdot)$ and the ground truth label z. To achieve a failure mode diagnosis, one possible idea is to construct a fusion model $f_t(\boldsymbol{L}_{i,.,t})$ that predicts the failure mode of unit i, such that the loss function $V(f_t(\boldsymbol{L}_{i,.,t}), z_i)$ can be minimized over the entire training dataset:

$$\text{obj} = \min_{f_t} \frac{1}{m} \sum_{i=1}^{m} \sum_{t=1}^{n_i} V(f_t(\boldsymbol{L}_{i,.,t}), z_i). \tag{4.1}$$

While this approach is intuitively sound, it produces a fusion model $f_t(\boldsymbol{L}_{i,.,t})$ at each observation time t, which is difficult to implement for new testing units during condition monitoring. As a result, instead we propose the following optimization problem with one fusion model $f(\boldsymbol{L}_{i,.,t})$ that

is independent of the observation time t:

$$\text{obj} = \min_f \frac{1}{m}\sum_{i=1}^{m}\sum_{t=1}^{n_i} a_{i,t} V\left(f\left(\boldsymbol{L}_{i,.,t}\right), z_i\right) + \lambda R(f). \quad (4.2)$$

Here, we define $x_{i,t} = f\left(\boldsymbol{L}_{i,.,t}\right)$ to be the FM-INDEX for unit i at time t. Thus, an ideal case is that the constructed FM-INDEX, $x_{i,t}$, can always retrieve the true failure mode label z_i, i.e., achieving the minimum value in the loss function $V\left(x_{i,t}, z_i\right)$. $R(f)$ is a regularization function to prevent over-fitting; λ is the regularization parameter; and $a_{i,t}$ is called the loss weight for unit i at time t, which is a key parameter of our proposed data fusion model in (4.2). In the next sub-section, we show how to construct the loss weights $\{a_{i,t}\}$, such that the failure mode diagnosis can be tailored to the needs of degradation modeling and prognostics. One of the main advantages of the proposed approach is the flexibility in choosing the loss function. Furthermore, under certain parameters, the proposed method can be effectively transformed to several known and powerful classifiers such as the logistic regression [70] and the Support Vector Machine (SVM) [71]. More details about the loss function and $R(f)$ are discussed in Section 4.3.3.

4.3.2 Setting the Loss Weights $\{a_{i,t}\}$

Denote M_k to be the set of training units that belong to failure mode k. In particular, we consider the following three unique requirements when setting the loss weights $\{a_{i,t}\}$:

(i) Equal importance of the unit from the same failure mode;

(ii) Equal importance for different failure modes; and

(iii) More accurate failure mode diagnostic result as the unit approaches failure.

The requirement (i) means that for each unit that belongs to the same failure mode, their importance should be equally considered in the proposed data fusion model. Thus, the following constraint must be satisfied:

$$\sum_{t=1}^{n_i} a_{i,t} = c_k, \forall i \in M_k, \forall k, \tag{4.3}$$

where c_k denotes a constant number whose value depends on failure mode k.

In practice, as the number of units from different failure modes may not be the same, the requirement (ii) must be also satisfied to deal with this imbalanced classification problem:

$$\sum_{i \in M_k} \sum_{t=1}^{n_i} a_{i,t} = |M_k| * c_k = c, \forall k, \tag{4.4}$$

where c denotes a constant value and $|M_k|$ is the cardinality of the set M_k (i.e., the number of training units that fail under failure mode k). This constraint ensures equal contributions from each failure mode in constructing the FM-INDEX.

Considering that the remaining life prediction becomes more critical as a unit approaches the end of its life, requirement (iii) must be satisfied as well. We impose this constraint by setting $\{a_{i,t}\}$ as a non-decreasing series:

$$a_{i,t+1} \geq a_{i,t}, i = 1, \dots, m, \ t = 1, 2, \dots, \ n_i - 1. \tag{4.5}$$

In this way, more penalty will be given to the loss function as time increases to ensure a more accurate failure mode diagnostic result as the unit approaches the end of life. Specifically, if we assume $a_{i,t}$ follows an arithmetic series $\left(2a_{i,t} = a_{i,t+1} + a_{i,t-1}\right)$, then we can show that the values of the loss weights for training units that fail under failure mode k satisfy the following equation:

$$a_{i,t} = a_{i,1} + (t-1)\frac{\frac{2m}{|M_k|} - 2a_{i,1}n_i}{(n_i - 1)n_i}, t = 1, \dots, n_i, \forall i \in M_k. \tag{4.6}$$

In the above equation, the constant term c_k in (3) is set to be $\frac{m}{|M_k|}$ and the constant term c in (4.4) is set to be m.

4.3.3 Hinge Loss Functions

Many of the existing loss functions (e.g., square loss, logistic loss, hinge loss) can be used in (4.2) to identify the failure mode during condition monitoring [9]. Here, we choose to focus on the hinge loss function as a demonstration. Recall that in Section 4.2, our problem of interest has two potential failure modes, and thus we concentrate on the two failure mode scenarios. A discussion on the extensions to the multiple failure modes is given at the end of this sub-section. In addition, to highlight our main idea, we focus on the linear fusion model as an illustration, i.e., $x_{i,t} = f(\boldsymbol{L}_{i,.,t}) = \boldsymbol{L}_{i,.,t}\boldsymbol{w} + b$. Without loss of generality, such a linear model can be easily extended to nonlinear models by either the creation of nonlinear features or the use of kernel methods [72].

Specifically, the hinge loss function can be expressed as:

$$V(x_{i,t}, z_i) = \xi_{i,t} = \max(0, 1 - z_i x_{i,t}), \tag{4.7}$$

where $\xi_{i,t}$ is called the mis-classification error. Based on the hinge loss function, the failure mode of training unit i is considered to be classified correctly at time t (i.e., with zero loss, $\xi_{i,t} = 0$), if $z_i x_{i,t} \geq 1$; otherwise, unit i is considered to be misclassified at time t with an error $\xi_{i,t} = 1 - z_i x_{i,t}$. It is known that the models solely relying on the hinge loss function are subject to over-fitting. To avoid this problem, a regularization function $R(f) = \|f\|^2 = \boldsymbol{w}^T\boldsymbol{w}$ is usually integrated with the hinge loss function. It can be shown that with the regularized hinge loss function, the proposed data-level fusion model in (4.2) simplifies to the weighted SVM approach [71]:

$$\text{obj} = \min_{\boldsymbol{w},b,\xi_{i,t}} C\sum_{i=1}^{m}\sum_{t=1}^{n_i}\left(a_{i,t} * \xi_{i,t}\right) + 0.5\boldsymbol{w}^T\boldsymbol{w}$$
$$\text{s.t. } z_i x_{i,t} \geq 1 - \xi_{i,t},\ \xi_{i,t} \geq 0,\ i = 1, 2, \ldots, m,\ t = 1,\ 2, \ldots, n_i \tag{4.8}$$

(See Appendix for details).

Here, C is a tuning parameter that controls the relative importance of minimizing the weighted misclassification errors $\sum_{i=1}^{m}\sum_{t=1}^{n_i}\left(a_{i,t} * \xi_{i,t}\right)$ and the regularization term $\boldsymbol{w}^T\boldsymbol{w}$. In practice, the value of C can be chosen by cross validation; and the weighted SVM problem can be solved efficiently by many algorithms, such as the gradient descent methods [73].

The two failure mode scenarios discussed above can also be extended to the multiple failure mode scenarios (see [74] for more details). For example, one possibility is to design multiple FM-INDICES and apply techniques like One-vs.-rest or One-vs.-one strategy [75].

4.3.4 Failure Mode Probability Model

Here, we discuss how to leverage the proposed FM-INDEX, $x_{i,t}$, to estimate the probability of the failure mode. In the existing literature, it is a common practice to use the sigmoid function to transform the outputs of a classification model to a probability function. The technique is called Platt scaling which was proposed by John Platt [76] in the context of support vector machines, and currently this technique is available in a lot of computer software (e.g., MATLAB) and packages. Here, we adopt the sigmoid function $g\left(x_{i,t}\right) = \frac{1}{1+e^{Ax_{i,t}+B}}$ in [76] to develop the failure mode (FM) probability model:

$$P\left(Z_{i,t} = k \middle| x_{i,t}\right) = \begin{cases} \frac{1}{1+e^{Ax_{i,t}+B}}, for\ k = 1 \\ \frac{e^{Ax_{i,t}+B}}{1+e^{Ax_{i,t}+B}}, for\ k = -1 \end{cases}, \tag{4.9}$$

where $Z_{i,t}$ denotes the predicted failure mode of unit i at time t; k is the failure mode label; and the parameters A and B can be estimated by the maximum likelihood estimation approach. Without loss of generality, this proposed FM probability model in (4.9) can be also extended to multiple failure modes (e.g., see [77]).

4.3.5 Real Time Failure Mode Probability Estimation

Previously, we developed an FM-INDEX and a probability estimation of the failure mode for a specific observation. Here, we show how to calculate the failure mode probability for a testing unit based on its available observations. In the reliability-based approach, survival models are often utilized to find the failure time and then competing risk analysis is performed to infer the failure mode probability [78]. However, this approach often focuses on estimating the Remaining Life Distribution (RLD) for the entire population instead of for each individual unit that is still operating in the field. While it is possible to online update the degradation model based on the *in-situ* sensor data and then project it up to the estimated failure threshold to obtain a failure time estimation for each failure model, the derived FM probability estimation depends on the selected failure time window. In fact, it is possible that the estimated failure times of two units from different failure modes exhibit similar results whereas the *in-situ* sensor data of these two units already present different characteristics. Therefore, in this Chapter, we rather choose a different path that predicts the FM probability of a unit in real time directly based on its *in-situ* multiple sensor measurements. Specifically, we first calculate the conditional probability $P\left(Z_{i,t} = k \middle| x_{i,t}\right)$ in (4.9) for unit i at time t. Then, we integrate the probability estimations up to the current time n_i as shown below:

$$P\left(Z_i = k|\boldsymbol{L}_{i,.,.}\right) = P\left(Z_i = k\middle|\boldsymbol{x}_{i,.}\right) = \sum_{t=1}^{n_i} c_{i,t}\, P\left(Z_{i,t} = k\middle|x_{i,t}\right). \quad (4.10)$$

Here, $c_{i,t}$ is the weight coefficient of the probability estimation $P\left(Z_{i,t} = k|x_{i,t}\right)$ for unit i at time t. Similar to the loss weights $\{a_{i,t}\}$ in Section 4.3.2, here we propose to also assign higher weights to the most recent observations:

$$c_{i,t} \geq c_{i,t-1} \geq 0\ , t = 1,2, \dots, n_i - 1. \quad (4.11)$$

In addition, according to the law of probability, the following constraint should be satisfied:

$$\sum_{t=1}^{n_i} c_{i,t} = 1, t = 1,2, \dots, n_i. \quad (4.12)$$

Similar to (4.6), if we assume $c_{i,t}$ follows an arithmetic series $\left(2c_{i,t} = c_{i,t+1} + c_{i,t-1}\right)$, then we can determine the value of $c_{i,t}$ using the following arithmetic series:

$$c_{i,t} = c_{i,1} + (t-1)\frac{2-2c_{i,1}n_i}{(n_i-1)n_i}, t = 1, \dots, n_i. \quad (4.13)$$

Considering that the fault diagnostic result may not be reliable at the early stage of degradation, we can further apply the idea of moving window to the arithmetic series in (4.13). Then, the value of $c_{i,t}$ is calculated by:

$$c_{i,t} = \begin{cases} c_{i,(n_i-r+1)} + (t - n_i + r - 1)\frac{2-2rc_{i,(n_i-r+1)}}{(r-1)r}, t = (n_i - r + 1), \dots, n_i \\ 0, t = 1, \dots, (n_i - r) \end{cases}, \quad (4.14)$$

where r is the moving window size.

4.3.6 Flow Chart of the Data-level Fusion Model for Failure Mode Diagnosis

Fig. 4.4 illustrates the flow chart of the proposed failure mode diagnostic approach. To summarize, we first construct a composite index, named FM-INDEX, which characterizes the failure mode of a unit based on multiple sensor data. Then, with the constructed FM-INDEX, the

FM probability model in (4.9) produces a probability estimation of the failure mode for each observation. Finally, we combine the probability estimations from the available observations of a unit to produce a single failure mode probability estimation for the unit as shown in (4.10).

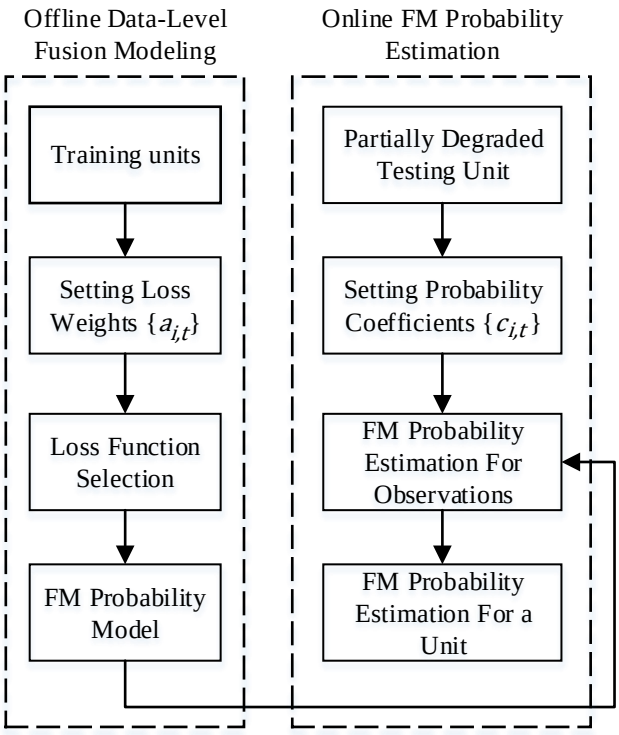

Fig. 4.4 Flow chart of the proposed data-level fusion model for failure mode diagnosis.

4.4 Health Index Construction under Multiple Failure Modes

From the motivation example in Fig. 4.3, we can see that different sensors may contain partial and dependent information about the same unit. In such cases, data fusion methods are often used that take advantage of the sensor data dependencies to better characterize the condition of a unit. For example, Liu and Huang [11] proposed to construct a health index via the combination of

multiple sensor data when there is only one failure mode. They identified two important properties that the health index should possess to ensure successful degradation modeling and prognostics:

Property 1: Given the same failure mode, the variance in the failure threshold of the developed health index should be minimal.

Property 2: Given the same failure mode, the model fitting errors for the developed health index should be minimal.

In particular, Liu and Huang [11] proposed a linear fusion model to construct a health index, $h_{i,t} = \boldsymbol{L}_{i,.,t}\boldsymbol{w}_h$. Here, $h_{i,t}$ is the developed health index for unit i at time t, and $\boldsymbol{w}_h$ is the fusion coefficients for combining the multiple sensor data $\boldsymbol{L}_{i,.,t}$. By optimizing the two identified properties mentioned above, Liu and Huang [11] further showed that the optimal weight $\boldsymbol{w}_h$ can be efficiently solved via a quadratic programing. While such a data-level fusion model has demonstrated a significant improvement for degradation modeling and prognostics, it is limited by the single failure mode assumption. To address this issue, we propose to construct a conditional health index for each failure mode k (i.e., $h_{i,t}^{(k)} = \boldsymbol{L}_{i,.,t}\boldsymbol{w}_h^{(k)}$), which can be further integrated with the developed failure mode diagnostic procedure. Specifically, $\boldsymbol{w}_h^{(k)}$ can be derived by using the proposed method in Liu and Huang [11] based on only the training units that fail under failure mode k. Our goal is that the constructed health index can better characterize the degradation status of a unit than relying on any original sensor data, thus leading to an improved remaining life prediction.

To model the evolution of the degradation process, one of the commonly used methods is the general mixed-effect degradation model proposed by Lu and Meeker [25]:

$$h_t = \eta(\boldsymbol{\varphi}, \boldsymbol{\theta}, t) + \varepsilon_t, \tag{4.15}$$

where $\eta(\cdot)$ is the parametric form of the degradation model; h_t is the health index that measures the underlying degradation status at time t; $\boldsymbol{\varphi}$ is a vector of fixed-effect parameters that represents common characteristics of the population; $\boldsymbol{\theta}$ is a vector of random-effect parameters that characterizes the unit-to-unit variability; and ε_t is an error term that represents the measurement noises. Depending on the parametric form of $\eta(\cdot)$, this degradation model can be used to describe different functional forms according to the evolution of the conditional health index. Without loss of generality, here we focus on the pth order polynomial degradation model as a demonstration:

$$h_{i,t}^{(k)} = \sum_{\alpha=0}^{p} \theta_{i,\alpha}^{(k)} t^{\alpha} + \varepsilon_{i,t}^{(k)} = \boldsymbol{\Gamma}_t \boldsymbol{\theta}_i^{(k)} + \varepsilon_{i,t}^{(k)}, \tag{4.16}$$

where p is the order of the polynomial model; $h_{i,t}^{(k)}$ is the constructed conditional health index for unit i at time t given failure mode k; $\boldsymbol{\theta}_i^{(k)} = \left[\theta_{i,0}^{(k)}, \dots, \theta_{i,p}^{(k)}\right]'$ is the random-effect parameters for unit i given that it degrades due to failure mode k and it is often assumed to follow a multivariate normal distribution, $\boldsymbol{\theta}_i^{(k)} \sim N_{p+1}(\boldsymbol{u}_k^0, \boldsymbol{\Sigma}_k^0)$; $\varepsilon_{i,t}^{(k)}$ is the random noise and follows $N(0, \sigma_k^2)$; and $\boldsymbol{\Gamma}_t = [1, t, \dots, t^p]$. There are two main reasons that we focus on the polynomial degradation model: First, many degradation models discussed in the existing literature can be transformed into (4.16), such as the exponential form models [10], [29], [79] and the random coefficient based models [66], [67]. Second, we implement the Bayesian approach that leverages the *in-situ* multiple sensor data of an operating unit to online update the parameters of the polynomial model. In this way, the updated model becomes more tailored to the unique degradation characteristics of each unit to ensure a good extrapolation performance for prognostics.

4.5 Estimation of the Remaining Useful Life Distribution

In this section, we propose a prognostic method to predict the remaining lifetime of a partially degraded unit in real time when there are multiple failure modes. Specifically, this method integrates the Bayesian updating approach and the failure mode diagnostic procedure proposed in Section 4.3.

For prognostic analysis, a common assumption is that a unit fails once its degradation signal crosses a predefined failure threshold. Consequently, a fundamental requirement here is to accurately predict the future evolution of the degradation signal for a unit up to the failure threshold. However, the central challenge is that the failure mode may affect the degradation path and thus the failure time of a unit. To address this issue, we propose to first offline estimate the values of $\boldsymbol{u}_k^0, \boldsymbol{\Sigma}_k^0$ and σ_k^2 in (4.16) based on the training units that fail under each failure mode k. Then, based on the *in-situ* data $\boldsymbol{L}_{i,:}$ collected from unit i, we calculate the conditional health index $\boldsymbol{h}_{i,\cdot}^{(k)}$ and further calibrate the random-effect parameters $\boldsymbol{\theta}_i^{(k)}$ by using the Bayesian updating approach proposed by Gebraeel [56], conditioning on that unit i fails under failure mode $Z_i = k$. Specifically, for $\boldsymbol{\theta}_i^{(k)} \sim N_{p+1}(\boldsymbol{u}_k^0, \boldsymbol{\Sigma}_k^0)$, we can show that the conditional posterior distribution of $\boldsymbol{\theta}_i^{(k)}$ still follows a multivariate normal distribution given the constructed conditional health index up to the current observation time n_i, $\boldsymbol{h}_{i,\cdot}^{(k)} = \left[h_{i,1}^{(k)}, \ldots, h_{i,n_i}^{(k)}\right]' \in \mathbb{R}^{n_i \times 1}$:

$$\boldsymbol{\theta}_i^{(k)} \,|\boldsymbol{h}_{i,\cdot}^{(k)} \sim N_{p+1}(\boldsymbol{u}_k^1, \boldsymbol{\Sigma}_k^1), \tag{4.17}$$

where $\boldsymbol{u}_k^1 = \left(\frac{\boldsymbol{\Psi}_i{}'\boldsymbol{\Psi}_i}{\sigma_k^2} + (\boldsymbol{\Sigma}_k^0)^{-1}\right)^{-1} (\frac{\boldsymbol{\Psi}_i{}'\boldsymbol{h}_{i,\cdot}^{(k)}}{\sigma_k^2} + (\boldsymbol{\Sigma}_k^0)^{-1}\boldsymbol{u}_k^0)$, $\boldsymbol{\Sigma}_k^1 = \left(\frac{\boldsymbol{\Psi}_i{}'\boldsymbol{\Psi}_i}{\sigma_k^2} + (\boldsymbol{\Sigma}_k^0)^{-1}\right)^{-1}$, and $\boldsymbol{\Psi}_i \in R^{n_i\times(p+1)} = \begin{bmatrix} 1 & \dots & 1 \\ \dots & \dots & \dots \\ 1 & \dots & t^p \\ \dots & \dots & \dots \\ 1 & \dots & n_i^p \end{bmatrix}$ (See Appendix for details).

Here, we consider the normal distribution for the random-effect parameters since it leads to a convenient and analytical solution when updating the random-effect parameters in real time. In fact, this normal assumption for the random-effect parameters has been widely adopted in the existing literature [11], [29], [30], [79]. On the other hand, it is also possible to utilize other distributions or even some simulation-based approaches to compute the posterior distribution and remaining lifetime. In such cases, our proposed ideas for estimating the failure mode can still be applied. Thus, to highlight our main ideas, in the remaining of the Chapter, we only focus on the normal distribution for the random-effect parameters.

Based on the updated random-effect parameters $\boldsymbol{\theta}_i^{(k)} | \boldsymbol{h}_{i,\cdot}^{(k)}$, we then derive the RLD. In particular, since $\tilde{h}_{i,n_i+t}^{(k)} = \theta_{i,0}^{(k)} + \cdots + \theta_{i,p}^{(k)} * (n_i + t)^p + \varepsilon_{i,n_i+t}^{(k)}$ follows a normal distribution with mean $\tilde{u}_{i,n_i+t,k} = [1, \dots, (n_i + t)^p]\boldsymbol{u}_k^1$ and variance $\tilde{\sigma}_{i,n_i+t,k}^2 = [1, \dots, (n_i + t)^p]\boldsymbol{\Sigma}_k^1[1, \dots, (n_i + t)^p]' + \sigma_k^2$, we can calculate the conditional cumulative distribution function (cdf) for the remaining lifetime $\tilde{T}_i$ based on the conditional health index $\boldsymbol{h}_{i,\cdot}^{(k)}$ given that unit i fails under failure mode $Z_i = k$:

$$P\left(\tilde{T}_i \le t \middle| \boldsymbol{L}_{i,\cdot,\cdot}, Z_i = k\right) = P\left(\tilde{T}_i \le t \middle| \boldsymbol{h}_{i,\cdot}^{(k)}\right) \tag{4.18}$$

$$= P\left(\tilde{h}_{i,n_i+t}^{(k)} \geq D_h^{(k)} \middle| \boldsymbol{h}_{i,\cdot}^{(k)}\right) = \Phi\left(\frac{\tilde{u}_{i,n_i+t,k} - u_{h,k}^d}{\sqrt{\tilde{\sigma}_{i,n_i+t,k}^2 + v_{h,k}^d}}\right),$$

where $\Phi(\cdot)$ is the standard normal cdf; and $D_h^{(k)}$ is the failure threshold for the conditional health index under failure mode k and it is assumed to follow a normal distribution with mean and variance $u_{h,k}^d$ and $v_{h,k}^d$, respectively. Their values can be estimated by the sample mean and the sample variance based on the last observations of the training units that fail under failure mode k.

In practice, the failure mode for a unit is often unknown *a priori*. Thus, we integrate (4.18) with the FM probability estimation $P\left(Z_i = k \middle| \boldsymbol{L}_{i,\cdot,\cdot}\right)$ in Section 3.1.5 to predict the RLD $F_{\tilde{T}_i|\boldsymbol{L}_{i,\cdot,\cdot}}(t)$ given the collected *in-situ* sensor data $\boldsymbol{L}_{i,\cdot,\cdot}$:

$$F_{\tilde{T}_i|\boldsymbol{L}_{i,\cdot,\cdot}}(t) = P\left(\tilde{T}_i \leq t \middle| \boldsymbol{L}_{i,\cdot,\cdot}\right) = \sum_{k\in K}\left\{P\left(\tilde{T}_i \leq t \middle| \boldsymbol{L}_{i,\cdot,\cdot}, Z_i = k\right) * P\left(Z_i = k \middle| \boldsymbol{L}_{i,\cdot,\cdot}\right)\right\} =$$

$$\sum_{k\in K}\left\{P\left(\tilde{h}_{i,n_i+t}^{(k)} \geq D_h^{(k)} \middle| \boldsymbol{h}_{i,\cdot}^{(k)}\right) * P\left(Z_i = k \middle| \boldsymbol{x}_{i,\cdot}\right)\right\}, \quad (4.19)$$

where the set K includes all the potential failure modes. Since the remaining lifetime for a testing unit should be greater than 0, we further consider the truncated cdf conditioning on $\tilde{T}_i \geq 0$:

$$P\left(\tilde{T}_i \leq t \middle| \boldsymbol{L}_{i,\cdot,\cdot}, \tilde{T}_i \geq 0\right) = \frac{P\left(0 \leq \tilde{T}_i \leq t \middle| \boldsymbol{L}_{i,\cdot,\cdot}\right)}{P\left(\tilde{T}_i \geq 0 \middle| \boldsymbol{L}_{i,\cdot,\cdot}\right)} =$$

$$\frac{\sum_{k\in K}\left\{P(Z_i = k|\boldsymbol{L}_{i,\cdot,\cdot}) * \left(P\left(\tilde{T}_i \leq t \middle| \boldsymbol{L}_{i,\cdot,\cdot}, Z_i = k\right) - P\left(\tilde{T}_i \leq 0 \middle| \boldsymbol{L}_{i,\cdot,\cdot}, Z_i = k\right)\right)\right\}}{1 - \sum_{k\in K}\left\{P(Z_i = k|\boldsymbol{L}_{i,\cdot,\cdot}) * P\left(\tilde{T}_i \leq 0 \middle| \boldsymbol{L}_{i,\cdot,\cdot}, Z_i = k\right)\right\}}. \quad (4.20)$$

Here, because the truncated cdf in (4.20) is skewed, we can use the median as the point estimator for the remaining lifetime. Numerically, this is equivalent to finding the observation time t such that $P\left(\tilde{T}_i \leq t \middle| \boldsymbol{L}_{i,\cdot,\cdot}, \tilde{T}_i \geq 0\right) = 0.5$.

4.6 Case Study

In this section, we evaluate our proposed data-level fusion method for failure mode diagnosis and prognostic analysis based on the dataset introduced in Section 4.2. In particular, we consider the following two benchmark methods when evaluating the performance of our proposed method: 1) the data-level fusion method in Liu and Huang [11] as introduced in Section 4.4 which ignores the effect of multiple failure modes; and 2) each original sensor data with the consideration of the effect of multiple failure modes.

4.6.1 Data Preprocessing and Sensor Selection

Before implementing the proposed data fusion model, it requires to first determine which sensors to use. To ensure a fair comparison with the existing literature, we adopt a similar sensor selection criterion as in Liu and Huang [11]. They chose to select a sensor if it shows a consistent increasing or decreasing degradation trend among all the training units that fail under the same failure mode. Accordingly, the following 14 sensors (i.e., $s = 14$) from Table 4.1 are selected: T24, T30, T50, P30, Nf, Nc, Ps30, phi, NRf, NRc, BPR, htBleed, W31 and W32. Given that these sensors measure different characteristics of the engine with different measurement units (e.g., psia, rpm, lbm/s), we further take a log transformation and then standardize the dataset before the data fusion as in Liu and Huang [11].

4.6.2 Failure Mode Diagnosis

In this sub-section, we will demonstrate how to construct the proposed FM-INDEX. In addition, we conduct a sensitivity analysis on the weights of the constructed FM-INDEX. Then, we evaluate

and compare the failure mode diagnostic accuracy by using the FM-INDEX and each selected sensor data. Finally, we illustrate how to estimate the probability of the failure mode for a testing unit in real time.

4.6.3 Parameter Setup

In this case study, we choose the value of the loss weight $\{a_{i,t}\}$ by assuming it follows the arithmetic series in (4.6). In addition, we select the hinge loss function in (4.7) with the regularization function $R(f) = \|f\|^2 = \boldsymbol{w}^T\boldsymbol{w}$. By solving the optimization problem in (4.8), the optimal weights $\boldsymbol{w}^*$ and the intercept term b^* for constructing the FM-INDEX are obtained and shown in Table 4.2.

TABLE 4.2. THE OPTIMAL WEIGHTS AND THE INTERCEPT TERM FOR COMBINING THE SELECTED SENSOR DATA WHEN CONSTRUCTING THE FM-INDEX

Name	T24	T30	T50	P30	Nf	Nc	Ps30	phi
Value	-0.2757	-0.2347	-0.4254	2.3447	-1.7481	-1.5456	-0.7011	3.1123
Name	NRf	NRc	BPR	htBleed	W31	W32	intercept	
Value	-1.6992	-1.8486	-2.213	-0.2213	0.4162	0.4461	-2.1222	

Recall that the two failure modes occur either at the HPC or at the fan. From Table 4.2, we can see that the sensors that directly monitor the HPC or the fan tend to exhibit higher weights, which is consistent with our intuition. For example, the sensors with the largest four weights are phi, P30, BPR, NRc. According to the description in Table 4.1, these sensors directly measure the characteristics of the HPC or the fan.

4.6.4 Sensitivity Study

In this sub-section, we conduct a sensitivity analysis on the optimal weights $\boldsymbol{w}^*$ of the FM-INDEX. In particular, we consider the missing data scenario by randomly hiding certain percentage of the observations. Then, we solve the optimization problem in (4.8) to study the sensitivity of the derived optimal weights $\boldsymbol{w}^*$ at different degrees of missing data.

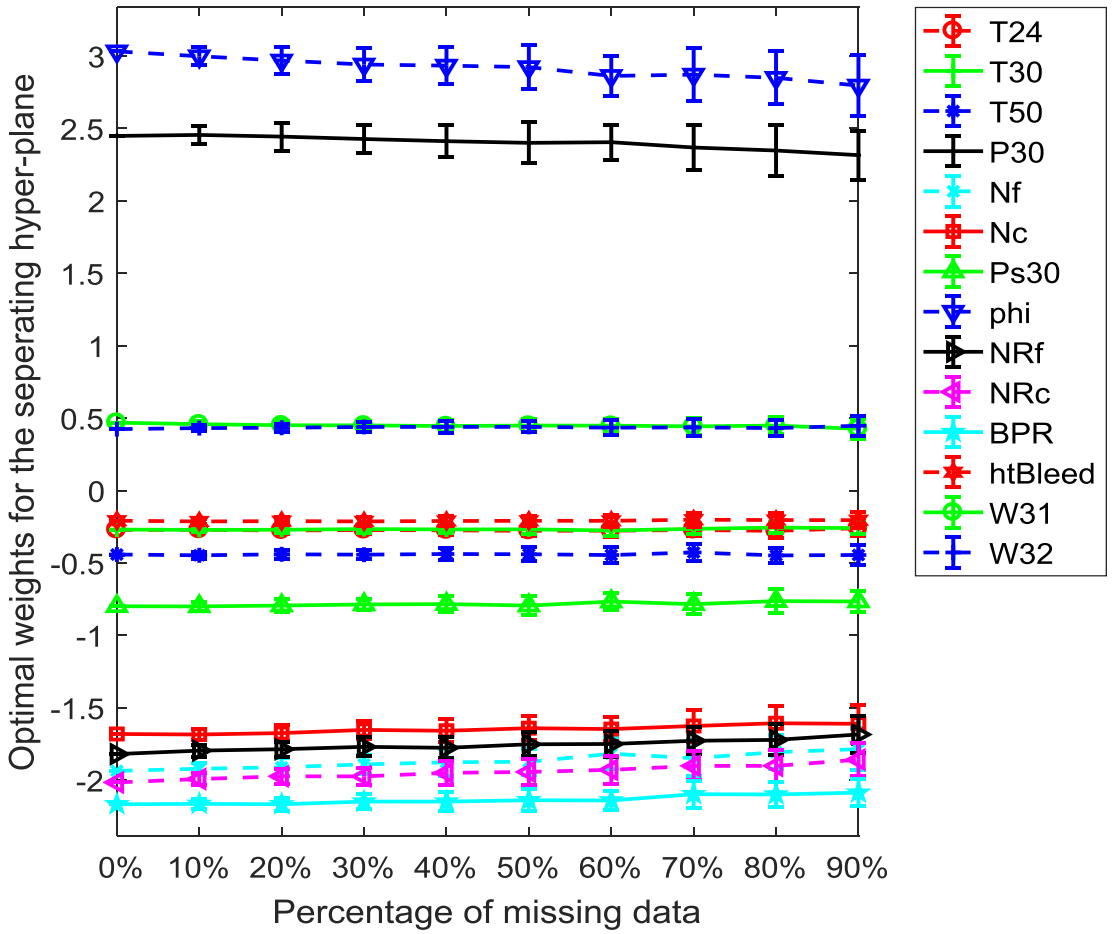

Fig. 4.5. The optimal weights learned at different degrees of missing data with one standard deviation bars.

Fig. 4.5 shows the changes in the optimal weights $\boldsymbol{w}^*$ based on different degrees of missing data with one standard deviation bars. For example, the label "30%" refers to the case where 30%

of the training data is randomly removed prior to learning the optimal weights. Here, the standard deviation bars are obtained by simulating 100 replications with the same percentage of missing data. From Fig. 4.5, we can see that the optimal weights for constructing the FM-INDEX are relatively stable at different degrees of missing data, especially when the missing data is less than or equal to 20%. This indicates that our proposed data-level fusion model allows for certain levels of missing data without great expense to the accuracy of the learned optimal weights.

4.6.5 FM-INDEX Demonstration

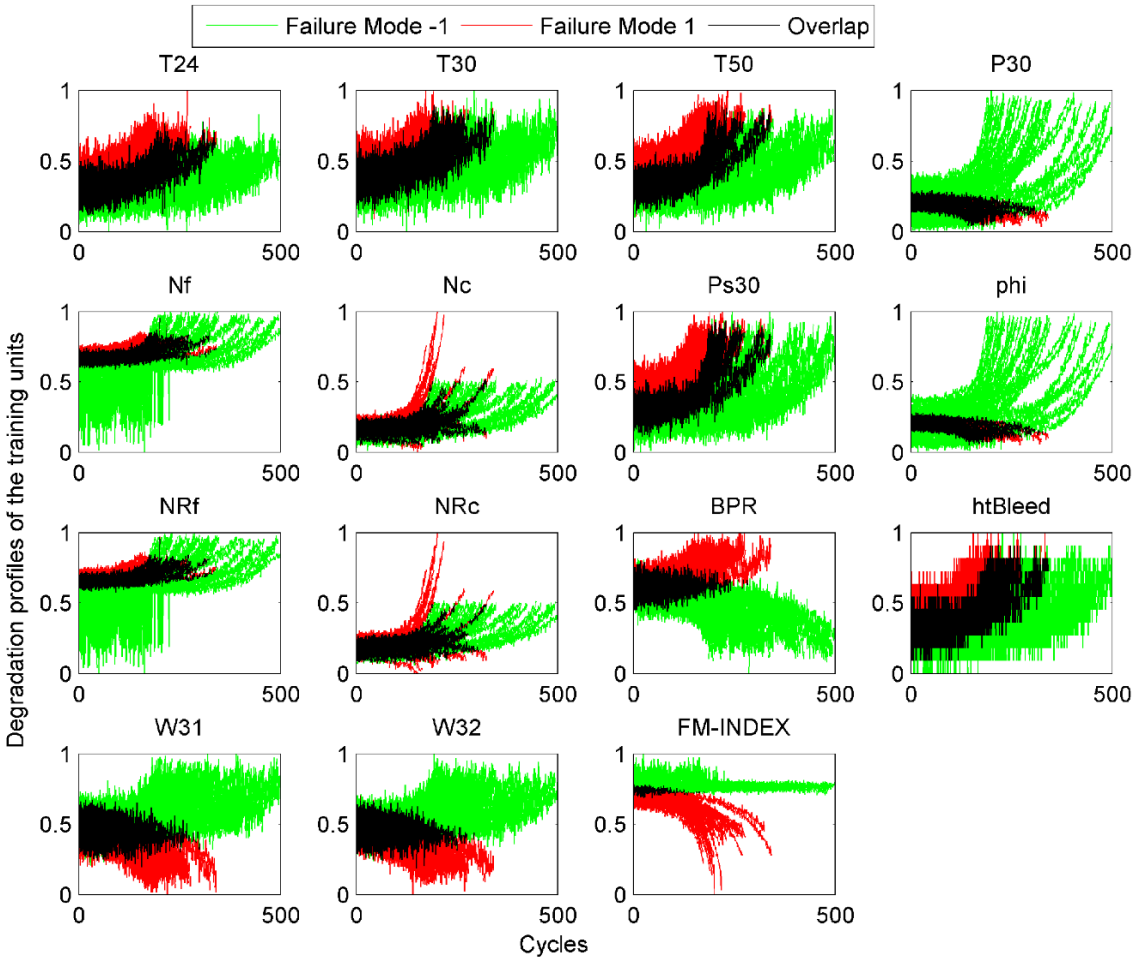

Fig. 4.6. An illustration of the degradation signals of each selected sensor data and the constructed FM-INDEX for the 100 training units.

Fig. 4.6 plots the degradation signals of each selected sensor data and the constructed FM-INDEX for all 100 training units. In particular, the green and the red color shows the units that belong to failure mode 1 and -1, respectively. The overlapping areas, which are highlighted in black, represent the undistinguished regions of the two failure modes.

From Fig. 4.6, we can see that the FM-INDEX better separates the units from these two failure modes than each selected sensor data. This is mainly due to the fact that the FM-INDEX leverages the dependent information from multiple sensor data, and thus it is more informative than the individual sensor data for failure mode diagnosis. In addition, the overlapping areas become much smaller (i.e., failure mode diagnostic result becomes more accurate) as the units approach the end of lives. This observation coincides with our model constraints, in which we consider a non-decreasing series in the key parameter $\{a_{i,t}\}$ when constructing the FM-INDEX.

Furthermore, since the true failure modes of the training units are known, we can evaluate the accuracy of the FM probability estimation over the training units. Specifically, we define the absolute error in the FM probability estimation for training unit i as:

$$p_i^{(err)} = \left|1 - P\left(Z_i = z_i \left| \boldsymbol{L}_{i,,,.}\right.\right)\right|. \tag{4.21}$$

The comparison results by using our proposed FM-INDEX and each selected sensor data are shown in Fig. 4.7, which provides the boxplots of the absolute errors in the FM probability estimation over the training units using a leave-one-out cross-validation technique. The result clearly shows that the FM-INDEX outperforms each individual sensor data for FM probability estimation.

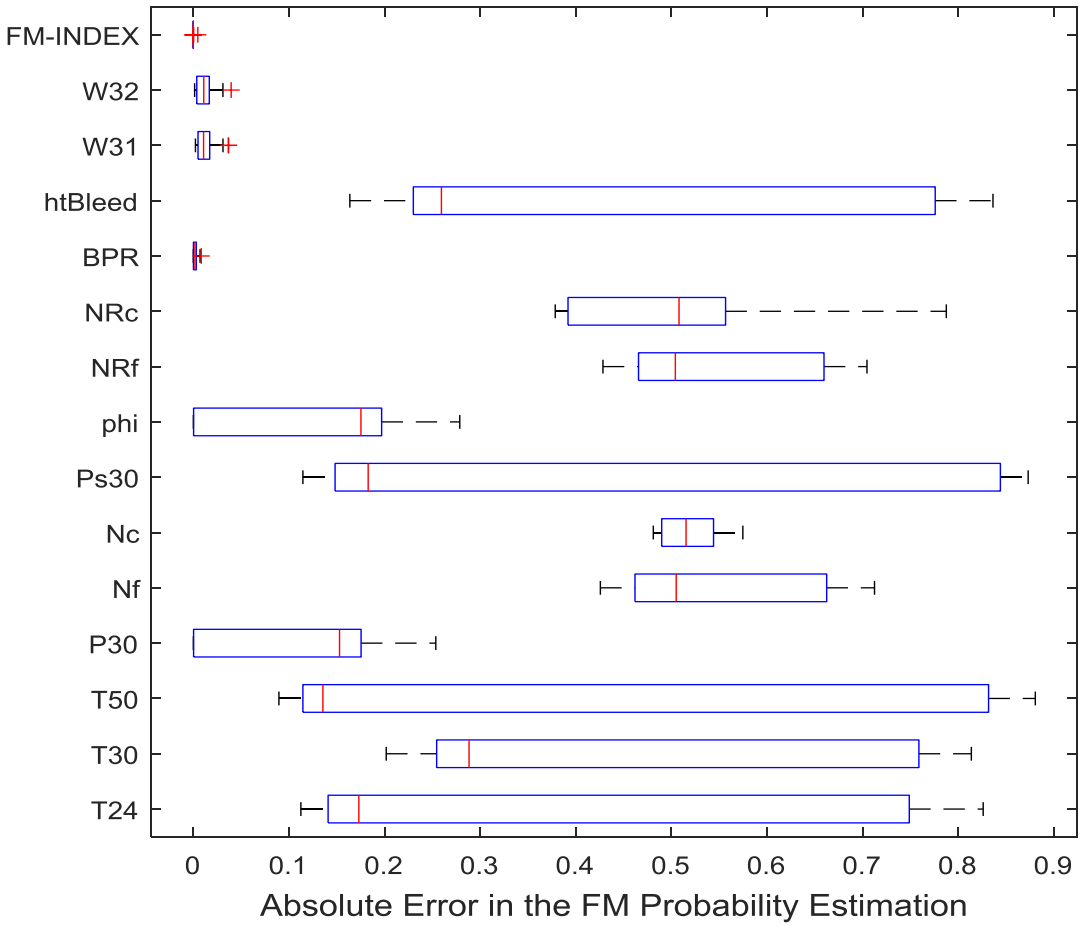

Fig. 4.7. The boxplots of the absolute errors in the FM probability estimation over the training units by using a leave-one-out cross validation.

4.6.6 Real Time Failure Mode Probability Estimation

In this section, we illustrate the failure mode probability estimation for a unit as more data are collected in real time by using the proposed method in Section 4.3.5. In particular, the coefficients $\{c_{i,t}\}$ are chosen by following the arithmetic series in (4.14) with a moving window size $r = 30$. Fig. 8 shows the changes of the probability estimation $P\left(Z_i = -1 \middle| \boldsymbol{L}_{i,\cdot,\cdot}\right)$ for unit i that is known to degrade due to failure mode -1. From Fig. 4.8, we can see that as more observations are collected,

the probability estimation of the failure mode improves and $P(Z_i = -1|\boldsymbol{L}_{i,\dots,})$ approaches 1. Furthermore, since it is difficult to identify the failure modes at an early stage of degradation, the figure shows fluctuations in the probability estimation at the beginning.

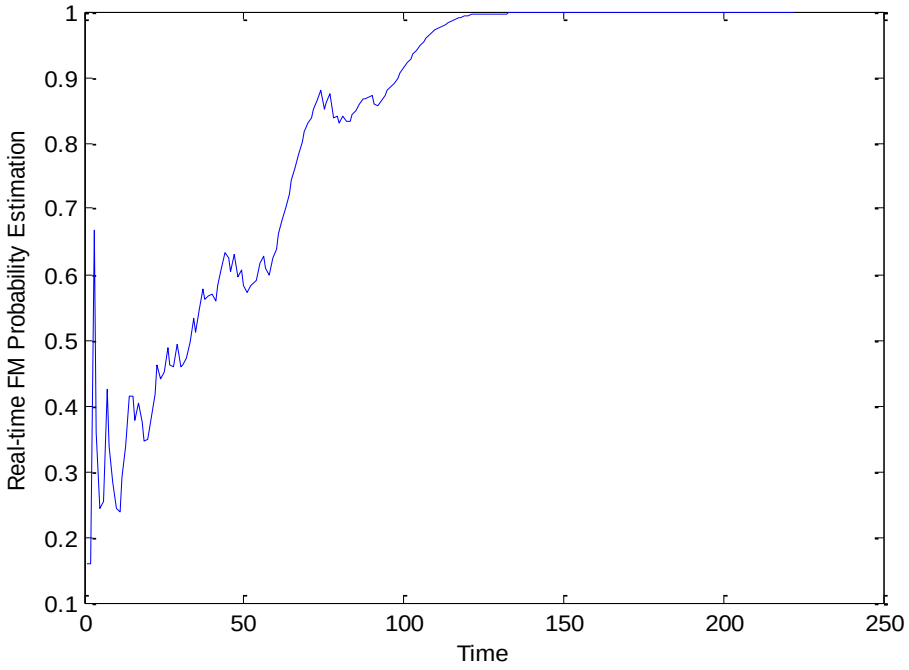

Fig. 4.8. An illustration of the probability estimation $P(Z_i = -1|\boldsymbol{L}_{i,\dots,})$ for a validation unit that fails under failure mode -1 as more data are collected in real time.

4.6.7 Estimation of the RUL Distribution

We estimate the remaining lifetime of the testing units by implementing the proposed method in Section 4.5. Recall that we assume the random-effect parameters follow a multivariate normal distribution. In this case study, we conducted the Henze-Zirkler test to check the normality assumption based on the training dataset for each failure mode. The result showed that for the constructed conditional health index under failure mode 1, we achieved a p-value of 0.49 and for the constructed conditional health index under failure mode 2, we achieved a p-value of 0.15. Thus,

this result further validates that it is satisfactory to adopt the normal distribution for the random-effect parameters.

To evaluate the prognostic performance under the proposed and benchmark methods, we consider the error metric in (4.22), err_i, which is defined as the relative error between the estimated and the actual remaining lifetime of unit i.

$$err_i = \frac{(n_i+\tilde{T}_i)-(n_i+T_i)}{n_i+T_i} = \frac{\tilde{T}_i-T_i}{n_i+T_i}, \tag{4.22}$$

where n_i is the time index of the last observation for testing unit i when it stops further usage prior to failure; T_i is the actual remaining lifetime for unit i; and $\tilde{T}_i$ is the estimated remaining lifetime for unit i.

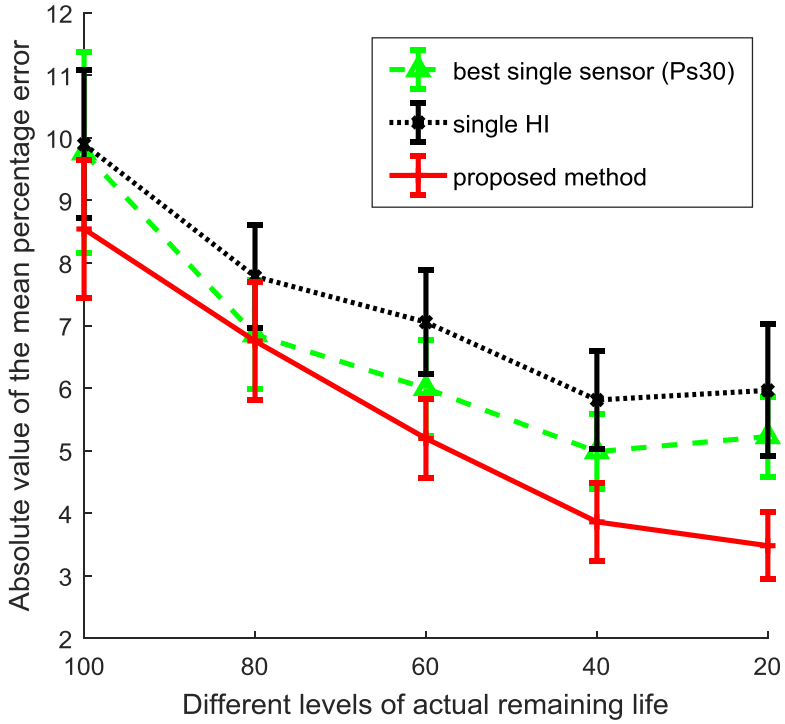

Fig. 4.9. Comparison results of the absolute value of the mean percentage error (%) for the proposed method and the other two benchmark methods. The bars correspond to one standard deviation in the mean percentage error.

Fig. 4.9 shows the comparison results of the absolute value of the mean percentage error (%) for the proposed method, the single HI and the best performing single sensor. In this example, the "100" label refers to the scenarios when only the testing units with actual remaining lifetime equal to or less than 100 are taken into consideration, etc. For the single sensors, we estimate the failure mode and remaining lifetime given only the data from the sensor of interest (i.e., ignoring the data from the other sensors). Based on the plots of the sensor data in Fig. 4.6, we can see that the degradation signals exhibit an exponential functional form. Thus, in the benchmark method, we consider the following degradation model for the log-transformed sensor measurement $L_{i,j,t}$ given the failure mode k:

$$L_{i,j,t} = \theta_{i,j,0}^{(k)} + \theta_{i,j,1}^{(k)} t + \theta_{i,j,2}^{(k)} t^2 + \varepsilon_{i,j,t}^{(k)}, \tag{4.23}$$

where $\varepsilon_{i,j,t}^{(k)}$ is the error term and often assumed to follow the normal distribution, which can be tested based on the fitted degradation model. For the single HI, we adopt the method by Liu and Huang (2016) to estimate the remaining lifetime, which ignores the presence of the multiple failure modes. In addition, to achieve a fair comparison, the random failure threshold approach in (4.18) is considered for all the three methods.

Based on Fig. 4.9, we can see that: (i) our proposed approach outperformed the other two benchmark methods; (ii) in general, the predictions get more accurate as the unit approaches failure in our proposed approach; and (iii) the best single sensor (Ps30) outperforms the single HI constructed by the model in Liu and Huang [11]. The observation (i) mainly results from two reasons. First, our proposed data fusion method takes the advantage of the multiple, dependent

sensor data, and thus it leads to a better remaining life prediction. Second, our proposed approach considers the distinct influence of different failure modes on the lifecycle path of a unit. Regarding the observation (ii), one possible reason is that less actual remaining lifetime typically indicates that more observational data have been collected, and thus we are more confident about the failure mode diagnosis and degradation modeling as the unit approaches the end of its life. In addition, we choose both the loss weights $\{a_{i,t}\}$ and the coefficients $\{c_{i,t}\}$ to be linearly increasing with time, and thus more weight is given to the observations that are closer to failure. However, because only partial degradation signals are provided for the testing units, different sets of testing units are used here to evaluate the prediction error at different levels of actual remaining lifetime; therefore, there is no guarantee to have a monotonically decreasing trend in Fig. 4.9. Finally, the main reason behind the observation (iii) is that the single HI does not consider the presence of two failure modes by assuming all units come from the same failure mode, and thus it leads to a poor prediction performance. This result further demonstrates the importance of our proposed method for online estimation of the failure mode when predicting the RLD.

4.7 Conclusions

With the rapid development of condition monitoring and sensor techniques, multiple sensors have been widely used to simultaneously monitor the degradation status of a unit. In such a big data environment, (i) each sensor may only contain partial and dependent information about the health status of a unit; and (ii) sensors may exhibit distinct characteristic patterns in different units if these units belong to different failure modes.

The main contribution of this Chapter is to address the challenges of degradation modeling and

prognostics in big data environments in presence of multiple sensor data and multiple failure modes. In particular, we construct an FM-INDEX that effectively and accurately diagnoses the failure mode for a unit during condition monitoring, which is tailored to the needs of degradation modeling and prognostic analysis. Next, we extend the existing degradation model and data-level fusion techniques to multiple failure mode scenarios by conditioning on knowing each failure mode, and further integrate with our proposed failure mode diagnostic approach to produce a final remaining life prediction. The developed methodology was tested and validated by using multiple sensor signals from aircraft gas turbine engines that contain two potential failure modes [27]. Our case study showed that: (i) the developed FM-INDEX better distinguishes the units from the two failure modes than each original sensor data; (ii) the failure mode diagnostic result becomes more accurate as a unit approaches failure; and (iii) the remaining life prediction by using the proposed method outperforms the related benchmarks.

There are several important topics for future research: First, it is worth exploring the performance of the developed FM-INDEX by using non-linear fusion functions (e.g., via kernel methods). Second, in this Chapter we derive the FM-INDEX based on historical records offline via a frequentist approach. For future research, it is worth investigating how to online update the FM-INDEX via a Bayesian approach.

4.8 Appendix

Here, we show that our proposed data fusion model $\min_{\boldsymbol{w},b} \frac{1}{m}\sum_{i=1}^{m}\sum_{t=1}^{n_i} a_{i,t} V\big(f\big(\boldsymbol{L}_{i,.,t}\big), z_{i,t}\big) + \lambda R(f)$ with the hinge loss function $V\big(x_{i,t}, z_{i,t}\big) = \max\big(0, 1 - z_{i,t}x_{i,t}\big)$, where $x_{i,t} = f\big(\boldsymbol{L}_{i,.,t}\big) = \boldsymbol{L}_{i,.,t}\boldsymbol{w} + b$ and $R(f) = \|f\|^2 = \boldsymbol{w}^T\boldsymbol{w}$ is equivalent to the weighted SVM approach:

$\min_{\boldsymbol{w},b,\xi_{i,t}} C\sum_{i=1}^{m}\sum_{t=1}^{n_i}\left(a_{i,t} * \xi_{i,t}\right) + 0.5\boldsymbol{w}^T\boldsymbol{w}$, s.t. $z_{i,t}x_{i,t} \geq 1 - \xi_{i,t}$ and $\xi_{i,t} \geq 0, \forall i = 1, 2, \ldots, m,\ t = 1,\ 2, \ldots, n_i$.

Let $\xi_{i,t} = V\left(x_{i,t}, z_{i,t}\right) = \max\left(0, 1 - z_{i,t}x_{i,t}\right)$. Then, $\xi_{i,t} \geq 0$ and $\xi_{i,t} \geq 1 - z_{i,t}x_{i,t}$. Consequently, $\min_{\boldsymbol{w},b}\max\left(0, 1 - z_{i,t}x_{i,t}\right) + \lambda\boldsymbol{w}^T\boldsymbol{w}$ is equivalent to $\min\limits_{\boldsymbol{w},b,\xi_{i,t}} \xi_{i,t} + \lambda\boldsymbol{w}^T\boldsymbol{w}$, subject to that $\xi_{i,t} \geq 0$ and $\xi_{i,t} \geq 1 - z_{i,t}x_{i,t}$.

In this way, $\min_{\boldsymbol{w},b}\frac{1}{m}\sum_{i=1}^{m}\sum_{t=1}^{n_i}\left(a_{i,t} * \max\left(0, 1 - z_{i,t}x_{i,t}\right)\right) + \lambda\boldsymbol{w}^T\boldsymbol{w}$ is equivalent to $\min_{\boldsymbol{w},b,\xi_{i,t}}\frac{1}{m}\sum_{i=1}^{m}\sum_{t=1}^{n_i}\left(a_{i,t} * \xi_{i,t}\right) + \lambda\boldsymbol{w}^T\boldsymbol{w}$, s.t. $z_{i,t}x_{i,t} \geq 1 - \xi_{i,t}$ and $\xi_{i,t} \geq 0$, $\forall i = 1, 2, \ldots, m,\ t = 1,\ 2, \ldots, n_i$. Finally, by scaling the objective function with a factor of $1/(2\lambda)$ and setting $C = 1/(2m\lambda)$, we can rewrite the objective function as $\min_{\boldsymbol{w},b,\xi_{i,t}} C\sum_{i=1}^{m}\sum_{t=1}^{n_i}\left(a_{i,t} * \xi_{i,t}\right) + 0.5\boldsymbol{w}^T\boldsymbol{w}$.

Here, we show that given the failure mode $Z_i = k$, the posterior distribution of $\boldsymbol{\theta}_{i,j}^{(k)}$ still follows a multivariate normal distribution with mean $\boldsymbol{u}_{j,k}^1 = \left(\frac{\boldsymbol{\Psi}_i{}'\boldsymbol{\Psi}_i}{\sigma_{j,k}^2} + (\boldsymbol{\Sigma}_{j,k}^0)^{-1}\right)^{-1}\left(\frac{\boldsymbol{\Psi}_i{}'\boldsymbol{L}_{i,j,\cdot}^{(k)}}{\sigma_{j,k}^2} + (\boldsymbol{\Sigma}_{j,k}^0)^{-1}\boldsymbol{u}_{j,k}^0\right)$ and variance $\boldsymbol{\Sigma}_{j,k}^1 = \left(\frac{\boldsymbol{\Psi}_i{}'\boldsymbol{\Psi}_i}{\sigma_{j,k}^2} + (\boldsymbol{\Sigma}_{j,k}^0)^{-1}\right)^{-1}$.

Given the failure mode $Z_i = k$, the probability density function of $\boldsymbol{\theta}_{i,j}^{(k)}$ can be expressed as $P\left(\boldsymbol{\theta}_{i,j}^{(k)}\right) \propto e^{-\frac{1}{2}(\boldsymbol{\theta}_{i,j}^{(k)} - \boldsymbol{u}_{j,k}^0)'(\boldsymbol{\Sigma}_{j,k}^0)^{-1}(\boldsymbol{\theta}_{i,j}^{(k)} - \boldsymbol{u}_{j,k}^0)}$. Thus,

$P\left(\boldsymbol{\theta}_{i,j}^{(k)}\middle|\boldsymbol{L}_{i,j,\cdot}^{(k)}\right) \propto P\left(\boldsymbol{L}_{i,j,\cdot}^{(k)}\middle|\boldsymbol{\theta}_{i,j}^{(k)}\right)P\left(\boldsymbol{\theta}_{i,j}^{(k)}\right) \propto$

$$e^{-\frac{1}{2\sigma_{j,k}^2}\left(\boldsymbol{L}_{i,j,\cdot}^{(k)}-\boldsymbol{\Psi}_i\boldsymbol{\theta}_{i,j}^{(k)}\right)'\left(\boldsymbol{L}_{i,j,\cdot}^{(k)}-\boldsymbol{\Psi}_i\boldsymbol{\theta}_{i,j}^{(k)}\right)-\frac{1}{2}\left(\boldsymbol{\theta}_{i,j}^{(k)}-\boldsymbol{u}_{j,k}^0\right)'\left(\boldsymbol{\Sigma}_{j,k}^0\right)^{-1}\left(\boldsymbol{\theta}_{i,j}^{(k)}-\boldsymbol{u}_{j,k}^0\right)} \propto$$

$$e^{\boldsymbol{\theta}_{i,j}^{(k)'}\left(\frac{\boldsymbol{\Psi}_i'\boldsymbol{\Psi}_i}{\sigma_{j,k}^2}+(\boldsymbol{\Sigma}_{j,k}^0)^{-1}\right)\boldsymbol{\theta}_{i,j}^{(k)}-2\left(\frac{\boldsymbol{L}_{i,j,\cdot}^{(k)}\boldsymbol{\Psi}_i}{\sigma_{j,k}^2}+\left(\boldsymbol{u}_{j,k}^0\right)'\left(\boldsymbol{\Sigma}_{j,k}^0\right)^{-1}\right)\boldsymbol{\theta}_{i,j}^{(k)}}.$$

It is known that the Gaussian family is the *conjugate prior* for the Gaussian likelihood function, and thus the posterior also follows a multivariate normal distribution. As a result, $\boldsymbol{\theta}_{i,j}^{(k)}|\boldsymbol{L}_{i,j,\cdot}^{(k)}\sim N_{p+1}(\boldsymbol{u}_{j,k}^1,\boldsymbol{\Sigma}_{j,k}^1)$ and $P\left(\boldsymbol{\theta}_{i,j}^{(k)}\middle|\boldsymbol{L}_{i,j,\cdot}^{(k)}\right)\propto e^{-\frac{1}{2}\left(\boldsymbol{\theta}_{i,j}^{(k)}-\boldsymbol{u}_{j,k}^1\right)'\left(\boldsymbol{\Sigma}_{j,k}^1\right)^{-1}\left(\boldsymbol{\theta}_{i,j}^{(k)}-\boldsymbol{u}_{j,k}^1\right)}$.

Comparing the above two equations for $P\left(\boldsymbol{\theta}_{i,j}^{(k)}\middle|\boldsymbol{L}_{i,j,\cdot}^{(k)}\right)$, we obtain the results $(\boldsymbol{\Sigma}_{j,k}^1)^{-1}=\left(\frac{\boldsymbol{\Psi}_i'\boldsymbol{\Psi}_i}{\sigma_{j,k}^2}+(\boldsymbol{\Sigma}_{j,k}^0)^{-1}\right)$ and $\left(\boldsymbol{u}_{j,k}^1\right)'\left(\boldsymbol{\Sigma}_{j,k}^1\right)^{-1}=\left(\frac{\boldsymbol{L}_{i,j,\cdot}^{(k)}\boldsymbol{\Psi}_i}{\sigma_{j,k}^2}+\left(\boldsymbol{u}_{j,k}^0\right)'\left(\boldsymbol{\Sigma}_{j,k}^0\right)^{-1}\right)$. This finishes the proof that $\boldsymbol{\theta}_{i,j}^{(k)}\middle|\boldsymbol{L}_{i,j,\cdot}^{(k)}\sim N_{p+1}(\boldsymbol{u}_{j,k}^1,\boldsymbol{\Sigma}_{j,k}^1)$, where $\boldsymbol{u}_{j,k}^1=\left(\frac{\boldsymbol{\Psi}_i'\boldsymbol{\Psi}_i}{\sigma_{j,k}^2}+(\boldsymbol{\Sigma}_{j,k}^0)^{-1}\right)^{-1}\left(\frac{\boldsymbol{\Psi}_i'\boldsymbol{L}_{i,j,\cdot}^{(k)}}{\sigma_{j,k}^2}+(\boldsymbol{\Sigma}_{j,k}^0)^{-1}\boldsymbol{u}_{j,k}^0\right)$ and $\boldsymbol{\Sigma}_{j,k}^1=\left(\frac{\boldsymbol{\Psi}_i'\boldsymbol{\Psi}_i}{\sigma_{j,k}^2}+(\boldsymbol{\Sigma}_{j,k}^0)^{-1}\right)^{-1}$.

Chapter 5

Sensory-based Failure Threshold Estimation for Remaining Useful Life Prediction

5.1 Introduction

As mentioned in Chapter 1, to achieve an appropriate estimation on the RUL distribution, there are two essential requirements that must be satisfied: (i) an accurate estimation of the failure threshold D_j; and (ii) a reliable degradation model that characterizes the degradation profile of a unit, such that the degradation status for operating unit j, L_{j,n_j+t}, can be accurately predicted at the future observation time $n_j + t$.

To the best of our knowledge, the existing literature still lacks a reliable approach that can effectively estimate the RUL distribution of a unit when the failure threshold is not known *a priori*. This Chapter seeks to fill this literature gap by developing a convex quadratic formulation that combines the historical population information and the condition monitoring data of an operating unit to online estimate its failure threshold. In other words, unlike many of the existing reliability studies that treat the failure threshold as a fixed value, this Chapter considers a random failure threshold for different units. In fact, this random failure threshold assumption has been also considered in the literature (e.g., [10], [11], [30], [31], [79], [80]). With the proposed efforts, we expect to attain more reliable and accurate RUL prediction.

5.2 Problem Formulation

To begin with, we assume that the degradation signals of a group of historical units during

lifecycle (run to failure) have been collected. These units can be used to build the mixed-effects degradation model in (2.5). Specifically, the degradation model of historical unit i is written as:

$$s_{i,t} = \eta(\boldsymbol{\varphi}_i, t) + \epsilon_{i,t}, \tag{5.1}$$

where $s_{i,t}$ is the sensor measurement of historical unit i at time t; η is the functional form of the degradation model; $\boldsymbol{\varphi}_i \in R^{p\text{x}1}$ is the random effects of historical unit i; and $\epsilon_{i,t}$ represents the random noise.

Without loss of generality, we further assume that under the same degradation conditions (e.g., failure mode, environmental condition), the degradation profiles (the fitted degradation model and the failure threshold) of units are expected to be the same, subject to some random variation. Under this assumption, our method will online estimate the failure threshold of an operating unit *j* by calculating the weight coefficient, w_{ij}, that measures the relative similarity between the degradation profiles of operating unit j and historical unit i compared to the other historical units. Please note that if such a weight coefficient w_{ij} is available, then we can reconstruct a degradation model for operating unit j via a weighted average of the fitted degradation models of the historical units:

$$E(\tilde{L}_{j,t}|\boldsymbol{L}_{j,\cdot}) = \sum_{i=1}^{m} \left(w_{ij}\eta(\hat{\boldsymbol{\varphi}}_i, t)\right), \tag{5.2}$$

where $\sum_{i=1}^{m} w_{ij} = 1$, and $w_{ij} \geq 0$ for $i = 1,2, \dots, m$; $\tilde{L}_{j,t}$ is the reconstructed signal for operating unit j at time t by using the proposed weight coefficient w_{ij}; $\boldsymbol{L}_{j,\cdot} = \left[L_{j,1}, \dots, L_{j,n_j}\right]^T$ is the available *in-situ* sensory data for unit j; $\hat{\boldsymbol{\varphi}}_i$ is the estimated parameters of historical unit i; and m is the number of historical units.

On the other hand, recall that the existing literature has also developed a powerful Bayesian

approach (e.g., Gebraeel *et al.* [29]) that can online calibrate the degradation model of an operating unit. Specifically, here we denote the updated degradation model via the Bayesian approach for operating unit j as:

$$E(\hat{L}_{j,t}|\boldsymbol{L}_{j,\cdot}) = \eta\Big(E(\boldsymbol{\theta}_j^{(1)}), t\Big), \tag{5.3}$$

where $\boldsymbol{\theta}_j^{(1)} \in R^{p\text{x}1}$ is the updated random effects of operating unit j; and $\hat{L}_{j,t}$ is the reconstructed signal for operating unit j at time t by using the Bayesian approach.

Our innovative idea is to find the weight coefficient, w_{ij}, such that the Bayesian updated model in (6) can be approximated by the reconstructed model in (5.2). In this way, we can achieve an accurate estimation of the failure threshold distribution by effectively leveraging the information of historical units. Specifically, the k^{th} moment of the failure threshold for operating unit j, D_j, can be estimated by the weighted sample moments:

$$E\big[D_j^k\big] = \textstyle\sum_{i=1}^{m}\big(w_{ij}s_{i,n_i}^k\big). \tag{5.4}$$

where $\sum_{i=1}^{m} w_{ij} = 1,$ and $w_{ij} \geq 0$ for $i = 1,2,\dots,m$; s_{i,n_i} is the last sensor measurement before failure of historical unit i. As a result, the key challenge is to estimate the w_{ij} accurately.

If we consider the sum of squared errors to characterize the goodness-of-fit between the reconstructed model in (5.2) and the Bayesian updated degradation in (5.3), then the following optimization formulation can be used to estimate w_{ij} given $\boldsymbol{L}_{j,\cdot}$:

$$\begin{aligned} &\min_{w_{ij}} \textstyle\sum_{t=1}^{n_j}\big(E[\hat{L}_{j,t}] - E[\tilde{L}_{j,t}]\big)^2 \\ &\text{s.t. } \textstyle\sum_{i=1}^{m} w_{ij} = 1 \\ &\quad w_{ij} \geq 0, \text{ for } i = 1,2,\dots,m, \end{aligned} \tag{5.5}$$

where n_j is the number of available observations from operating unit j; and $\sum_{i=1}^{m} w_{ij} = 1$ means that the reconstructed degradation model is a weighted sum of the degradation models of the historical units.

Theorem 1: If $\eta(\boldsymbol{\theta}_j, t) = \boldsymbol{X}_t \boldsymbol{\theta}_j = \sum_{k=1}^{p} X_{t,k}\, \theta_{j,k}$, where $\boldsymbol{X}_t$ contains time-dependent covariates, and $E[\boldsymbol{\theta}_j^{(1)}]$ lies in the convex hull of the set $\{\widehat{\boldsymbol{\varphi}}_1, \dots, \widehat{\boldsymbol{\varphi}}_m\}$, then there exists at least one set of $\{w_{1j}, \dots, w_{mj}\}$ such that $E[\tilde{L}_{j,t}] = E[\hat{L}_{j,t}]\ for\ \forall t$, in which $\sum_{i=1}^{m} w_{ij} = 1$ and $w_{ij} \geq 0$ (see Appendix for proof).

Theorem 2: If $\eta(\boldsymbol{\theta}_j, t) = \boldsymbol{X}_t \boldsymbol{\theta}_j = \sum_{k=1}^{p} X_{t,k}\, \theta_{j,k}$, where $\boldsymbol{X}_t$ contains time-dependent covariates, and $E[\boldsymbol{\theta}_j^{(1)}]$ lies in the convex hull of the set $\{\widehat{\boldsymbol{\varphi}}_1, \dots, \widehat{\boldsymbol{\varphi}}_m\}$, the optimal solution from formulation (8) guarantees that $E[\tilde{L}_{j,t}] = E[\hat{L}_{j,t}]\ for\ t = 1,2, \dots, n_j$ (see Appendix for proof).

Theorem 1 indicates that if the convex combination of the random effects of the historical units spans the parameter space of the degradation models, then it is always possible to reconstruct the Bayesian updated model by using the fitted degradation models of historical units. Theorem 2 then shows that such a reconstruction can be achieved by solving the optimization problem in (8). Please note that with more historical units available, it is more likely that a convex combination of the degradation models of historical units will span the entire parameter space of the possible degradation models.

Fig. 5.1 shows an example satisfying the conditions of Theorems 1 and 2. There are three historical units present in Fig. 5.1, in which the Bayesian updated degradation model of an operating unit by (5.3) is shown in the blue circle, and the reconstructed model for the operating unit by solving formulation (5.5) is shown in the red cross. From Fig. 5.1, we can see that the

Bayesian updated degradation model is successfully reconstructed via the convex combination of the degradation models of historical units, which is an expected consequence of Theorem 2.

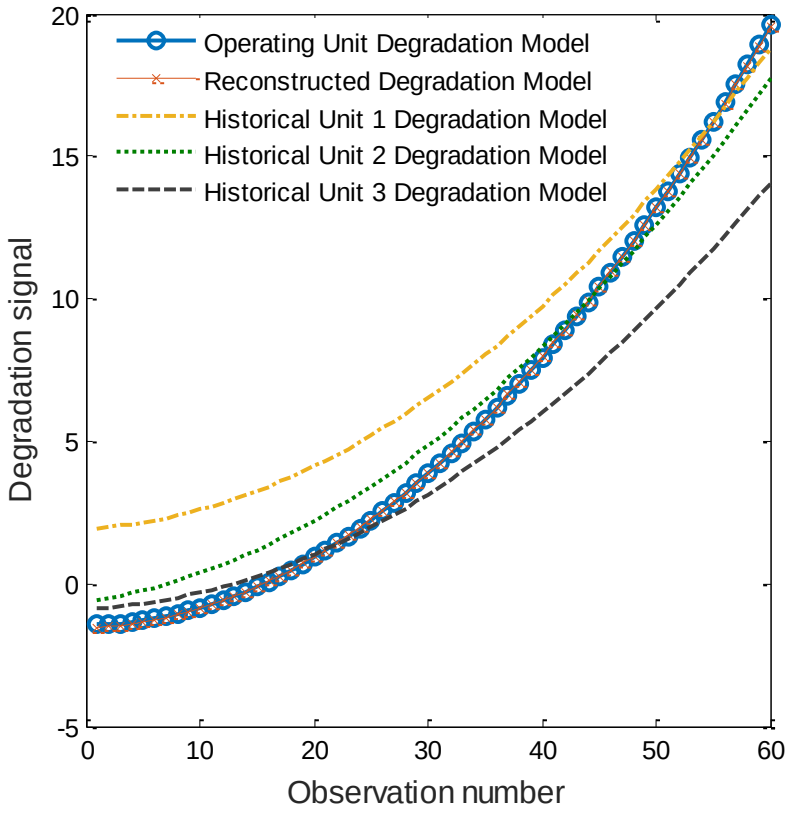

Fig. 5.1. An illustration of reconstructing the degradation model of an operating unit by the convex combination of the degradation models from the historical units

While Theorems 1 and 2 lay a theoretical foundation for the proposed method, direct implementation of the model in (5.5) may lead to over-fitting problems, especially when the number of available observations n_j is small for operating unit j (i.e., in the initial stage of condition monitoring). To address this issue, we further introduce the following regularization function in (5.6):

$$R(w_{1j}, \dots, w_{mj}) = \sum_{i=1}^{m} \left(w_{ij} \{ E[\hat{L}_{j,n_i}] - s_{i,n_i} \}^2 \right), \quad (5.6)$$

where $E[\hat{L}_{j,n_i}]$ is the expected sensor measurement of operating unit j at the failure time of historical unit i, and it can be calculated by projecting the Bayesian updated degradation model in (6) up to time n_i; and s_{i,n_i} is the sensor measurement when historical unit i fails. We can understand the proposed regularization function in the following way. Recall that we estimate the expected failure threshold of operating unit j based on the weighted average of the last sensor measurements of historical units in (5.4), where the weight coefficient w_{ij} controls the relative importance of historical unit i for estimating the failure threshold of operating unit j. As a result, if the projected sensor measurement of operating unit j at time n_i is noticeably different from the sensor measurement of historical unit i at time n_i, then historical unit i is less informative to estimate the failure threshold of operating unit j. Therefore, a larger penalty should be given to historical unit i as shown in this regularization function.

By combining formulation (8) and the regularization function in (9), then our proposed model becomes:

$$\min_{w_{ij}} \sum_{t=1}^{n_j} \frac{\left(E[\hat{L}_{j,t}] - E[\tilde{L}_{j,t}] \right)^2}{n_j} + \lambda R(w_{1j}, \dots, w_{mj})$$

$$\text{s.t. } \sum_{i=1}^{m} w_{ij} = 1$$

$$w_{ij} \geq 0, \text{ for } i = 1,2, \dots, m, \quad (5.7)$$

where λ is the regularization parameter and can be obtained via the K-Fold cross validation method. Here, formulation (5.7) considers minimizing both the mean squared error over the observed time domain for operating unit j, and the differences between the projected degradation

signal for operating unit j and the sensor measurements of historical units at their corresponding failure times.

We can further simplify the formulation in (5.7) to the following convex quadratic program:

$$\min_{\boldsymbol{w}_j} \boldsymbol{w}_j{}^T \boldsymbol{B}^T \boldsymbol{B} \boldsymbol{w}_j + \left(\lambda n_j \boldsymbol{c}^T - 2\boldsymbol{a}^T \boldsymbol{B}\right) \boldsymbol{w}_j$$

$$\text{s.t. } \boldsymbol{o}^T \boldsymbol{w}_j = 1$$

$$\boldsymbol{w}_j \geq \boldsymbol{0}, \tag{5.8}$$

where $\boldsymbol{w}_j = \left[w_{1j}, \ldots, w_{mj}\right]^T \in R^{m\text{x}1}$; $\boldsymbol{c}_j = \left[\left\{E\left[\hat{L}_{j,n_1}\right] - s_{1,n_1}\right\}^2 \quad \cdots \quad \left\{E\left[\hat{L}_{j,n_m}\right] - s_{m,n_m}\right\}^2\right]^T \in R^{m\text{x}1}$; $\boldsymbol{B}_j = \begin{bmatrix} \eta(\widehat{\boldsymbol{\varphi}}_1, 1) & \ldots & \eta(\widehat{\boldsymbol{\varphi}}_m, 1) \\ \vdots & \ddots & \vdots \\ \eta(\widehat{\boldsymbol{\varphi}}_1, n_j) & \ldots & \eta(\widehat{\boldsymbol{\varphi}}_m, n_j) \end{bmatrix} \in R^{n_j\text{x}m}$; $\boldsymbol{o} \in R^{m\text{x}1}$ is a vector of all ones; and $\boldsymbol{a}_j = \left[E[\hat{L}_{j,1}] \quad \ldots \quad E[\hat{L}_{j,n_j}]\right]^T \in R^{n_j\text{x}1}$ (see Appendix for details).

The formulation in (5.8) can be solved efficiently by many existing numeric solvers [73], which ensures that our proposed method can be effectively used during condition monitoring. Recall that once the optimal solution $\boldsymbol{w}_j$ is obtained, we can then estimate the failure threshold for operating unit j by using (5.4). In the next sub-section, we will investigate how to derive the RUL distribution of an operating unit during condition monitoring based on the proposed model in (5.8).

5.3 Estimation of the RUL distribution

Recall that our proposed model relies on comparing the Bayesian updated degradation model for the operating unit and the fitted degradation models of the historical units. Without loss of generality, below we focus on the polynomial functional form of the degradation model when deriving the RUL of the operating unit during condition monitoring. In fact, many degradation

models discussed in the existing literature can be transformed into the polynomial form [10], [67], such as the exponential form models [56], and random coefficient growth models [66]. In particular, the p^{th} order polynomial degradation model can be expressed as:

$$L_{j,t} = \boldsymbol{X}_t \boldsymbol{\theta}_j + \epsilon_{j,t} = \sum_{k=0}^{p} \theta_{j,\alpha} t^k + \epsilon_{j,t}, \tag{5.9}$$

where p is the order of the polynomial model and can be determined based on different criteria like the Akaike information criterion and the Bayesian information criterion [21]; $\boldsymbol{\theta}_j$ is the vector of random effects for the degradation model of operating unit *j* and it is assumed to follow a multivariate normal distribution $N_{p+1}(\boldsymbol{u}_j\ , \boldsymbol{\Sigma}_j\ )$; $\epsilon_{j,t}$ is the random noise and assumed to follow $N(0, \sigma_j^2)$; and $\boldsymbol{X}_t = [1, t, \dots, t^p]$. Here, the prior distribution of the random effects can be estimated by fitting the degradation profiles of historical units:

$$\boldsymbol{\theta}_j^{(0)} \sim N_{p+1}(\boldsymbol{u}_j^0, \boldsymbol{\Sigma}_j^0), \tag{5.10}$$

where $\boldsymbol{u}_j^0$ and $\boldsymbol{\Sigma}_j^0$ are the prior mean and variance of the random effects, respectively. Then, we can calculate the posterior distribution, $\boldsymbol{\theta}_j^{(1)}$ by leveraging the *in-situ* sensory data $\boldsymbol{L}_{j,\cdot}$ collected from operating unit *j*. For example, if we follow the Bayesian updating approach introduced in Gebraeel *et al.* [29], then the updated random effects $\boldsymbol{\theta}_j^{(1)}$ will still follow the normal distribution:

$$\boldsymbol{\theta}_j^{(1)} = \boldsymbol{\theta}_j | \boldsymbol{L}_{j,\cdot} \sim N_{p+1}(\boldsymbol{u}_j^1, \boldsymbol{\Sigma}_j^1), \tag{5.11}$$

where $\boldsymbol{u}_j^1 = \left(\frac{\boldsymbol{\Psi}_j{}^T \boldsymbol{\Psi}_j}{\sigma_j^2} + (\boldsymbol{\Sigma}_j^0)^{-1}\right)^{-1} \left(\frac{\boldsymbol{\Psi}_j{}^T \boldsymbol{L}_{j,\cdot}}{\sigma_j^2} + (\boldsymbol{\Sigma}_j^0)^{-1} \boldsymbol{u}_j^0\right)$,

$\boldsymbol{\Sigma}_j^1 = \left(\frac{\boldsymbol{\Psi}_j{}^T \boldsymbol{\Psi}_j}{\sigma_j^2} + (\boldsymbol{\Sigma}_j^0)^{-1}\right)^{-1}$, and $\boldsymbol{\Psi}_j \in R^{n_j \times (p+1)} = \begin{bmatrix} 1 & \dots & 1 \\ \dots & \dots & \dots \\ 1 & \dots & t^p \\ \dots & \dots & \dots \\ 1 & \dots & n_j^p \end{bmatrix}$.

Here, our use of the normal distribution for the random effects is primarily to take advantage of the convenient and closed-form solution provided for the posterior distribution when updating the random effects in real time. In fact, this normal assumption for the random effects has been widely adopted in the existing literature [2], [9-12], [17]. Please note that it is also possible to utilize other distributions or even some simulation-based approaches to compute the posterior distribution and RUL. In such cases, our proposed method for estimating the failure threshold is still applicable. Thus, to highlight our main ideas, for the remainder of the Chapter we only focus on the normal distribution for the random effects.

Consequently, the degradation signal of operating unit j at time t given $\boldsymbol{\theta}_j^{(1)}$ will follow:

$$L_{j,t}\left|\boldsymbol{\theta}_j^{(1)}\right. \sim N\left(\boldsymbol{X}_t\boldsymbol{u}_j^1, \boldsymbol{X}_t\boldsymbol{\Sigma}_j^1\boldsymbol{X}_t^T + \sigma_j^2\right). \tag{5.12}$$

Recall that the mean u_j^d and variance v_j^d of the failure threshold D_j for operating unit j can be estimated by $u_j^d = \sum_{i=1}^{m}\left(w_{ij}s_{i,n_i}\right)$ and $v_j^d = \sum_{i=1}^{m}\left(w_{ij}s_{i,n_i}^2\right) - \left(\sum_{i=1}^{m}\left(w_{ij}s_{i,n_i}\right)\right)^2$ using (7). If we assume the failure threshold D_j follows a normal distribution as well, i.e., $D_j \sim N\left(u_j^d, v_j^d\right)$, then the cumulative distribution function (CDF) of the RUL $\tilde{T}_j$ for operating unit j based on the available *in-situ* sensory data $\boldsymbol{L}_{j,:}$ can be calculated as:

$$P\left(\tilde{T}_j \le t\left|\boldsymbol{L}_{j,:}\right.\right) = P\left(L_{j,n_j+t} \ge D_j\left|\boldsymbol{L}_{j,:}\right.\right) = \Phi\left(\frac{\boldsymbol{X}_{n_j+t}\boldsymbol{u}_j^1 - u_j^d}{\sqrt{\boldsymbol{X}_{n_j+t}\boldsymbol{\Sigma}_j^1\boldsymbol{X}_{n_j+t}^T + \sigma_j^2 + v_j^d}}\right) = \Phi\left(g(t)\right). \tag{5.13}$$

Here, $\Phi(\cdot)$ is the CDF of the standard normal distribution.

Given that the RUL for operating unit j should be greater than 0, we further consider the truncated CDF conditioning on $\tilde{T}_j \geq 0$:

$$P\left(\tilde{T}_j \leq t \middle| \tilde{T}_j \geq 0, \boldsymbol{L}_{j,\cdot}\right) = \frac{P\left(0 \leq \tilde{T}_j \leq t \middle| \boldsymbol{L}_{j,\cdot}\right)}{P\left(\tilde{T}_j \geq 0 \middle| \boldsymbol{L}_{j,\cdot}\right)} = \frac{\Phi(g(t)) - \Phi(g(0))}{1 - \Phi(g(0))}. \tag{5.14}$$

As the truncated CDF in (5.14) is skewed, we can use the median as the point estimator for the RUL prediction. Numerically, this is equivalent to finding the observation time t such that $P\left(\tilde{T}_j \leq t \middle| \tilde{T}_j \geq 0, \boldsymbol{L}_{j,\cdot}\right) = 0.5$.

5.4 Flow Chart Summary

Fig. 5.2 illustrates the flow chart of the proposed methodology. The solid boxes and lines correspond to offline analyses. The dashed boxes and lines correspond to online analyses as the new condition monitoring data from an operating unit are observed. From the offline analysis, we extract the information regarding the last sensor measurement (s_{i,n_i}) and the degradation models of the historical units. Specifically, we estimate the failure threshold of historical unit i by its last sensor measurement before failure (s_{i,n_i}). We also fit a parametric model for the degradation signal of each historical unit as shown in (5.1) for a generic functional form and in (5.9) for the polynomial functional forms. At the end of the offline analysis, we obtain a prior distribution of the parameters in the degradation model based on the historical units. Particularly, in this Chapter, we focus on the normal distribution in (5.10) with a mean and a variance calculated as the sample mean and the sample variance of the parameters from the historical units. Then, during condition monitoring, we first update the degradation model for a new operating unit via the Bayesian approach (i.e., find the posterior distribution of the parameters) based on the *in-situ* sensor

measurements for that unit as shown in (5.11). Using this updated degradation model, we then estimate the failure threshold of this operating unit by solving (5.8). Finally, the updated degradation model and the estimated failure threshold are utilized to infer the RUL of this operating unit by (5.13) and (5.14).

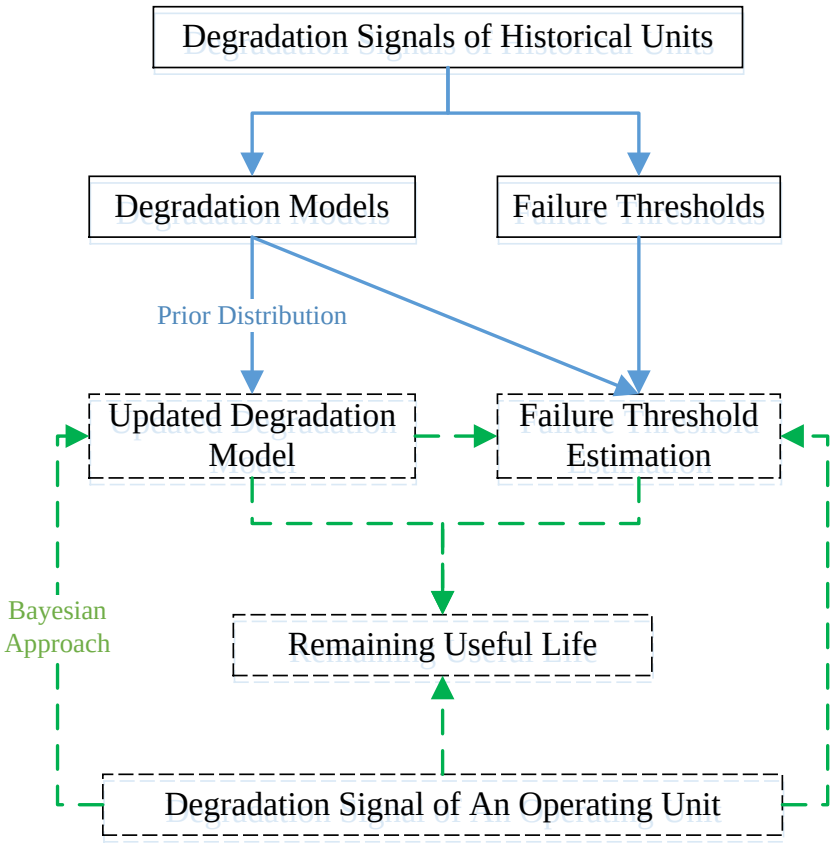

Fig. 5.2. A flow chart that illustrates the flow of the proposed methodology.

5.5 Simulation Studies

In this section, we investigate the performance of the proposed methodology and compare it with the existing literature in which the failure threshold distribution is estimated based on the population-wide characteristics. In fact, the existing data-driven approach for estimating the failure

threshold [10] can be regarded as a special case of our proposed method with equal weight coefficient $w_{ij} = 1/m$, for $i = 1,2, \dots, m$, i.e., treating each historical unit with equal importance. On the contrary, our proposed method will derive the optimal weight coefficient w_{ij}, for $i = 1,2, \dots, m$ by solving (5.8) when estimating the failure threshold of operating unit j.

5.5.1 Simulation Setup

To challenge our developed algorithm, in this simulation study, we consider the scenario where units degrade and fail according to one of the three possible failure modes, which all follow the following quadratic degradation model:

$$L_{i,t} = \theta_0 + \theta_1 t + \theta_2 t^2 + \epsilon_{i,t}. \tag{5.15}$$

TABLE 5.1. THE PARAMETERS OF THE DEGRADATION MODELS FOR THESE THREE FAILURE MODES

Failure mode	$\boldsymbol{\theta}_i^{(0)}$	$\epsilon_{i,t}$	D_i
1	$N\left(\begin{bmatrix} 0 \\ 0.025 \\ 0.004 \end{bmatrix}, \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 10^{-4} & 0 \\ 0 & 0 & 10^{-6} \end{bmatrix}\right)$	$N(0,9)$	$N(110,9)$
2	$N\left(\begin{bmatrix} 30 \\ -0.002 \\ -0.005 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 10^{-4} & 0 \\ 0 & 0 & 10^{-6} \end{bmatrix}\right)$	$N(0,9)$	$N(-150,9)$
3	$N\left(\begin{bmatrix} -100 \\ 0.004 \\ 0.008 \end{bmatrix}, \begin{bmatrix} 0.5 & 0 & 0 \\ 0 & 10^{-4} & 0 \\ 0 & 0 & 10^{-6} \end{bmatrix}\right)$	$N(0,9)$	$N(110,9)$

In addition, we assume that units within the same failure mode (i) share the same expected failure threshold subject to some variations; and (ii) their random effects come from the same distribution. The detailed simulation setup is summarized in Table 5.1. In this simulation study, we further

assume that the failure modes of the training units and the operating unit are unknown, which is often the case in real applications. As we will show below, our proposed method has an advantage that it can automatically infer the failure threshold and the RUL of the operating unit even without knowledge of the failure mode of the training units.

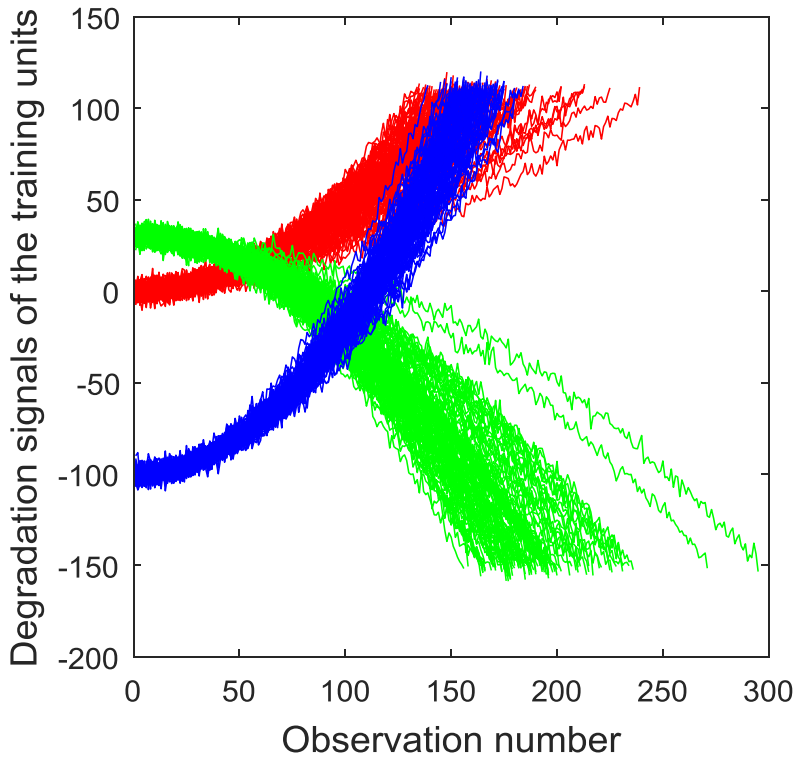

Fig. 5.3. The degradation signals of training units. Red, green, and blue corresponds to different failure modes.

We randomly generated 100 training units and 50 testing units from each of the three possible failure modes summarized in Table 5.1. Each training unit runs to failure, whereas for testing unit, we evaluate the performance of RUL prediction with various amounts of available observations (20, 40, 60, 80, 100 and 120 observations) before failure. The generated degradation signals for the training and testing units are shown in Fig. 5.3 and Fig. 5.4, respectively.

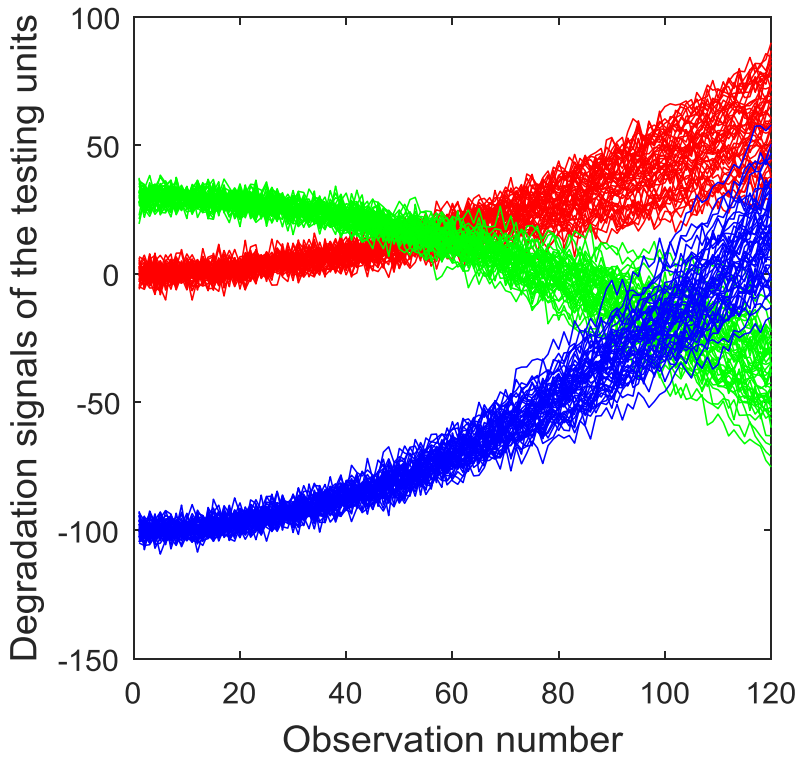

Fig. 5.4. The degradation signals of testing units up to 120 observations

In this Chapter, we consider the following two evaluation metrics:

(i) The average prediction error of the failure threshold (FT_e) as defined in (5.16).

(ii) The average prediction error of the RUL (RL_e) as defined in (5.17).

Metric (i) measures the average prediction error between the estimated failure threshold and the true failure threshold for the testing units:

$$FT_e(\%) = \frac{100}{N} * \sum_{j=1}^{N} \frac{\left|E[D_j] - D_j^{True}\right|}{D_j^{True}}, \tag{5.16}$$

where N is the number of testing units; $E\left[D_j\right]$ is the estimated failure threshold for testing unit j using (5.4); and D_j^{True} is the true failure threshold for testing unit j.

Metric (ii) measures the average prediction error of the RUL:

$$RL_e\ (\%) = \frac{100}{N} * \sum_{j=1}^{N} \left| \frac{(T_j + n_j) - (\tilde{T}_j + n_j)}{T_j + n_j} \right| = \frac{100}{N} * \sum_{j=1}^{N} \left| \frac{T_j - \tilde{T}_j}{T_j + n_j} \right|, \quad (5.17)$$

where T_j is the true RUL for testing unit j; $\tilde{T}_j$ is the estimated RUL for testing unit j; n_j is the number of available observations for testing unit j; and $T_j + n_j$ is the total lifetime for testing unit j.

5.5.2 Estimation of the Weight Coefficient, Failure Threshold and RUL

For the estimation of the weight coefficient, failure threshold, and RUL, we assume that the simulation setup in Section 5.5.1 is not given. Then, we follow the procedure shown in Section 5.4. First, we estimate the failure threshold of training unit i by its last observation (s_{i,n_i}), and we construct a 2nd order polynomial model for each training unit i and denote its parameters by $\boldsymbol{\varphi}_i = [\varphi_{i,0}, \varphi_{i,1}, \varphi_{i,2}]^T$. Next, we assume that the parameters of testing unit j follow a normal distribution $\boldsymbol{\theta}_j^{(0)} \sim N_{p+1}(\boldsymbol{u}_j^0, \boldsymbol{\Sigma}_j^0)$, where $\boldsymbol{u}_j^0$ and $\boldsymbol{\Sigma}_j^0$ are estimated by the sample mean and the sample variance of the parameters $\boldsymbol{\varphi}_i$ of training units, respectively. We then find the posterior distribution $\boldsymbol{\theta}_j^{(1)}$ based on the *in-situ* observations $\boldsymbol{L}_{j,\cdot}$ from testing unit j as shown in (5.11). Consequently, we solve (11) to calculate the weight coefficient w_{ij}. In this way, we estimate the failure threshold of testing unit j with mean $u_j^d = \sum_{i=1}^{m}(w_{ij} s_{i,n_i})$ and variance $v_j^d = \sum_{i=1}^{m}(w_{ij} s_{i,n_i}^2) - \left(\sum_{i=1}^{m}(w_{ij} s_{i,n_i})\right)^2$. Finally, the RUL is estimated by (5.14).

Fig. 5.5 shows the average prediction error of the failure threshold as a function of the number of available observations in the testing units. In addition, one standard deviation bars of the average

errors are shown in the plot as well. Here, we denote the existing approach for the RUL estimation as the "benchmark method" which estimates the failure threshold distribution based on population-wide characteristics. Recall that the existing method can be regarded as a special case of our proposed failure threshold estimation approach by taking equal weight coefficient $w_{ij} = 1/m$.

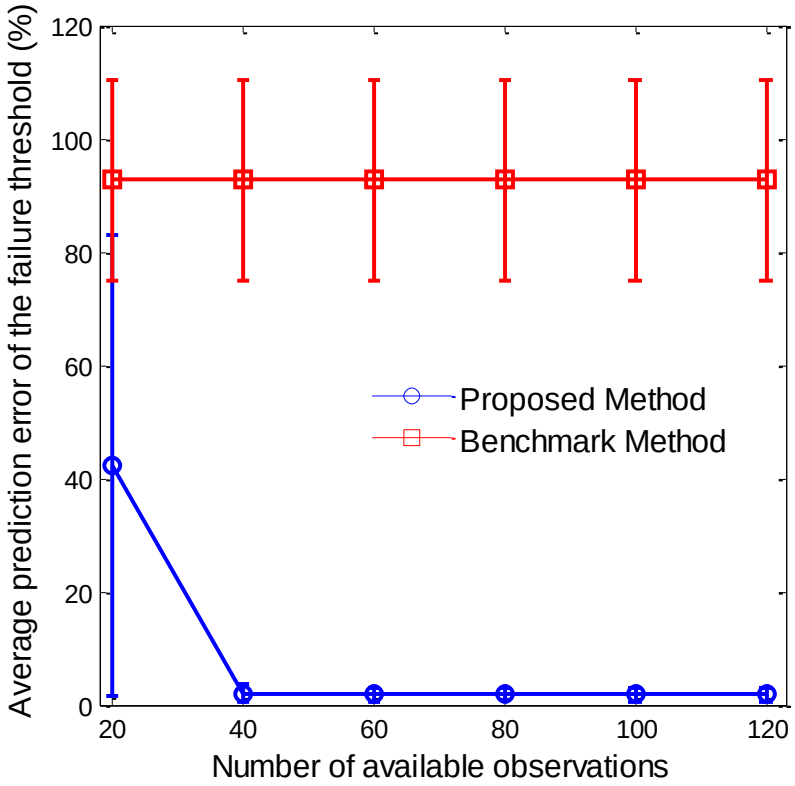

Fig. 5.5. The average prediction error (%) of the failure threshold as a function of the number of available observations in the testing units.

From Fig. 5.5, we can see that the benchmark method consistently has poor performance and does not improve with the increase of observations available, as the failure threshold is estimated based on the population-wide characteristics from training units. On the contrary, the failure threshold is accurately estimated by our proposed method with an error of approximately <2%

when there are 40 or more observations collected in the testing units. Specifically, the results show that our proposed methodology can estimate the failure threshold accurately even at a considerably early stage of the degradation. We also observe that there tends to be a large prediction error when there are only 20 observations. This is most likely because the Bayesian updated model in (6) may not fully capture the unique degradation characteristics of each testing unit from the few available observations. However, as more data are observed during condition monitoring, Fig. 5.5 clearly shows that the estimation accuracy of the failure threshold is significantly improved using our proposed method.

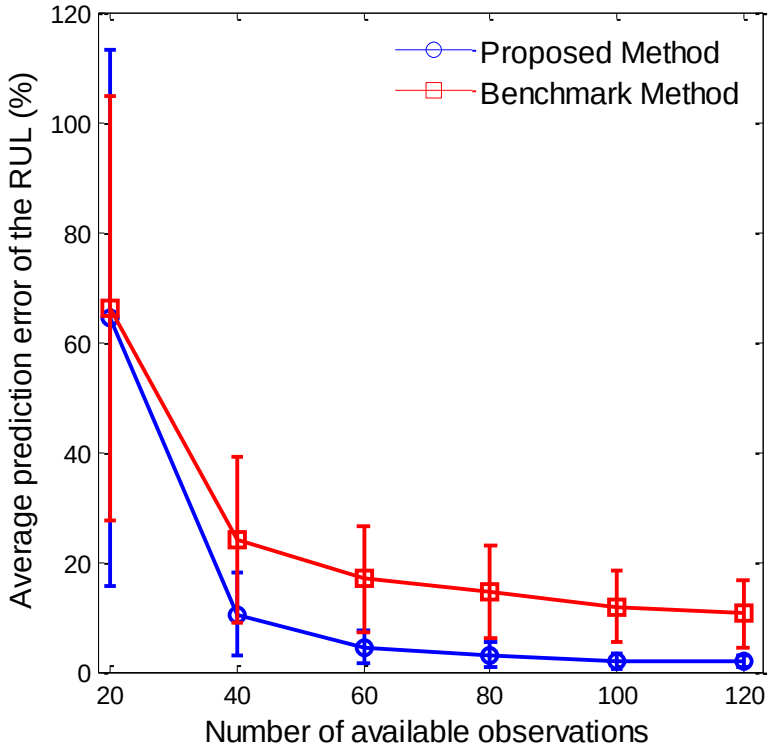

Fig. 5.6. The average prediction error (%) of the RUL as a function of the number of available observations in the testing units.

Moreover, Fig. 5.6 shows the average prediction error of the RUL and the corresponding one standard deviation bars as a function of the number of available observations in the testing units.

By comparing Fig. 5.5 and Fig. 5.6, we can see that the prediction accuracy is highly dependent on the failure threshold accuracy. Specifically, as the estimation accuracy of the failure threshold for the operating unit improves, the RUL prediction using our proposed method also improves and clearly outperforms the benchmark approach. Moreover, from the one standard deviation bar, we can see that the proposed method has a lower variance in the average prediction error than the benchmark approach as well.

To further evaluate the accuracy of the derived weight coefficient w_{ij}, we propose to consider the following metric:

$$p_j^{(k)} = \sum_{i \in FM_k} w_{ij}, \tag{5.18}$$

where FM_k represents the set of training units that fail under failure mode k. Generally speaking, if operating unit *j* degrades under failure mode k, then it is expected that the degradation profile of operating unit *j* should be relatively more similar to those of the training units that fail under failure mode k. In other words, we expect to see $p_j^{(k)}$ close to 1. In this sense, $p_j^{(k)}$ can be considered as a probability estimation that unit j fails under failure mode k.

The results are summarized in Figs. 5.7, 5.8 and 5.9. In those Figures, for better illustration, we sort the unit # such that the first 50 units (i.e., unit #1 to #50) degrade under failure mode 1, the next 50 units degrade under failure mode 2, and the last 50 units degrade under failure mode 3.

Clearly, we can see that the probability estimation of the failure mode in (5.18) provides very accurate results when the available number of observations is greater than or equal to 40 (i.e., approximately about 25% of the lifecycle of units). One possible reason is due to the high accuracy in the failure threshold estimation after 40 observations as shown in Fig. 5.5. This observation

further shows an additional advantage of our proposed method that it also provides a real time estimation about the failure mode of an operating unit (when the failure modes of training units, i.e., FM_k, are assumed to be known). Note that if the failure modes of the training units are known, it is possible to add additional structural constraints that force the pairwise product between any two training units under different failure modes to be zero or close to zero. This will be of high value for units with limited number of observations. We discuss a similar approach for structural degradation modeling in Chapter 6.

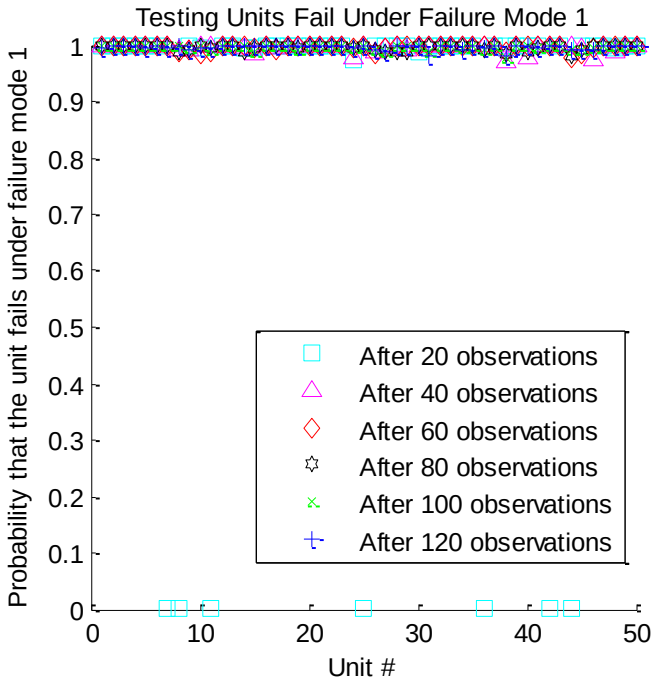

Fig. 5.7. The estimated probability that an operating unit fails under failure mode 1 given that its true failure mode is 1.

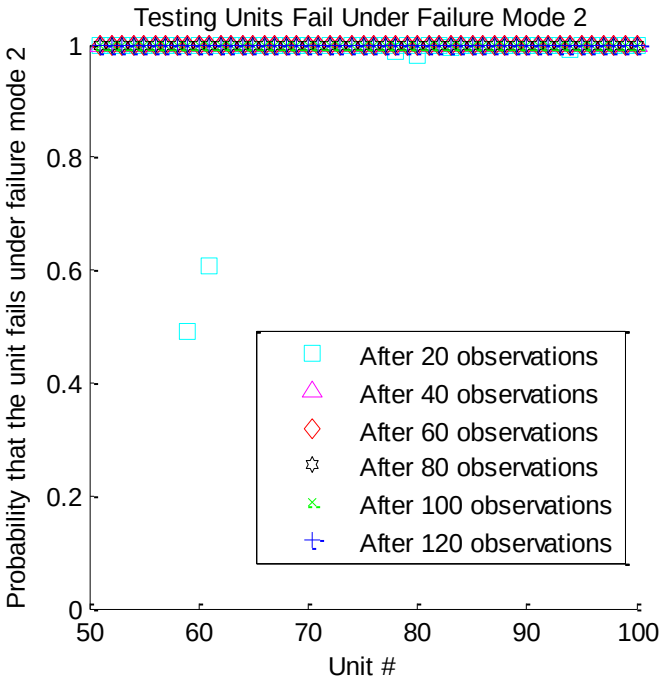

Fig. 5.8. The estimated probability that an operating unit fails under failure mode 2 given that its true failure mode is 2.

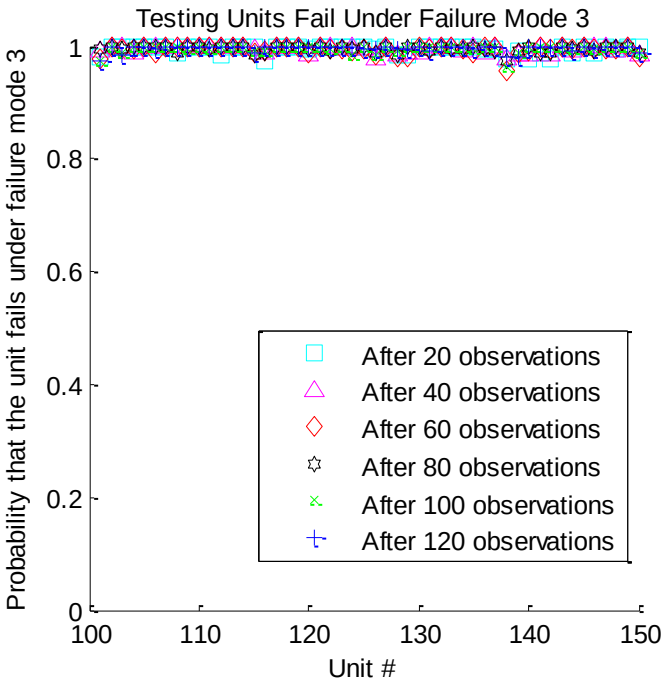

Fig. 5.9. The estimated probability that an operating unit fails under failure mode 3 given that its true failure mode is 3.

5.6 Case Study

In this section, we further evaluate our proposed method for prognostic analysis based on a degradation dataset for aircraft engines. The dataset is generated from commercial modular aero-propulsion system simulator (C-MAPSS), which is discussed in Section 4.2. In particular, the dataset contains $m = 100$ training units with a total of $\sum_{i=1}^{m} n_i = 33560$ observations. Units are subject to two potential failure modes: (i) a fault at the high-pressure compressor (HPC), or (ii) a fault at the fan. In addition, the dataset also includes 100 testing units that contain a total of 21350 observations with unknown failure modes. Unlike the training units, we do not have the full sensor signals until failure for the testing units. Instead, the dataset contains only partial sensor signals for the testing units to some points before failure. The actual RUL for the testing units is stored in a separate file. In this way, we are able to numerically evaluate the performance of our proposed method and compare with the existing benchmark. For each unit, there are 21 sensors that measure different characteristics of the engine performance [32]. The descriptions of the 21 sensors are given in Table 5.2.

5.6.1 Degradation Modeling

Based on the preliminary analysis [10], the degradation signals exhibit an exponential functional form. Thus, we consider the following exponential degradation model:

$$y_{i,j,t} = \gamma_j + e^{\theta_{i,j,0} + \theta_{i,j,1} t + \theta_{i,j,2} t^2 + \varepsilon_{i,j,t}}, \tag{5.19}$$

where $y_{i,j,t}$ is the sensor measurement for unit i, sensor j at time t; $\theta_{i,j,0}$, $\theta_{i,j,1}$, and $\theta_{i,j,2}$ are the random effects for unit i, sensor j; γ_j is the fixed-effect parameter for sensor j; and $\varepsilon_{i,j,t}$ is the random noise. Similar to Gebraeel [56] and Liu *et al.* [10], we use a log-transformation to the

original data and focus on modeling the logged signals:

$$L_{i,j,t} = \ln\left(y_{i,j,t} - \gamma_j\right) = \theta_{i,j,0} + \theta_{i,j,1}t + \theta_{i,j,2}t^2 + \varepsilon_{i,j,t}. \quad (5.20)$$

TABLE 5.2. DETAILED SENSOR DESCRIPTIONS [27].

Symbol	Description	Units
T2	Total temperature at fan inlet	°R
T24	Total temperature at LPC outlet	°R
T30	Total temperature at HPC outlet	°R
T50	Total temperature at LPT outlet	°R
P2	Pressure at fan inlet	Psia
P15	Total pressure in bypass-duct	Psia
P30	Total pressure at HPC outlet	psia
Nf	Physical fan speed	rpm
Nc	Physical core speed	rpm
epr	Engine pressure ratio (P50/P2)	--
Ps30	Static pressure at HPC outlet	psia
phi	Ratio of fuel flow to Ps30	pps/psi
NRf	Corrected fan speed	rpm
NRc	Corrected core speed	rpm
BPR	Bypass Ratio	--
farB	Burner fuel-air ratio	--
htBleed	Bleed Enthalpy	--
Nf_dmd	Demanded fan speed	rpm
PCNfR_dmd	Demanded corrected fan speed	rpm
W31	HPT coolant bleed	lbm/s
W32	LPT coolant bleed	lbm/s

5.6.2 Estimation of the RUL distribution

To numerically evaluate the performance of our proposed methodology for prognostic analysis, we follow a similar procedure as described in Section 5.5.3 and then compute the RUL of the testing units by implementing the proposed method in Section 5.4. Please note that here we have conducted the Henze-Zirckler's test based on the training dataset for each failure mode and the

results show that it is satisfactory to use the normal distribution as the prior for the random effects as described in (5.10).

TABLE 5.3. THE PERFORMANCE MEASURE OF RL_e BY USING THE PROPOSED AND THE BENCHMARK METHODOLOGIES.

Sensor	Our proposed method	Benchmark method	Improvement (%)
T24	21.30%	26.73%	+20.32%
T50	22.06%	27.25%	+19.06%
P30	43.21%	44.91%	+3.79%
Nf	24.04%	38.39%	+37.37%
Ps30	20.80%	23.40%	+11.11%
phi	45.01%	46.52%	+3.26%
NRf	24.66%	36.62%	+32.67%
BPR	31.55%	34.04%	+7.30%
htBleed	21.78%	29.37%	+25.83%
W31	33.25%	39.26%	+15.31%
W32	32.17%	34.78%	+7.50%
Average	29.08%	34.66%	+16.68%

Given the actual RUL of the testing units, we then calculate the average prediction error of the RUL, RL_e as defined in (5.17), by using the sensor signals from all of the 100 testing units. The results are summarized in Table 5.3. In particular, here we consider 11 sensors, which are selected in [10] and shown to be highly related to the degradation mechanism.

From Table 5.3, it can be seen that our proposed method improves the RUL prediction for all selected sensors when compared to the benchmark method. Please note that because the failure modes of training units are not available in the dataset, to achieve a fair comparison, here we employ the same benchmark approach used in the simulation study.

To further understand and evaluate the performance of the proposed and benchmark methods regarding RUL prediction, we calculate the average percentage error (defined in (5.17)) at different

levels of actual RUL for the testing units. The results for the proposed and the benchmark methodologies are shown in Figs. 5.10 and 5.11, respectively. Specifically, "100" refers to the average prediction error for the testing units that have 100 or less actual RUL. The maximum actual RUL among all the 100 testing units is 145. Thus, "145" refers to the average prediction error based on all the testing units. Note that we are only given partial signals for the testing units; therefore, different sets of testing units are used here to evaluate the RUL prediction at different levels of the actual RUL.

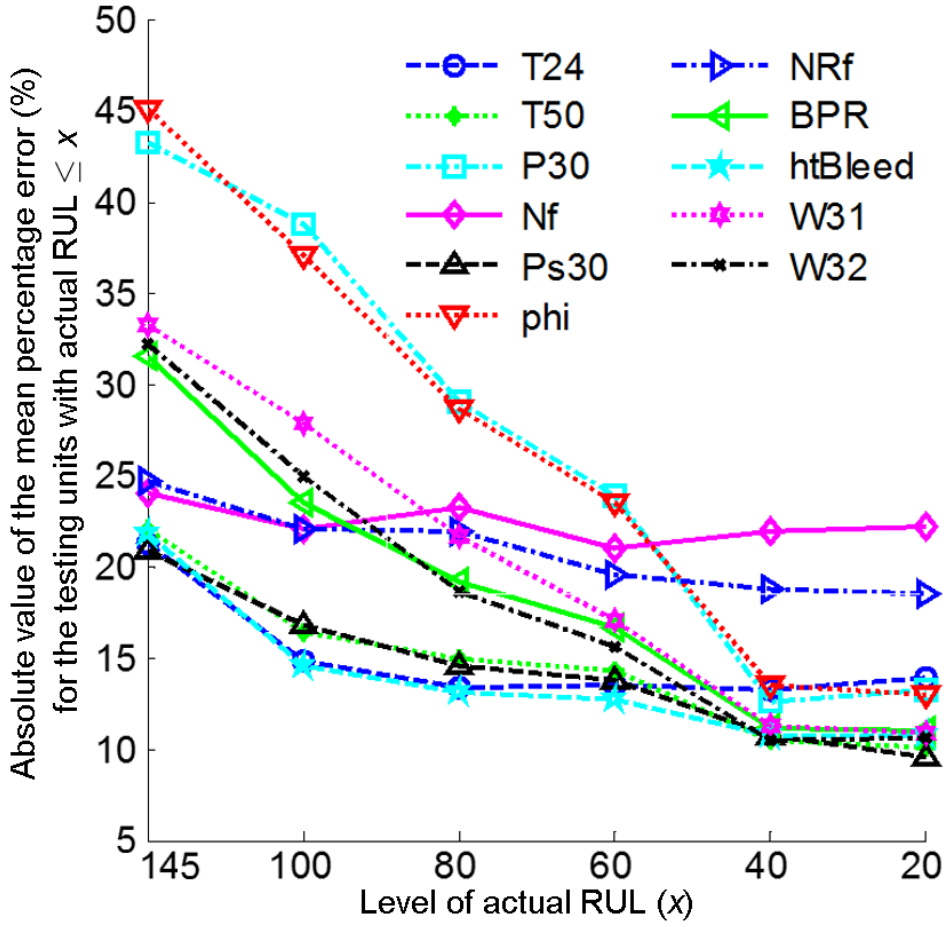

Fig. 5.10. Average prediction error at different levels of actual RUL

Based on Fig. 5.10, we can see that in general the RUL predictions become more accurate as the unit approaches failure. One possible reason is that less actual RUL typically means that more sensory data have been collected, and thus we are more confident about the updated Bayesian degradation model as well as the estimated failure threshold for the operating unit.

Fig. 5.11 further shows the average percentage errors for the proposed and the benchmark methodologies by using the best and the worst performing sensors according to the results in Table 5.3. From Fig. 5.11, we clearly see that the proposed method outperforms the benchmark method in both sensors. This case study further demonstrates the effectiveness of our proposed method for RUL prediction.

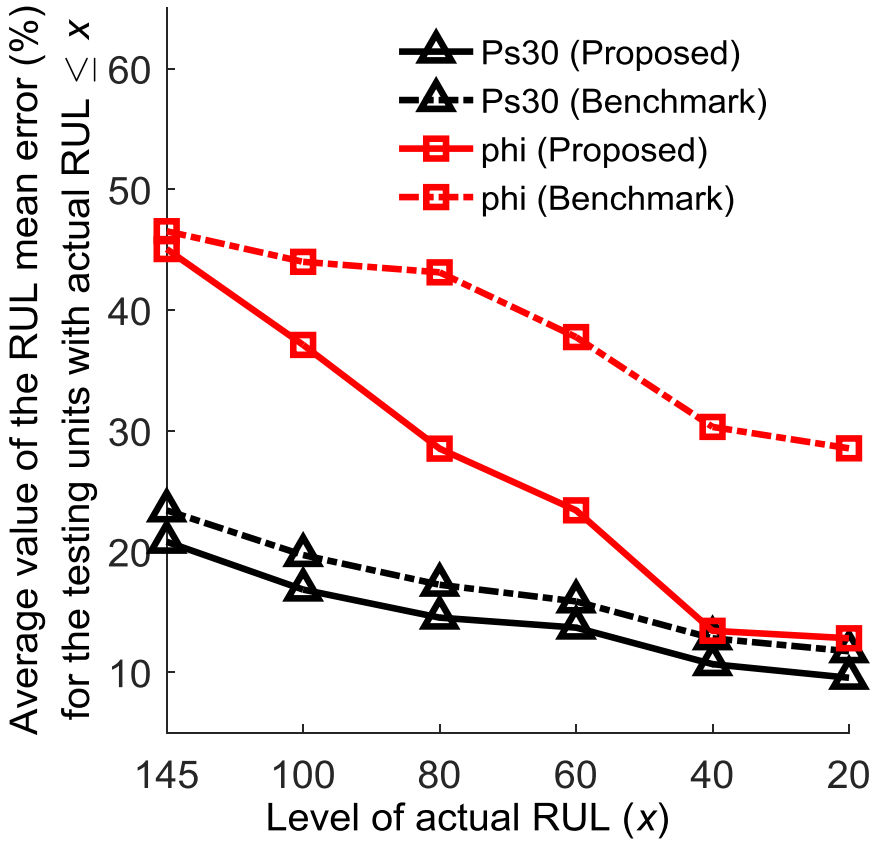

Fig. 5.11. Comparison of the average prediction error between the proposed and benchmark methodologies by using the best performing sensor (Ps30) and the worst performing sensor (phi)

5.7 Conclusions

Existing literature has been focused on developing advanced degradation modeling techniques to improve prognostics during condition monitoring. While beneficial, these methods often assume that the failure threshold is either already known *a priori* or simply estimated based on the population information. Such an assumption limits the accuracy and applicability of the prognostic analysis in many applications.

This Chapter aims to address this issue by real time estimating the failure threshold of an operating unit during condition monitoring when the failure threshold is not known *a priori*. In particular, we developed a convex quadratic formulation that fuses the degradation models of historical units along with the *in-situ* sensory data of an operating unit to online estimate its failure threshold. Such information can be effectively integrated with the existing degradation modeling techniques for better prognostic analysis. As demonstrated in the simulation studies, the proposed method provided an accurate estimation of the failure threshold, even at an early stage of the degradation. With the improved estimation of the failure threshold, a more accurate prediction of the RUL can be achieved. These findings were further reaffirmed in a case study that involves a degradation dataset of aircraft turbofan engines. These results not only show the effectiveness of our proposed methodology, but also shed light on the importance of online estimating the failure threshold for an operating unit, which has been previously ignored in the existing literature.

Given the importance of the topic, we believe this study will stimulate follow-up research in the near future. For example, it would be of interest to also consider a Bayesian framework for updating the failure threshold, as compared to the frequentist approach proposed in this Chapter.

Also, there is high potential for future studies to utilize the proposed approach for better diagnostic inferences, maintenance scheduling [3], work in progress regulations [81], etc.

5.8 Appendix

Proof for Theorem 1:

If $\eta(\boldsymbol{\theta}_j, t) = \boldsymbol{X}_t\boldsymbol{\theta}_j$, then $E[\hat{L}_{j,t}] = E[\eta(\boldsymbol{\theta}_j, t)] = E[\boldsymbol{X}_t\boldsymbol{\theta}_j] = \boldsymbol{X}_t E[\boldsymbol{\theta}_j] = \eta(E[\boldsymbol{\theta}_j], t)$. If $E[\boldsymbol{\theta}_j]$ lies in the convex hull of the set $\{\boldsymbol{\varphi}_1, \dots, \boldsymbol{\varphi}_m\}$, then $E[\boldsymbol{\theta}_j] = \sum_{i=1}^{m}(w_{ij}\boldsymbol{\varphi}_i)$ such that $\sum_{i=1}^{m} w_{ij} = 1$ and $w_{ij} \geq 0$, for $i = 1,2, \dots, m$. Also, $E[\hat{L}_{j,t}] = \eta(E[\boldsymbol{\theta}_j], t) = \eta(\sum_{i=1}^{m}(w_{ij}\boldsymbol{\varphi}_i), t) = \boldsymbol{X}_t \sum_{i=1}^{m}(w_{ij}\boldsymbol{\varphi}_i) = \sum_{i=1}^{m}(w_{ij}\boldsymbol{X}_t\boldsymbol{\varphi}_i) = \sum_{i=1}^{m}\left(w_{ij}\eta(\boldsymbol{\varphi}_i, t)\right) = \sum_{i=1}^{m}\left(w_{ij}\eta(\boldsymbol{\varphi}_i, t)\right) = E[\tilde{L}_{j,t}]$. This finished the proof that there exists at least one set of $\{w_{1j}, \dots, w_{mj}\}$ such that $E[\tilde{L}_{j,t}] = E[\hat{L}_{j,t}]\ for\ \forall t$, in which $\sum_{i=1}^{m} w_{ij} = 1$ and $w_{ij} \geq 0$.

Proof for Theorem 2:

From Theorem 1, we assume that there is one set of $\{w_{1j}, \dots, w_{mj}\}$, in which $\sum_{i=1}^{m} w_{ij} = 1$ and $w_{ij} \geq 0$, for $i = 1,2, \dots, m$ such that $E[\tilde{L}_{j,t}] = E[\hat{L}_{j,t}]\ \forall t = 1,2, \dots, n_j$.

Now, we show that the solution of formulation (5.5) guarantees that $E[\tilde{L}_{j,t}] = E[\hat{L}_{j,t}]$. First, it is straightforward to see that $\sum_{t=1}^{n_j}\left(E[\hat{L}_{j,t}] - E[\tilde{L}_{j,t}]\right)^2 = 0$ is true if and only if $E[\tilde{L}_{j,t}] = E[\hat{L}_{j,t}]\ \forall t = 1,2, \dots, n_j$.

Accordingly, there exists at least one set $\{w_{ij}\}$ such that $\sum_{i=1}^{m} w_{ij} = 1$ and $w_{ij} \geq 0$, for $i =$

$1,2,\dots,m$, where $\sum_{t=1}^{n_j}\left(E[\hat{L}_{j,t}]-E[\tilde{L}_{j,t}]\right)^2=0$. As a result of $\min_{w_{ij}}\sum_{t=1}^{n_j}\left(E[\hat{L}_{j,t}]-E[\tilde{L}_{j,t}]\right)^2\geq 0$, then the optimal solution from formulation (5.5) is achieved at $\sum_{t=1}^{n_j}\left(E[\hat{L}_{j,t}]-E[\tilde{L}_{j,t}]\right)^2=0$.

Therefore, if $\eta(\boldsymbol{\theta}_j,t)=\boldsymbol{X}_t\boldsymbol{\theta}_j$ and $E[\boldsymbol{\theta}_j]$ lies in the convex hull of the set $\{\boldsymbol{\varphi}_1,\dots,\boldsymbol{\varphi}_m\}$, then the solution of formulation (8) guarantees that $E\left[\tilde{L}_{j,t}\right]=E\left[\hat{L}_{j,t}\right]\ \forall t=1,2,\dots,n_j$.

Proof (5.8) is equivalent to (5.7) and that the formulation is convex:

First, the constraint $\sum_{i=1}^{m}w_{ij}=1$ can be written as a vector multiplication $[1\quad\dots\quad 1]*\begin{bmatrix}w_{1j}\\ \vdots\\ w_{mj}\end{bmatrix}=\boldsymbol{o}^T\boldsymbol{w}_j=1$ and $w_{ij}\geq 0,$ for $i=1,2,\dots,m$ can be written as $\begin{bmatrix}w_{1j}\\ \vdots\\ w_{mj}\end{bmatrix}\geq\begin{bmatrix}0\\ \vdots\\ 0\end{bmatrix}$. Thus, the constraints in formulation (5.8) are equivalent to the constraints in formulation (5.7). On one hand, the first piece of the objective function in formulation (5.7) can be expressed as: $\sum_{t=1}^{n_j}\frac{\left(E[L_{j,t}]-E[\tilde{L}_{j,t}]\right)^2}{n_j}=\sum_{t=1}^{n_j}\frac{\left(E[L_{j,t}]-[\eta(\boldsymbol{\varphi}_1,\mathrm{t})\quad\dots\quad\eta(\boldsymbol{\varphi}_m,\mathrm{t})]\begin{bmatrix}w_{1j}\\ \vdots\\ w_{mj}\end{bmatrix}\right)^2}{n_j}=\frac{(\boldsymbol{a}-\boldsymbol{B}\boldsymbol{w}_j)^T(\boldsymbol{a}-\boldsymbol{B}\boldsymbol{w}_j)}{n_j}=$

$\frac{\boldsymbol{a}^T\boldsymbol{a}-2\boldsymbol{a}^T\boldsymbol{B}\boldsymbol{w}_j+\boldsymbol{w}_j{}^T\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{w}_j}{n_j}$, where $\boldsymbol{B}=\begin{bmatrix}\eta(\boldsymbol{\varphi}_1,1) & \dots & \eta(\boldsymbol{\varphi}_m,1)\\ \vdots & \ddots & \vdots\\ \eta(\boldsymbol{\varphi}_1,n_j) & \dots & \eta(\boldsymbol{\varphi}_m,n_j)\end{bmatrix}\in R^{n_j\mathrm{x}m}$ and $\boldsymbol{a}=$ $\left[E[L_{j,1}]\quad\dots\quad E[L_{j,n_j}]\right]^T\in R^{n_j\mathrm{x}1}$. On the other hand, the second piece of the objective function in formulation (5.7) can be expressed as: $\lambda\,R(\boldsymbol{w}_j)=\lambda\boldsymbol{c}^T\boldsymbol{w}_j$, in which $\boldsymbol{c}=\left[\left\{E\left[L_{j,n_1}\right]-s_{1,n_1}\right\}^2\quad\cdots\quad\left\{E\left[L_{j,\mathrm{n}_m}\right]-s_{m,n_m}\right\}^2\right]^T\in R^{m\mathrm{x}1}$, then $\min_{w_{ij}}\sum_{t=1}^{n_j}\frac{\left(E[L_{j,t}]-E[\tilde{L}_{j,t}]\right)^2}{n_j}+$ $\lambda\,R(\boldsymbol{w}_j)$ can be re-written as: $\min_{\boldsymbol{w}_j}\frac{\boldsymbol{a}^T\boldsymbol{a}-2\boldsymbol{a}^T\boldsymbol{B}\boldsymbol{w}_j+\boldsymbol{w}_j{}^T\boldsymbol{B}^T\boldsymbol{B}\boldsymbol{w}_j}{n_j}+\lambda\boldsymbol{c}^T\boldsymbol{w}_j$. Thus, after rescaling the

objective function by a factor n_j and removing the constant term, the objective function can be thus rewritten into the quadratic optimization formulation in (5.8): $\min_{\boldsymbol{w}_j} \boldsymbol{w}_j{}^T \boldsymbol{B}^T \boldsymbol{B}\boldsymbol{w}_j + \left(\lambda n_j \boldsymbol{c}^T - 2\boldsymbol{a}^T \boldsymbol{B}\right)\boldsymbol{w}_j$, subject to $\boldsymbol{o}^T \boldsymbol{w}_j = 1$ and $\boldsymbol{w}_j \geq \boldsymbol{0}$.

Given that $\boldsymbol{B}^T \boldsymbol{B}$ is a positive semi-definite matrix because $\boldsymbol{x}^T \boldsymbol{B}^T \boldsymbol{B}\boldsymbol{x} = (\boldsymbol{B}\boldsymbol{x})^T \boldsymbol{B}\boldsymbol{x} = \|\boldsymbol{B}\boldsymbol{x}\|_2^2 \geq 0 \; \forall x \in R^{m\text{x}1}$. Therefore, we can conclude that (5.8) is a convex quadratic optimization problem.

Chapter 6

Structural Degradation Modeling Framework for Sparse Datasets with an application on Alzheimer's Disease

Unlike most of the existing literature focus on degradation modeling and prognostics with rich data environments, in this Chapter, we tackle the problem of having a sparse data environment. Specifically, here the sparse data environment refers to the case when there are many units while the information from most of the available units are limited.

6.1 Introduction

Accurate modeling and prediction of the future degradation evolution has been a critically important task in many applications. For example, in manufacturing equipment, an unexpected failure may lead to significant economic losses, production downtime, customer dissatisfaction, and safety issues. Also in health-care applications, an unexpected disease onset may lead to severe medical complexities, ineffective treatment planning, and long-term side effects. Therefore, it is crucial to accurately monitor the health status of a unit (e.g., system, equipment, patient) and understand its degradation process. To achieve this goal, condition-based techniques have been rapidly developed, which aim to fully understand the degradation mechanism of each individual unit so that optimal intervention decisions can be made [82]–[86]. For example, in health-care applications, appropriate implementation of the condition-based techniques can significantly improve early disease diagnosis, treatment effect monitoring and evaluation, and reduce medical and economic costs [87], [88]. Similarly, in manufacturing applications, effectively employing the

condition-based strategy can extend the life of a unit, deliver quantum leaps in productivity, lower the total cost of ownership, enhance equipment safety, and improve operator experience [8], [89]–[91].

Fortunately, the accessibility and development of information technologies (e.g., clouds) have facilitated the collection and storage of information from a massive number of units nowadays, which provides a great opportunity to better understand the degradation processes. For example, the Alzheimer's Disease Neuroimaging Initiative (ADNI) has been collecting longitudinal measurements of biomarkers from hundreds of participating patients. While the data environment is *rich* in the number of studied patients, the available observations from each patient are often quite limited. The "sparse data environments" can be resulted from several reasons such as (i) data loss during transmission; (ii) high cost or limited feasibility of acquiring the biomarkers; and (iii) the patient being a new participant in the study. Unfortunately, most of the existing literature on degradation modeling is not specifically designed for such sparse datasets (Lu and Meeker, 1993; Gebraeel *et al.*, 2005; Peng *et al.*, 2012; Zhou *et al.*, 2012; Chen and Tsui, 2013; K. Liu *et al.*, 2015; Kosasih *et al.*, 2014; Moghaddass and Zuo, 2014).

Specifically, the existing literature on data-driven degradation modeling can be classified into two main categories. The first category directly constructs an individual-level degradation model for a unit using only the available data from that unit such as the least squares model [92] and the maximum likelihood estimation model [94]. These methods often require a large amount of historical data to maintain a certain level of model accuracy. Another category focuses on the population characteristics across a set of units and then leverages the available data from the unit of interest to construct an individual-level degradation model [10], [25], [29]. For example, a

mixed-effects model that describes an entire population was introduced by Lu and Meeker (1993). Then, Gebraeel *et al.* (2005) extended the model by leveraging the available data of an individual unit of interest via Bayesian approaches to produce an individual-level degradation model for that unit. In general, direct implementation of the above-mentioned techniques to the sparse datasets may lead to several issues, including: (i) the high sensitivity of the estimated parameters of the degradation model to the given data; (ii) the lack of interpretability of the degradation model; and (iii) the poor prediction accuracy of the degradation model. In Chapter 2, we discussed in depth the conventional approaches for degradation modeling as well as some recent studies that are related to dealing with the challenges of sparse data environments.

To the best of our knowledge, the existing literature still lacks a reliable degradation modeling approach that is suitable for the units with a limited number of observations. This Chapter seeks to fill this gap and tackle the unique challenges in sparse data environments by developing a structural degradation modeling (SDM) framework. Our approach is inspired by the recommender system, which recommends an item to a user based on (i) other recommenders who rated that item before and (ii) the historical ratings of the user. In particular, first we define each available unit with historical observations as a recommender and the unit of interest as the user; and then we model the degradation status of each interested unit as a combination of the recommenders by taking into account of (i) the available data from the unit of interest; (ii) the population characteristics; (iii) the relationships between the recommenders; and (iv) the precision of the recommenders. Essentially, our proposed method integrates two important ideas: (i) leveraging the available data from the interested unit to improve the modeling fitting of the individual unit over the observed time domain; and (ii) considering the relationship between the available units to

extract proper and accurate population characteristics to address the challenge of limited observations.

The rest of the Chapter is structured as follows. Sections 6.2, 6.3 and 6.4 propose a degradation modeling approach as a recommender system for sparse datasets. Section 6.5 demonstrates the effectiveness of the proposed method regarding degradation modeling based on simulation studies. Section 6.6 demonstrates the proposed method for degradation modeling and prognostics based on the ADNI dataset. Section 6.7 draws a conclusion and discusses future research directions.

6.2 Model Formulation

To adopt the recommender system, we assume all the available units with historical observations excluding the unit of interest as recommenders. Then our main idea is to construct the degradation model of unit j of interest as a combination of the expected values from the recommenders by the following:

$$s_{j,t} = \sum_{i=1}^{m}\left(w_{ij}E\left[s_{i,t}\middle|\boldsymbol{s}_{i,\cdot}\right]\right) + b_j + \epsilon_{j,t}. \tag{6.1}$$

Here, $\sum_{i=1}^{m}\left(w_{ij}E\left[s_{i,t}\middle|\boldsymbol{s}_{i,\cdot}\right]\right) + b_j$ is the predicted degradation status for unit j at time t; $\boldsymbol{s}_{i,\cdot}$ is the vector of available observations for recommender i; m is the number of available recommenders; w_{ij} is the weight of recommender i in construction of the model of unit j and it also quantifies the similarity between the degradation profiles for unit j and recommender i; b_j is a bias term; and $\epsilon_{j,t}$ quantifies the measurement errors and is assumed to follow a normal distribution $N(0, \sigma_j^2)$.

In the above equation, w_{ij} is a key parameter because it quantifies the similarity between the degradation profiles for unit j and recommender i. Once w_{ij} is accurately estimated, we will then be able to leverage the degradation profiles of the recommenders to accurately predict the future

degradation status for the unit of interest. It may be intuitive to calculate w_{ij} via the existing methods such as (i) maximizing the likelihood, or (ii) minimizing the sum of squared errors over the observed time domain (i.e., least squares approach). However, such approaches do not provide any guarantee that w_{ij} quantifies the similarity between the degradation profiles for unit j and recommender i and therefore the interpretability is limited. In addition, for a unit j with limited observations, there is high uncertainty involved in the estimation of w_{ij} as it is only based on the limited observations from unit j.

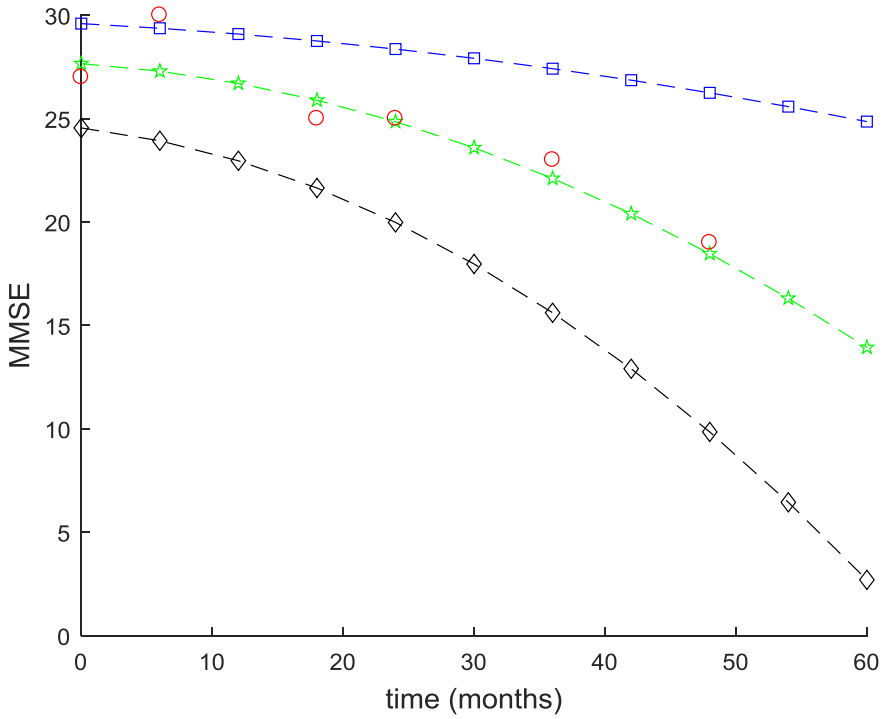

Fig. 6.1. Illustration of the challenges for modeling the AD. Blue squares represent the mild (early-stage), green stars represent the moderate (middle-stage), black diamonds represent the severe (late-stage), and the red circles show the observed biomarker (MMSE) measurements of a patient

In the following, we will first provide an example to illustrate the challenges for modeling the AD with limited observations. Specifically, Figure 6.1 shows in red circles the Mini-Mental Score

Examination (MMSE) measurements for an interested Alzheimer's Disease (AD) patient, in blue squares a recommender for the MMSE of a typical mild (early-stage) AD patient, in green stars a recommender for the MMSE of a typical moderate (middle-stage) AD patient, and in black diamonds a recommender for the MMSE of a typical severe (late-stage) AD patient. Our goal is to estimate the degradation model of the patient of interest using equation (6.1). With the maximum likelihood estimator, the model parameters learned for this example are $\begin{bmatrix} b \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 1.2 \\ 0.86 \\ -1.0 \\ 0.89 \end{bmatrix}$, which minimizes the sum of squared errors. However, this model is limited in interpretation, which cannot provide any insights on the level of sickness. To be specific, the results indicate that the degenerative process of the patient of interest is almost equally close to the degradation processes of the severe patient as well as the mild patient ($w_1 = 0.86$ and $w_3 = 0.89$). In other words, this model indicates that the patient should follow the path of a moderate patient. Contradictorily, the results also show that the patient's degradation model is far from that of the moderate patient ($w_2 = -1$). Furthermore, to understand the sensitivity of the model to newly observed data, we intentionally hide the last observation and derive the model parameters again: $\begin{bmatrix} b \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} -73 \\ 2.8 \\ 1.9 \\ 0.95 \end{bmatrix}$. It can be seen that there is a huge difference in the estimated parameters after hiding the last observation, which indicates that the estimated parameters are highly sensitive to newly observed data and the derived degradation model is not reliable for extrapolation.

6.3 Structural Degradation Modeling (SDM) Framework

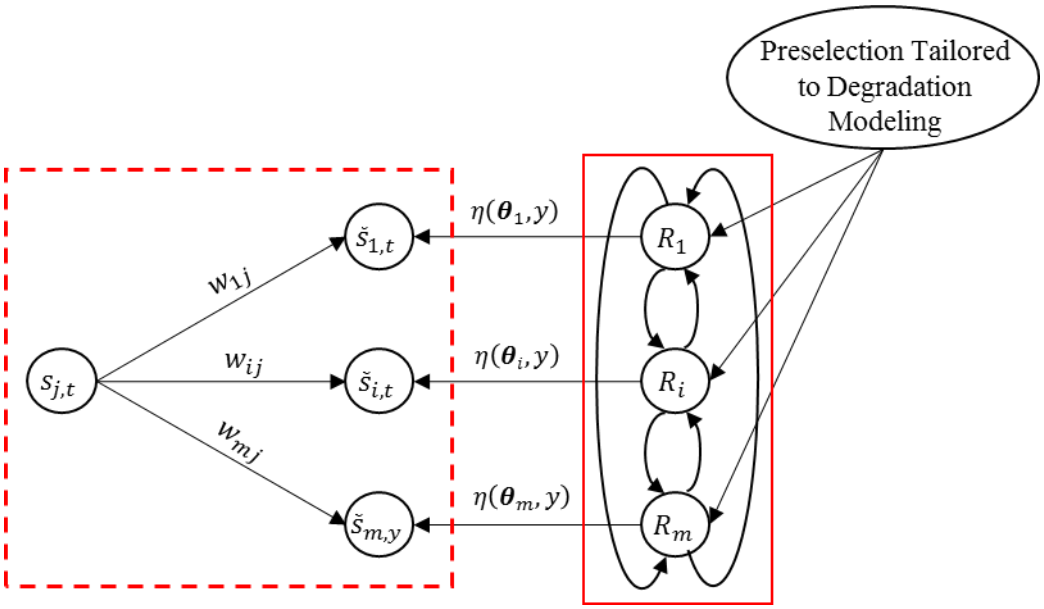

Fig. 6.2. Structural Degradation Modeling (SDM) Framework

To address the challenges in sparse datasets, we propose to model the degradation status of the unit of interest based on the recommender system framework by taking into account of the precision of the recommenders and the relationship between the recommenders as shown in the solid box in Figure 6.2. Similar to the conventional recommender system, our proposed approach allows preselecting recommenders and adding artificial recommenders based on domain knowledge. It is important to note that such a preselection process is critical to decrease the computational efforts for dealing with big data. However, unlike the conventional recommender system, our proposed approach models the degradation profile of each recommender i as $N(\eta(\boldsymbol{\theta}_i, t), \sigma_i^2)$ such that $\eta(\boldsymbol{\theta}_i, t)$ is monotonically increasing or decreasing for $t \geq 0$ based on the domain knowledge of degradation process. Mathematically, (i) $\boldsymbol{\theta}_i$ is the solution of $\min_{\boldsymbol{\theta}_i} \sum_{r=1}^{n_i} \{\eta(\boldsymbol{\theta}_i, t_i(r)) - s_{i,t_i(r)}\}^2$ such that

$\begin{cases} d\eta(\boldsymbol{\theta}_i, t)/dt \geq 0, \ \forall t \geq 0 & \text{if montonically increasing} \\ d\eta(\boldsymbol{\theta}_i, t)/dt \leq 0, \ \forall t \geq 0 & \text{if monotonically decreasing} \end{cases}$; and (ii) $\sigma_i^2 = \frac{\sum_{r=1}^{n_i}\left\{\eta(\boldsymbol{\theta}_i, t_i(r)) - s_{i,t_i(r)}\right\}^2}{\text{degrees of freedom}}$.

For example, if we consider the quadratic model for each recommender that can be written as $s_{i,t} = \theta_{i,0} + \theta_{i.1}t + \theta_{i,2}t^2 + \epsilon_t$ and given that the degradation trends are monotonically decreasing, then (i) $\boldsymbol{\theta}_i$ is the solution of $\min_{\boldsymbol{\theta}_i} \sum_{r=1}^{n_i}\left\{\eta\left(\boldsymbol{\theta}_i, t_i(r)\right) - s_{i,t_i(r)}\right\}^2$ such that $\theta_{i,1} \leq 0$ and $\theta_{i,2} \leq 0$; and (ii) $\sigma_i^2 = \frac{\sum_{r=1}^{n_i}\left\{\eta(\boldsymbol{\theta}_i, t_i(r)) - s_{i,t_i(r)}\right\}^2}{n_i - 3}$.

Without loss of generality, we assume that a set of recommenders $\{1, \dots, m\}$ have been preselected. Then, we propose the following SDM formulation to estimate the parameters w_{ij} and b_j in equation (6.1):

$$\min_{w_{ij}, b_j} \sum_{r=1}^{n_j}\left\{E\left[s_{j,t_j(r)}\right] - s_{j,t_j(r)}\right\}^2 + h\left(n_j\right)\left\{\sum_{i=1}^{m}\sum_{k=1}^{m}\left(w_{ij}d_{ik}w_{kj}\right) + \sum_{i=1}^{m} w_{ij}\sigma_i\sigma_i w_{ij}\right\}$$

$$\text{s.t.} \quad \sum_{i=1}^{m} w_{ij} = 1,$$

$$w_{ij} \geq 0, i = 1,2, \dots, m,$$

$$z_{ik}\left(w_{ij} - w_{kj}\right) = 0, i = 1,2, \dots, m \text{ and } k = 1,2, \dots, m,$$

$$z_{ik} = \begin{cases} 1 & P\left(s_{i,t}|\boldsymbol{s}_i, . = y\right) = P\left(s_{k,t}|\boldsymbol{s}_k, . = y\right) \forall y \in R \\ 0 & \text{otherwise} \end{cases}, \tag{6.2}$$

Here, $E\left[s_{j,t}\right] = \sum_{i=1}^{m}\left(w_{ij}E\left[s_{i,t}|\boldsymbol{s}_{i,\cdot}\right]\right) + b_j$ is the predicted degradation status for unit j at time t, where $s_{i,t}|\boldsymbol{s}_{i,\cdot}$ is the random variable that represents the degradation status of recommender i given its historical degradation information and it is assumed to follow a normal distribution $N(\eta(\boldsymbol{\theta}_i, t), \sigma_i^2)$; z_{ik} is an indicator variable to check if recommenders i and k share the same degradation model; n_j is the number of available observations for unit j; $h\left(n_j\right)$ is a parameter that

depends on n_j; $t_j(r)$ is the time that the r^{th} observation is obtained for unit j; and $d_{ki} = d_{ik} = \sqrt{d_{ik}} * \sqrt{d_{ki}} \geq 0$ is the dissimilarity between recommender i and recommender k. Thus, in the objective function, $w_{ij}d_{ik}w_{kj} = w_{ij}\sqrt{d_{ik}} * \sqrt{d_{ki}}w_{kj}$ is the weighted dissimilarity between recommenders i and k, which depends on the estimated weights w_{ij} and w_{kj}.

Unlike the existing methods in degradation modeling, the SDM framework takes into consideration of all of the following:

(i) ***The precision of the recommenders*** via the term $\sum_{i=1}^{m} w_{ij}\sigma_i\sigma_i w_{ij}$;

(ii) ***The relationship between the recommenders*** via the term $\sum_{i=1}^{m}\sum_{k=1}^{m}\left(w_{ij}d_{ik}w_{kj}\right)$;

(iii) ***The population characteristics*** via modeling the unit of interest *j* as a combination of the recommenders and by the definition of the weight w_{ij} that it quantifies the similarity between the degradation profiles for unit *j* and recommender i.

(iv) ***The individual characteristics*** via minimizing the sum of squared errors $\sum_{r=1}^{n_j}\left\{E\left[s_{j,t_j(r)}\right] - s_{j,t_j(r)}\right\}^2$.

Below we provide detailed discussions for each term considered above. For (i), when predicting the degradation status of unit j, it is preferred to rely more on precise recommenders with low variance. In other words, if σ_i^2 of recommender i is relatively higher with respect to that of other recommenders, then we prefer to assign a lower value for w_{ij}. Therefore, we add the term $\sum_{i=1}^{m} w_{ij}\sigma_i\sigma_i w_{ij}$ to force to choose more precise recommenders.

For (ii), the similarity measure is a commonly used technique for clustering and classification such as DBSCAN [96], k-NN [97], k-means [98], [99], ratio-cut [100]. Here, in addition to

constructing an accurate model based on the available observations from the unit of interest, we also try to set $w_{ij} \geq 0$ for recommenders that are highly similar (i.e., recommenders that belong to the same cluster). Note that the weighted dissimilarity is important for controlling the stability in the constructed degradation model. Considering two non-similar recommenders i and k (i.e., d_{ik} is large), a large value in $w_{ij}d_{ik}w_{kj}$ means that large weights w_{ij} and w_{kj} are assigned to construct the degradation model for unit j, which thus leads to an uninterpretable and unstable model similar to the case in the motivation example in Figure 6.1 (w_1 and w_3 are very large and close to each other). To avoid this issue, we propose to minimize the overall pairwise weighted dissimilarity between the recommenders, $\sum_{i=1}^{m}\sum_{k=1}^{m}(w_{ij}d_{ik}w_{kj})$, to enhance the stability and interpretation of the constructed degradation model. Ideally, we prefer $w_{ij}d_{ik}w_{kj} \to 0$ when recommenders i and k do not share similar degradation characteristics (i.e., d_{ik} is large). This is equivalent to mitigating the contribution of recommender i (i.e., $w_{ij} \to 0$), or recommender k (i.e., $w_{kj} \to 0$), or both recommenders (i.e., $w_{ij} \to 0$ and $w_{kj} \to 0$). Regarding the dissimilarity metric d_{ii}, it is commonly set to 0 because there is no dissimilarity between the recommender and itself. However, in this Chapter, we propose a modification of $d_{ii} = \sum_{\substack{k=1 \\ k\neq i}}^{m} d_{ik}$. The rationale is that if recommender i was mistakenly chosen, then it adds up to the uncertainty of the constructed model. This uncertainty would be quantitatively high if recommender i is totally off from the true degradation model, which is estimated as a combination of the remaining recommenders' models. Accordingly, if recommender i is highly non-similar from the remaining recommenders, then it is risky to include recommender i in the constructed model. Therefore, we propose setting $d_{ii} =$

$\sum_{\substack{k=1 \\ k\neq i}}^{m} d_{ik}$ which is the sum of dissimilarities between recommender i and the remaining recommenders. For conducting the dissimilarity measure, extensive studies have been done in the literature [101]. Without loss of generality, in the Chapter, we consider the square of the Mahalanobis distance, $d_{ki} = (\boldsymbol{\theta}_i - \boldsymbol{\theta}_k)^T \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\boldsymbol{\theta}_i - \boldsymbol{\theta}_k)$ for $i \neq k$ to measure the dissimilarity between recommenders i and k. Here, $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ is the covariance matrix of $\boldsymbol{\theta}$ and it can be estimated by the sample covariance matrix from the parameters of the recommenders.

On the other hand, for (iv), $\sum_{r=1}^{n_j} \left(E\left[s_{j,t_j(r)} \right] - s_{j,t_j(r)} \right)^2$ is the sums of squared errors over the observed time domain, which is used to ensure the constructed model accurately characterizes the degradation evolution of interested unit j. In the objective function, the weighted dissimilarity $\sum_{i=1}^{m} \sum_{k=1}^{m} \left(w_{ij} d_{ik} w_{kj} \right) + \sum_{i=1}^{m} w_{ij} \sigma_i \sigma_i w_{ij}$ is more focused on the population characteristics, whereas the sum of squared errors $\sum_{r=1}^{n_j} \left(E\left[s_{j,t_j(r)} \right] - s_{j,t_j(r)} \right)^2$ is more focused on the individual characteristics of the unit of interest. $h(n_j)$ balances this trade-off. In this Chapter, we propose to consider $h(n_j)$ as a decreasing positive function with respect to n_j. This is because when there are more observations available from unit j, we will be more confident to rely on the individual observations to characterize the unit of interest. Specifically, as an illustration, we set $h(n_j) = \lambda / n_j$, where λ is a tuning parameter and we will discuss how to estimate λ in details in Section 6.4.

Finally, by the definition of w_{ij} that it quantifies the similarity between the degradation profiles for unit j and recommender i, the following constraints should be satisfied: (i) $w_{ij} \geq 0$ for $i = 1, \ldots, m$; and (ii) weights from equivalent recommenders should be equal, which mathematically

can be written as $z_{ik}(w_{ij} - w_{kj}) = 0$ such that $z_{ik} = 1$ if recommenders i and k are identical; otherwise, $z_{ik} = 0$. Since we are only given a finite set of recommenders, we focus on the normalized similarity measure by forcing $\sum_{i=1}^{m} w_{ij} = 1$.

With all the above efforts, we expect the proposed model to be more interpretable and robust to over-fitting. As an illustration, we apply the proposed framework to the example in Figure 6.1. The results show that the optimal solution is $\begin{bmatrix} b \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 0.97 \\ 0.09 \\ 0.77 \\ 0.14 \end{bmatrix}$ before hiding the last observation; and $\begin{bmatrix} b \\ w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 0.88 \\ 0.10 \\ 0.78 \\ 0.12 \end{bmatrix}$ after hiding the last observation. From these results, we can see that the proposed method has a nice interpretation that the patient is more likely to have a moderate AD status. In addition, the weights do not vary much compared to the least squares approach before and after hiding the last observation. In Sections 6.5 and 6.6, we will further thoroughly study our proposed method under different scenarios.

For simplicity, we can rewrite formulation (6.2) to the following:

$$\min_{\boldsymbol{w}_j} \ \boldsymbol{w}_j{}^T \boldsymbol{\Psi}^T \boldsymbol{\Psi} \boldsymbol{w}_j - 2\boldsymbol{s}_{j,\cdot}{}^T \boldsymbol{\Psi} \boldsymbol{w}_j + \boldsymbol{s}_{j,\cdot}{}^T \boldsymbol{s}_{j,\cdot} + h(n_j) \boldsymbol{w}_j{}^T \boldsymbol{D}^* \boldsymbol{w}_j$$

$$\text{s.t.} \quad \boldsymbol{o}^T \boldsymbol{w}_j = 1$$

$$\boldsymbol{A} \boldsymbol{w}_j \geq \boldsymbol{0}, \tag{6.3}$$

where $\boldsymbol{s}_{j,\cdot} = \left[s_{j,t_j(1)}, s_{j,t_j(2)}, \ldots, s_{j,t_j(n_j)} \right]^T$ is the vector of available observations for operating unit j; $\boldsymbol{w}_j = \left[b_j, w_{1j}, \ldots., w_{mj} \right]^T \in R^{(m+1)\times 1}$ is a vector that contains both the bias term and the

weights; $\boldsymbol{\Psi} = \begin{bmatrix} 1 & \eta\left(\boldsymbol{\theta}_1, t_j(1)\right) & \dots & \eta\left(\boldsymbol{\theta}_m, t_j(1)\right) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \eta\left(\boldsymbol{\theta}_1, t_j(n_j)\right) & \dots & \eta\left(\boldsymbol{\theta}_m, t_j(n_j)\right) \end{bmatrix} \in R^{n_j \times (m+1)}$ is the design matrix; $\boldsymbol{o} = [0, 1, \dots, 1]^T \in R^{(m+1)\times 1}$ is a vector containing all ones except for the first entry to be 0; and $\boldsymbol{A} = diag(\boldsymbol{o})$ is a diagonal matrix with a diagonal vector $\boldsymbol{o}$; and $\boldsymbol{D}^* = \begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & d_{11} + \sigma_1^2 & \dots & d_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & d_{m1} & \dots & d_{mm} + \sigma_m^2 \end{bmatrix} \in R^{(\mathrm{m}+1)\times(\mathrm{m}+1)}$ is the augmented dissimilarity matrix with $d_{ii} = \sum_{\substack{k=1 \\ k\neq i}}^{m} d_{ik}$ and $d_{ki} = d_{ik} \geq 0$. The constraints $z_{ik}\left(w_{ij} - w_{kj}\right) = 0, i = 1,2, \dots, m$ and $k = 1,2, \dots, m$ as well as $z_{ik} = \begin{cases} 1 & \text{if recommenders i and k are identical} \\ 0 & \text{otherwise} \end{cases}$ are removed by the results of Lemma 6.1.

In Lemma 6.1, we further prove that any two equivalent recommenders will have the same weights. This Lemma (i) allows removing the last two constraints in formulation (6.2) and maintaining the definition of w_{ij}; and (ii) helps reducing the number of unknown variables by substituting w_{k^*j} with w_{i^*j} if $\theta_{i^*} = \theta_{k^*}$ and $\sigma_{i^*}^2 = \sigma_{k^*}^2$.

Lemma 6.1: If recommenders i^* and k^* are identical (i.e., $\theta_{i^*} = \theta_{k^*}$ and $\sigma_{i^*}^2 = \sigma_{k^*}^2$), then the resulting weights w_{i^*j} and w_{k^*j} from the SDM formulation (6.3) are also identical (see Appendix for details).

In Lemma 6.2, we prove that the proposed SDM formulation (6.3) is convex. Therefore, the formulation can be efficiently solved at a low computational cost by many existing solvers [73].

Lemma 6.2: The SDM formulation (6.3) is convex (see Appendix for details).

6.4 Evaluation of the Tuning Parameter

Given a dataset $\mathfrak{D}$, we randomly partition the dataset into a validation dataset $\mathfrak{D}_v$ and a training dataset $\mathfrak{D}_t$. Then, we follow Algorithm 1 below to estimate the tuning parameter λ, which aims to minimize the validation error. For demonstration purposes, we define the validation error by the 2nd norm of the difference between the predicted and true degradation measurements of the validation dataset $e = \left\|\mathfrak{D}_v - \widehat{\mathfrak{D}}_v\right\|_2 = \sqrt{\sum_{r=1}^{n_j^v}\left(E\left[s_{j,t_j(r)}\right] - s_{j,t_j(r)}\right)^2}$, where n_j^v is the number of validation observations from unit j. The algorithm is iterative and terminates either after exceeding a predefined maximum number of iterations or when the difference in the validation accuracy between two consecutive iterations is below a certain threshold err (i.e., $|e - e_o| < err$). At each iteration, λ gets updated (by $\alpha * \lambda$, where α is the step size parameter) if a better performance is achieved; otherwise, we shrink the step size by a factor greater than 1 (e.g., $\alpha = \alpha/1.5$). In the simulation and case studies, we only utilize Algorithm 1 once to calculate the tuning parameter λ; however, it is possible to run the algorithm at different initial values of λ for a better estimation of the tuning parameter.

ALGORITHM 1: ESTIMATION OF THE TUNING PARAMETER

1.	Define the max allowable iterations i_{max}, the initial value of λ, a large initial value of validation error $e_o = 1000$, the minimally allowable difference in the validation error err between two consecutive iterations for termination, and an initial step size $\alpha = 2$.
2.	For $iter = 1$ *and* $iter < i_{max}$ (LOOP 1)
3.	$\lambda = \alpha * \lambda$
4.	For $i = 1$ *and* $i < m$: Loop over all the units (LOOP 2)
5.	Calculate $\boldsymbol{w}_j$ via the SDM framework given the training dataset $\mathfrak{D}_t$

6.	Calculate the validation error $e = \left\|\mathfrak{D}_v - \widehat{\mathfrak{D}}_v\right\|_2$
7.	IF $\|e - e_o\| < err$
8.	Stop and return λ
9.	END IF
10.	IF $e_o > e$ (then update e_o)
11.	$e_o = e$
12.	ELSE (there is no improvement because the step size α is large)
13.	$\lambda = \lambda/\alpha$ (do not update λ)
14.	$\alpha = \alpha/1.5$ (shrink the step size to reach a local minimum)
15.	END IF
16.	END LOOP 2
17.	END LOOP 1

6.5 Simulation Studies

In this section, we investigate the performance of the proposed SDM via simulation studies and compare them to the following benchmark methods: 1) Recommender system Model (RM), 2) Bayesian Mixed-Effects Model (BMEM), 3) Least Squares Model (LSM), 4) collaborative multi-output Gaussian process with no basis functions for the outputs (MOGP), and 5) collaborative multi-output Gaussian process with quadratic basis functions for the outputs (MOGP-Q). For the RM, we estimate w_{ij} by its restricted maximum likelihood estimator subject to $\sum_{i=1}^{m} w_{ij} = 1$ and $w_{ij} \geq 0$ for $i = 1, 2, .., m$. In our simulation studies, we focus on the AD application and simulate the MMSE measurements for AD patients. The MMSE ranges between 0 (worst condition) and 30 (best condition) and it is expected to decrease with time for all patients. In total 100 patients were simulated over a period of 36 months with measurements taken monthly. The MMSE measurement

for each simulated patient is assumed to follow a 2nd order polynomial model [102], [103]:

$$M_{i,t} = \theta_{i,0} + \theta_{i.1}t + \theta_{i,2}t^2 + \epsilon_{i,t}, \tag{6.4}$$

where $M_{i,t}$ is the simulated MMSE measurement for patient i at time t; $\boldsymbol{\theta}_i = [\theta_{i,0}, \theta_{i,1}, \theta_{i,2}]^T$ are the parameters for patient i and assumed to follow the normal distribution $N_3(\boldsymbol{u}, \boldsymbol{\Sigma})$ with $\boldsymbol{u} = [23;\ -0.05; -0.005]^T$ and $\boldsymbol{\Sigma} = \begin{bmatrix} 10 & -10^{-4} & -10^{-5} \\ -10^{-4} & 10^{-4} & 10^{-5} \\ -10^{-5} & 10^{-5} & 10^{-5} \end{bmatrix}$; $\epsilon_{i,t}$ represents the random noise in the MMSE measurements and is assumed to follow a normal distribution $N(0, \sigma_i^2)$; and t is the time in months. Figure 6.3 shows the underlying degradation models for AD patients over 36 months.

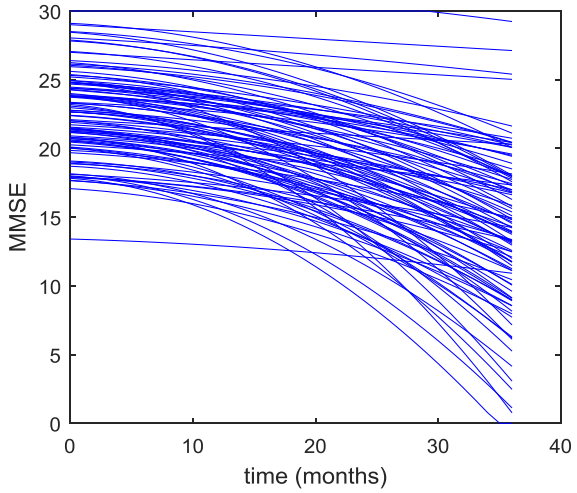

Figure 6.3. The underlying degradation models for AD patients over 36 months

In our simulation studies, 9 different scenarios are considered with two varying factors: (i) the variance of the random noise σ_i^2 with values 1, 9 and 16; and (ii) the percentage of missing data (sparsity level) by randomly hiding 20%, 50% and 80% of the observations from the full dataset. For each scenario, we (i) calculate the tuning parameter based on Section 6.4; and (ii) evaluate the metrics in equations (6.5) and (6.6) below.

(i) For each simulated patient j, we define the mean square error (MSE) of the hidden observations as the following:

$$MSE(j) = \frac{1}{n_j^h}\sum_{t=1}^{n_j^h}\left(M_{j,t} - \widehat{M}_{j,t}\right)^2, \tag{6.5}$$

where n_j^h is the number of hidden measurements for patient j; and $\widehat{M}_{j,t}$ is the estimated measurement for patient j at time t (i.e., $\widehat{M}_{j,t} = \hat{\theta}_{j,0} + \hat{\theta}_{j,1}t + \hat{\theta}_{j,2}t^2$ via the LSM and BMEM approaches, and $\widehat{M}_{j,t} = \sum_{i=1}^{m} w_{ij}\widehat{M}_{i,t} + b_j$ using the SDM and RM frameworks).

(ii) Since the true parametric simulation model is known, we can also calculate the mean square error (pMSE) of the parameters for patient j as the following:

$$pMSE(j) = \frac{(\boldsymbol{\theta}_j - \widehat{\boldsymbol{\theta}}_j)^T(\boldsymbol{\theta}_j - \widehat{\boldsymbol{\theta}}_j)}{p+1}, \tag{6.6}$$

where p is the order of the polynomial degradation model and it is equal to 2 in our simulation study; and $\widehat{\boldsymbol{\theta}}_j = \left[\hat{\theta}_{j,0}, \hat{\theta}_{j,1}, \hat{\theta}_{j,2}\right]^T$ is the estimated set of degradation parameters for patient j based on the available observations (i.e., excluding the hidden observations). Note that even if a constructed model fits the available observations of a unit accurately but fails to estimate the true degradation model, it is expected that the future predictions from the constructed model will still be inaccurate. Therefore, here, we utilize pMSE to measure the long-term prediction performance of the constructed model, such that a lower error in the model parameters is expected to produce better predictions in the future. On the contrary, the MSE metric in equation (6.5) characterizes the prediction performance within the observation window.

Table 6.1 summarizes the performance comparisons regarding MSE, which are based on the hidden observations. For a better visual illustration, Figure 6.4 shows the sample mean of the MSE for the scenarios with a noise variance equals to 9.

Table 6.1. The sample mean (μ_{MSE}) and sample standard deviation (σ_{MSE}) of the MSE under different simulated scenarios with the best performing model in bold

Noise Variance	Missing Data (%)	SDM $[\mu_{MSE}, \sigma_{MSE}]$	RM $[\mu_{MSE}, \sigma_{MSE}]$	BMEM $[\mu_{MSE}, \sigma_{MSE}]$	LSM $[\mu_{MSE}, \sigma_{MSE}]$	MOGP $[\mu_{MSE}, \sigma_{MSE}]$	MOGP-Q $[\mu_{MSE}, \sigma_{MSE}]$
1	20	[0.25, 0.15]	[0.25, 0.15]	[0.25, 0.15]	[0.26, 0.16]	[0.73, 0.81]	**[0.25, 0.15]**
1	50	[1.13, 0.60]	[1.13, 0.58]	**[1.12, 0.56]**	[1.19, 0.63]	[3.09, 3.52]	[1.13, 0.58]
1	80	**[7.38, 10.0]**	[9.67, 22.1]	[7.96, 10.9]	[22.5, 83.7]	[21.3, 25.0]	[9.74, 22.4]
9	20	**[2.24, 1.51]**	[2.26, 1.51]	[2.27, 1.52]	[2.35, 1.53]	[2.73, 2.08]	[2.28, 1.51]
9	50	**[9.83, 4.16]**	[10.1, 4.42]	[10.1, 4.22]	[11.0, 5.02]	[11.3, 5.68]	[10.2, 4.44]
9	80	**[48.5, 30.0]**	[54.9, 45.6]	[55.7, 32.4]	[107, 201]	[57.8, 34.3]	[55.0, 46.6]
16	20	**[3.96, 2.67]**	[4.01, 2.68]	[4.03, 2.71]	[4.16, 2.72]	[4.44, 3.19]	[4.04, 2.68]
16	50	**[17.3, 7.17]**	[17.9, 7.80]	[17.8, 7.37]	[19.5, 8.96]	[18.6, 8.21]	[18.0, 7.83]
16	80	**[84.1, 48.8]**	[95.1, 73.7]	[96.1, 53.6]	[166, 245]	[91.5, 48.8]	[95.4, 75.8]

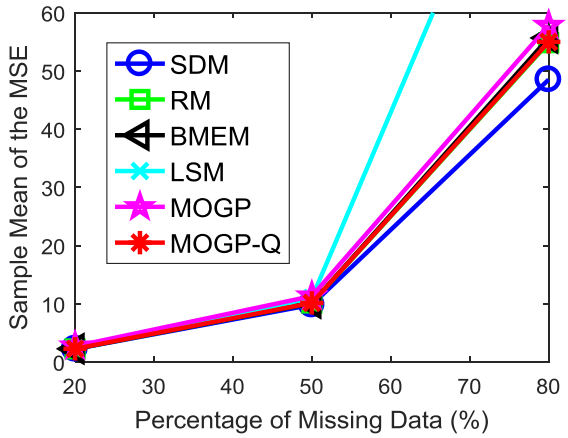

Figure 6.4. Sample mean of the MSE when the noise variance equals to 9

From Table 6.1, we can see that the SDM shows a relatively better performance compared to the benchmark methods at high percentages of missing data. This is an expected consequence because (i) the constructed models from the LSM and BMEM become unstable and highly sensitive to the variance of the available data; and (ii) the existing MOGPs may falsely characterize

the correlation between the units due to the limited available data. However, the SDM framework (i) focus on precise recommenders to achieve a less sensitive degradation model; and (ii) considers the relationships between the recommenders, which stabilizes the constructed degradation model. As a conclusion from the MSE metric results, the proposed SDM framework is expected to have a better performance than the benchmark methods for sparse datasets and/or in presence of high variance noise. Note that noisy datasets (i.e., high variance datasets) share similar challenges with sparse datasets such as potential over-fitting and lack of interpretability. This is because noisy datasets require more observations than a non-noisy dataset to achieve a comparable model fitting performance.

Figure 6.4 provides a better illustration which clearly shows that the SDM outperforms the benchmark methods at high percentages of missing data. Note that the MSE metric is based on the error in the randomly hidden observations, therefore it measures the interpolation performance of the methods because those hidden observations are scattered and not concentrated in a specific time window. To better understand the predictive performance for the long run, we further evaluate the results of the pMSE metric, which are present in Table 6.2. Here, we do not show the results from the MOGPs because they are based on data and do not give estimates for the parametric model. For a better visual illustration, Figure 6.5 shows the sample mean value of the pMSE for the scenarios with a noise variance equals to 9.

From Table 6.2, we can see that (i) the proposed SDM framework outperforms the benchmark methods in all of the scenarios; and (ii) the benchmark methods perform poorly at high percentages of missing data. The first observation is expected because the SDM framework considers the relationship between the recommenders, which ensures the stableness of the constructed

degradation model even at high levels of missing data. The second observation stems from the facts that (i) the RM and LSM tend to over-fit the limited observations, which thus leads to a good performance over the observed time domain but not for future predictions; and (ii) in presence of limited observations, the BMEM tends to focus more on the population characteristics that may not accurately realize the unique characteristics of the interested unit. As a conclusion, this simulation study clearly shows the advantage of our proposed method for sparse data environments where there exists many units that have limited number of available observations.

Table 6.2. The sample mean (μ_{pMSE}) and sample standard deviation (σ_{pMSE}) of the pMSE under different simulated scenarios with the best performing model in bold

Noise Variance	Missing Data (%)	SDM $[\mu_{pMSE}, \sigma_{pMSE}]$	RM $[\mu_{pMSE}, \sigma_{pMSE}]$	BMEM $[\mu_{pMSE}, \sigma_{pMSE}]$	LSM $[\mu_{pMSE}, \sigma_{pMSE}]$
1	20	**[0.030, 0.038]**	[0.061, 0.070]	[0.043, 0.050]	[0.097, 0.11]
1	50	**[0.055, 0.078]**	[0.10, 0.14]	[0.089, 0.11]	[0.21, 0.31]
1	80	**[0.17, 0.24]**	[0.28, 0.49]	[0.56, 2.86]	[5.59, 35.4]
9	20	**[0.28, 0.37]**	[0.45, 0.64]	[0.35, 0.42]	[0.83, 1.02]
9	50	**[0.38, 0.47]**	[0.72, 1.03]	[0.63, 0.76]	[1.53, 2.29]
9	80	**[0.81, 1.06]**	[1.59, 2.33]	[3.49, 5.35]	[20.9, 119]
16	20	**[0.44, 0.57]**	[0.74, 1.08]	[0.59, 0.71]	[1.48, 1.82]
16	50	**[0.56, 0.70]**	[1.19, 1.64]	[1.07, 1.27]	[2.73, 4.07]
16	80	**[1.31, 1.69]**	[2.59, 3.66]	[6.13, 9.33]	[37.2, 212]

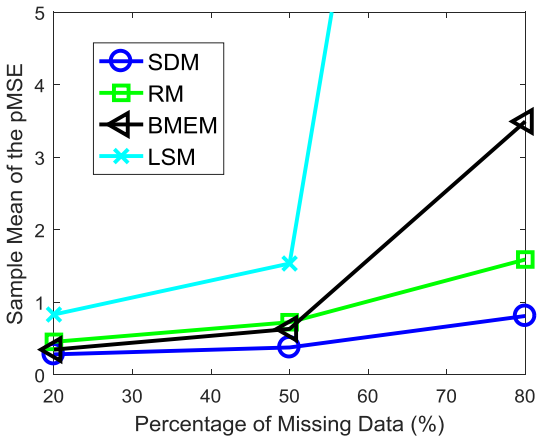

Figure 6.5. Sample mean of the pMSE when the noise variance equals to 9

6.6 Case Study

This section further investigates the performance of the proposed SDM framework based on the ADNI dataset [104]. We only consider the RM, BMEM, and LSM as the three benchmark methods and we exclude the MOGPS. This is because there are little shared observational time points between the patients, which results in ill-conditioned MOGPS as mentioned in the literature review. The dataset contains longitudinal measurements of examinations and biomarkers, and personal information for the participating patients. Here, we focus on the MMSE because it has been commonly used to predict the AD status of a patient in practice. Furthermore, we only consider the subset of participating patients that have 4 or more MMSE measurements. This requirement is needed to ensure the construction of valid models via the LSM and BMEM approaches. For a better visualization, Figure 6.6 shows the MMSE degradation curves for a subset of the patients. We can observe that different patients may have a different number of available observations and the observations are collected at different time points. In addition, the dataset of MMSE measurements is sparse in nature because (i) MMSE measurement is recorded semi-annually for each patient; (ii) patients may join the program at different AD stages; and (iii) some patients may skip the semi-annual visit to the clinic at some time points. Thus the given MMSE dataset presents the real challenge in practice and also provides a good example to test the efficacy of the proposed structural model for sparse datasets.

In this case study, we model the MMSE measurements as a 2nd order polynomial model [102], [103]:

$$M_{i,t} = \theta_{i,0} + \theta_{i.1}t + \theta_{i,2}t^2 + \epsilon_{i,t}, \tag{6.7}$$

where $M_{i,t}$ is the observed MMSE measurement for patient i at time t under the proposed model; and $\boldsymbol{\theta}_i =$

$[\theta_{i,0}, \theta_{i,1}, \theta_{i,2}]^T$ are the parameters of the degradation model for patient i. Here, to measure the prognostic performance of the degradation models, we conduct a leave one patient out cross-validation. For each validating patient, we hide the last two observations and calculate the root of the squared difference (rSD) between the predicted and true measurements for each of the hidden observations. Mathematically, rSD is defined as:

$$rSD(i,t) = \sqrt{\left(M_{i,t} - \widehat{M}_{i,t}\right)^2}, \tag{6.8}$$

where $\widehat{M}_{i,t} = \hat{\theta}_{i,0} + \hat{\theta}_{i.1}t + \hat{\theta}_{i,2}t^2$ is the estimated MMSE measurement for patient i at time t.

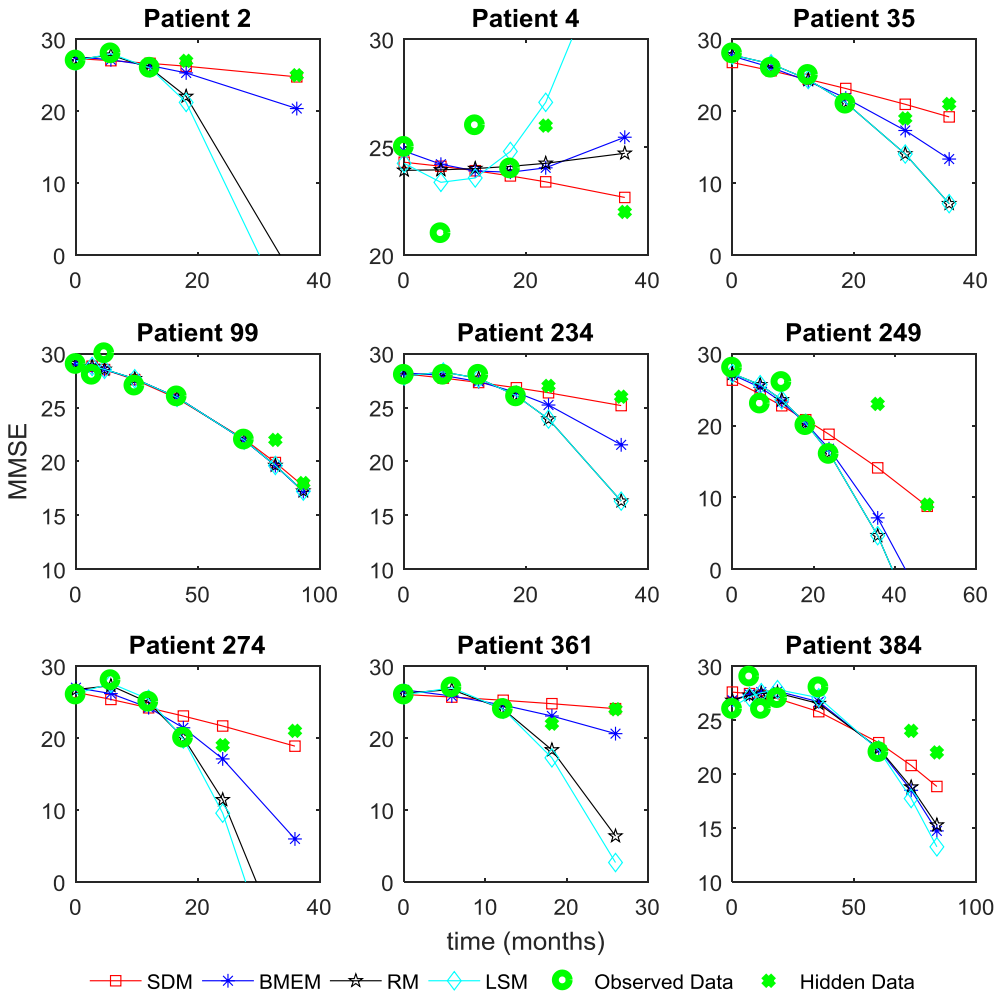

Fig. 6.6. The true and predicted MMSE measurements from a subset of patients

6.6.1 Prognostic Performance

To better visualize and compare the prognostic performance of the proposed SDM framework and the benchmark methods, Figure 6.6 shows the prediction results of the MMSE for some patients, and Figures 6.7 and 6.8 show the boxplots of the root of the squared difference, rSD, of the two hidden measurements for all the considered participating patients. Similar to the simulation study, we applied the proposed algorithm in Section 6.4 and found $\lambda = 4.16$.

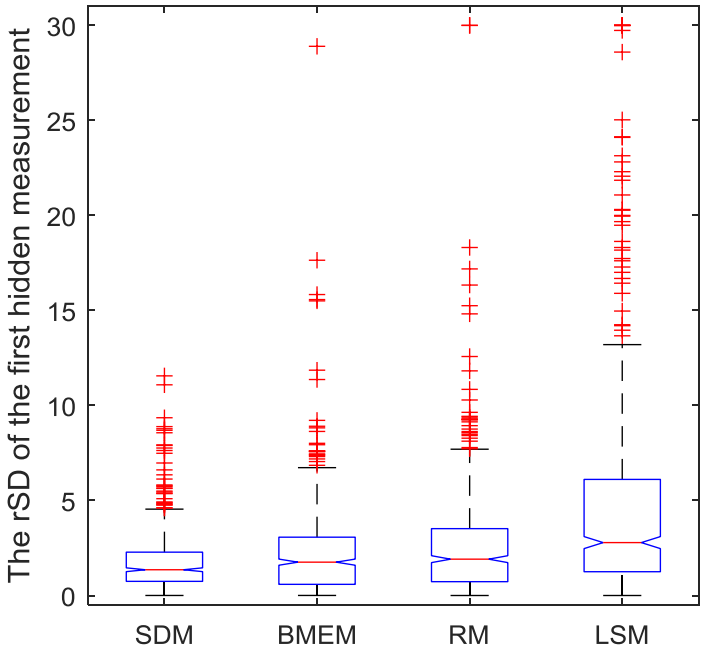

Fig. 6.7. The boxplot for the rSD of the first hidden measurement

From Figures 6.6, 6.7 and 6.8, we can see that (i) the SDM framework results in lower means and lower variances for the rSD of both hidden measurements compared to the benchmark methods; and (ii) the SDM method shows more stable results and smaller increases in the errors from the first hidden measurement to the second hidden measurement than the benchmark methods. The first observation stems from the fact the SDM framework takes into consideration

of (i) the relationship between the recommenders and (ii) the precision of the recommenders. The second observation shows the importance of considering (i) the relationship between the user and the recommenders and (ii) the pairwise relationships between the recommenders to construct an accurate degradation model that shows consistently satisfactory prognostic performance. In summary, this case study further validates our conclusion that for sparse datasets, the proposed SDM framework performs better than the benchmark methods.

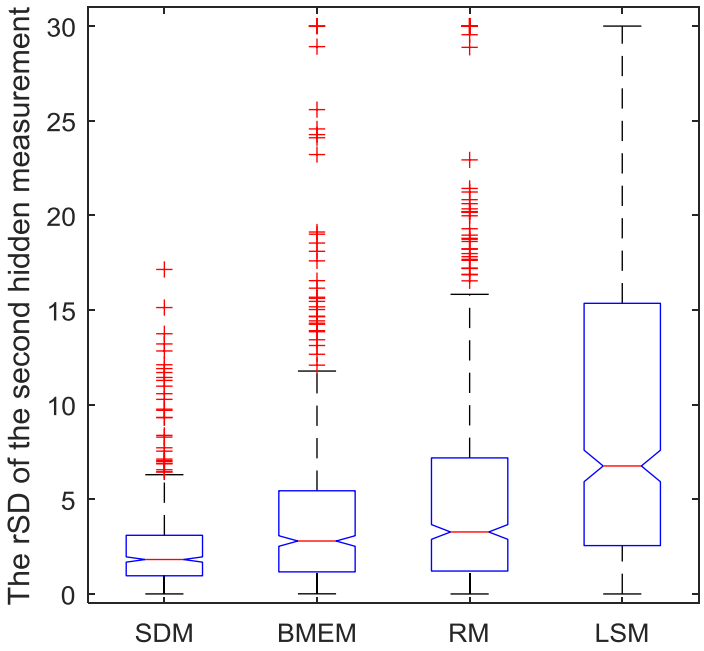

Fig. 6.8. The boxplot for the rSD of the second hidden measurement

6.6.2 Tuning Parameter

In this subsection, we conduct a sensitivity analysis for the tuning parameter λ in the SDM framework. Recall that in the case-study of Section 6.6.1, we applied the proposed algorithm in Section 6.4 and found $\lambda = 4.16$. To better understand the sensitivity of the proposed method on

the tuning parameter, we repeated the prognostic analysis for different values of λ and the results are summarized in Figures 6.9 and 6.10.

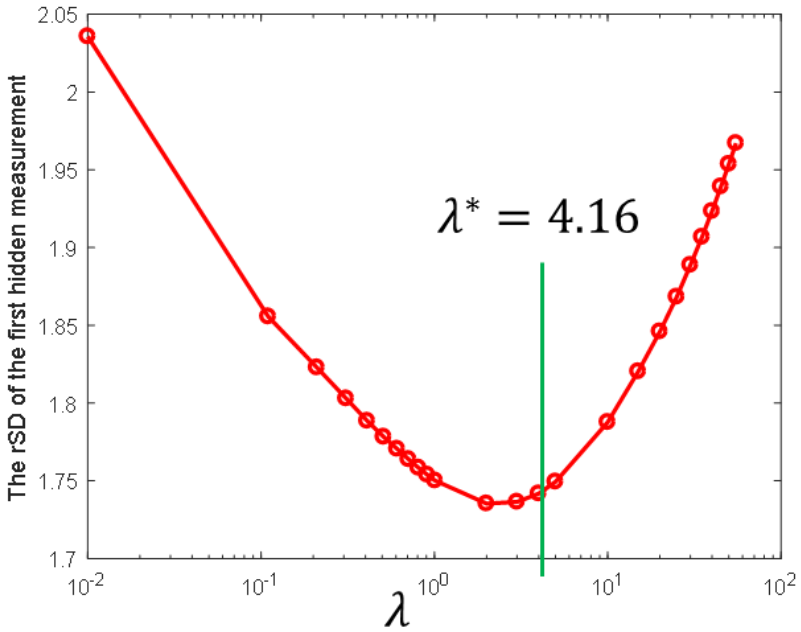

Fig. 6.9. The rSD of the first hidden observation at different values of the tuning parameter λ

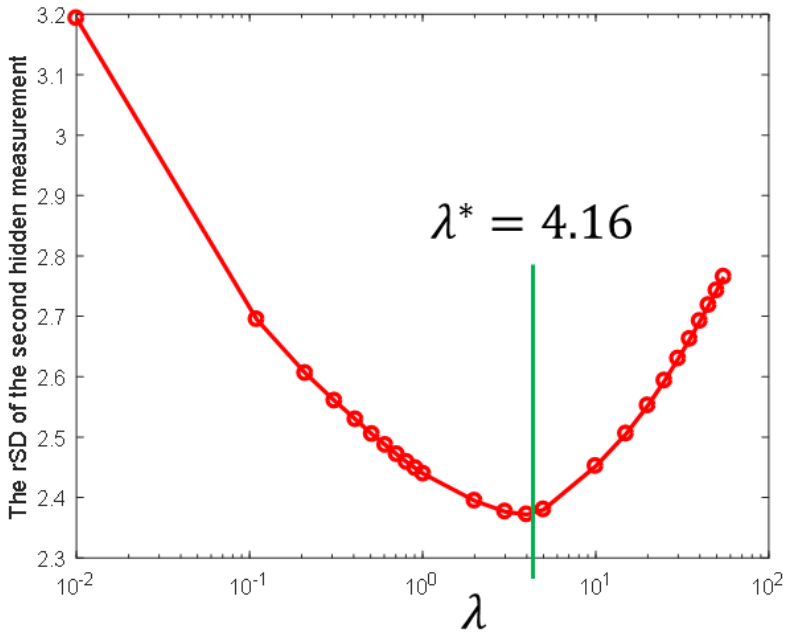

Fig. 6.10. The rSD of the second hidden observation at different values of the tuning parameter λ

From Figures 6.9 and 6.10, we can see that for this case study (i) the rSD is smooth and convex with respect to λ; and (ii) the value of λ evaluated by the algorithm in Section 6.4 is very close to the optimal value of λ that minimizes the rSD of the hidden measurements. The first observation suggests that for this case study reaching a local minimum is sufficient to find the optimal value of λ. The second observation shows the efficacy of the algorithm proposed in Section 6.4.

6.7 Conclusions

Predicting and modeling the progression of degradation is important and critical to a wide set of applications and industries. The development of information technologies (e.g., clouds) have facilitated the collection and storage of information from a massive number of operating units. While the number of recorded units can be large, many units often have limited available observations due to different reasons such as data loss during transmission, and high cost or limited feasibility of data acquisition. Direct implementation of the existing degradation modeling approaches in such sparse data environments may lead to (i) a low model accuracy for prediction; (ii) a high uncertainty of the model parameters; and (iii) a lack of interpretability of the model.

This Chapter aims to fill the literature gap by developing an SDM framework that addresses the unique challenges of the sparse data environments. Specifically, the degradation model of each individual unit is structured as a combination of a set of recommenders by taking into consideration of (i) the available data from the unit of interest; (ii) the relationship between the recommenders; (iii) the precision of the recommenders based on past performance; and (iv) the population characteristics. The developed framework is tested and validated by simulation studies as well as

the ADNI dataset and the results show that the SDM framework outperforms the benchmark methods for degradation modeling.

There are several important topics for future research: First, it is important to further explore non-parametric approaches beyond the discussed MOGPs to model the degradation status of the recommenders because, for some applications, it is complicated to learn the true parametric functional form of the degradation profile. Second, in this Chapter, we consider $h(n_j) = \lambda/n_j$ for demonstration. In future studies, it would be interesting to investigate more on the choice of $h(n_j)$.

6.8 Appendix

Proof for Lemma 6.11 that formulation (6.3), $\min_{\boldsymbol{w}_j} \boldsymbol{w}_j{}^T \boldsymbol{\Psi}^T \boldsymbol{\Psi} \boldsymbol{w}_j - 2\boldsymbol{s}_{j,\cdot}{}^T \boldsymbol{\Psi} \boldsymbol{w}_j + \boldsymbol{s}_{j,\cdot}{}^T \boldsymbol{s}_{j,\cdot} + h(n_j) \boldsymbol{w}_j{}^T \boldsymbol{D}^* \boldsymbol{w}_j$ subject to $\boldsymbol{o}^T \boldsymbol{w}_j = 1$ and $\boldsymbol{A}\boldsymbol{w}_j \geq \boldsymbol{0}$, is convex.

Since $h(n_j) \geq 0$ by definition, then it is sufficient to prove that $\boldsymbol{\Psi}^T \boldsymbol{\Psi}$ and $\boldsymbol{D}^*$ are positive semi-definite (PSD) matrices. First, $\boldsymbol{\Psi}^T \boldsymbol{\Psi}$ is PSD because for any non-zero vector $\boldsymbol{w}_j$, we have $\boldsymbol{w}_j{}^T \boldsymbol{\Psi}^T \boldsymbol{\Psi} \boldsymbol{w}_j = \|\boldsymbol{\Psi} \boldsymbol{w}_j\|_2^2 \geq 0$. Second, we proof that $\boldsymbol{D}^*$ is PSD by showing that $\boldsymbol{w}_j{}^T \boldsymbol{D}^* \boldsymbol{w}_j \geq 0$ for any $\boldsymbol{w}_j$.

The dissimilarity matrix can be written as $\boldsymbol{w}_j{}^T \boldsymbol{D}^* \boldsymbol{w}_j = \left(\sum_{i=1}^m \sum_{k=1}^m w_{ij} d_{ik} w_{kj}\right) = \sum_{i=1}^m w_{ij}^2 \sigma_i^2 + \sum_{i=1}^m w_{ij}^2 d_{ii} + \sum_{i=1}^m \sum_{\substack{k=1 \\ k \neq i}}^m w_{ij} d_{ik} w_{kj}$. Recall that $d_{ii} = \sum_{\substack{k=1 \\ k \neq i}}^m d_{ik}$, then $\boldsymbol{w}_j{}^T \boldsymbol{D}^* \boldsymbol{w}_j = \sum_{i=1}^m w_{ij}^2 \sigma_i^2 + \sum_{i=1}^m w_{ij}^2 \sum_{\substack{k=1 \\ k \neq i}}^m d_{ik} + \sum_{i=1}^m \sum_{\substack{k=1 \\ k \neq i}}^m w_{ij} d_{ik} w_{kj}$. The second part of $\boldsymbol{w}_j{}^T \boldsymbol{D}^* \boldsymbol{w}_j$ can be written as: $\sum_{i=1}^m w_{ij}^2 \sum_{\substack{k=1 \\ k \neq i}}^m d_{ik} = w_{1j}^2 (d_{12} + d_{13} + \cdots + d_{1m}) + w_{2j}^2 (d_{21} + d_{23} + \cdots + d_{2m}) + w_{3j}^2 (d_{31} + d_{32} + \cdots + d_{3m}) + \cdots + w_{mj}^2 \left(d_{m1} + d_{m2} + \cdots + d_{m(m-1)}\right) = \left(w_{1j}^2 + w_{2j}^2\right) d_{12} +$

$\left(w_{1j}^2 + w_{3j}^2\right)d_{13} + \left(w_{2j}^2 + w_{3j}^2\right)d_{23} + \cdots + \left(w_{(m-1)j}^2 + w_{mj}^2\right)d_{(m-1)m} = \sum_{i=1}^{m}\sum_{k=i+1}^{m}\left(w_{ij}^2 + w_{kj}^2\right)d_{ik}$. Similarly, the third part of $\boldsymbol{w}_j{}^T\boldsymbol{D}^*\boldsymbol{w}_j$ can be written as: $\sum_{i=1}^{m}\sum_{\substack{k=1\\k\neq i}}^{m} w_{ij}d_{ik}w_{kj} = \sum_{i=1}^{m}\sum_{k=i+1}^{m} 2d_{ik}\,w_{ij}w_{kj}$ because $d_{ik} = d_{ki}$. Then, for any $\boldsymbol{w}_j$, $\boldsymbol{w}_j{}^T\boldsymbol{D}^*\boldsymbol{w}_j = \sum_{i=1}^{m} w_{ij}^2\,\sigma_i^2 + \sum_{i=1}^{m}\sum_{k=i+1}^{m}\left(w_{ij}^2 + w_{kj}^2 + 2w_{ij}w_{kj}\right)d_{ik} = \sum_{i=1}^{m} w_{ij}^2\,\sigma_i^2 + \sum_{i=1}^{m}\sum_{k=i+1}^{m}\left(w_{ij} + w_{kj}\right)^2 d_{ik} \geq 0$ because $d_{ik} \geq 0$. This concludes our proof that the SDM formulation is convex.

Proof for Lemma 6.2 that if $\exists i^*, k^*$ such that $\boldsymbol{\theta}_{i^*} = \boldsymbol{\theta}_{k^*}$ and $\sigma_{i^*}^2 = \sigma_{k^*}^2$, then the solution of formulation (6.3) guarantees $w_{i^*j} = w_{k^*j}$. First, we split the objective function of the SDM formulation (6.3) into two pieces $q\left(w_{ij}\right) = \boldsymbol{w}_j{}^T\boldsymbol{\Psi}^T\boldsymbol{\Psi}\boldsymbol{w}_j - 2\boldsymbol{s}_{j,\cdot}{}^T\boldsymbol{\Psi}\boldsymbol{w}_j + \boldsymbol{s}_{j,\cdot}{}^T\boldsymbol{s}_{j,\cdot} = \sum_{r=1}^{n_j}\left(\sum_{i=1}^{m}\left(w_{ij}\eta\left(\boldsymbol{\theta}_i, t_j(r)\right)\right) + b_j - s_{j,t_j(r)}\right)^2$ and $f\left(w_{ij}\right) = \boldsymbol{w}_j{}^T\boldsymbol{D}^*\boldsymbol{w}_j$. Next, we isolate the indices i^* and k^* in $q\left(w_{ij}\right)$ and $f\left(w_{ij}\right)$. This is equivalent to writing $q\left(w_{ij}\right) = \sum_{r=1}^{n_j}\left(\left(\sum_{\substack{i=1\\i\neq i^*,k^*}}^{m} w_{ij}\eta\left(\boldsymbol{\theta}_i, t_j(r)\right)\right) + w_{i^*j}\eta\left(\boldsymbol{\theta}_{i^*}, t_j(r)\right) + w_{k^*j}\eta\left(\boldsymbol{\theta}_{k^*}, t_j(r)\right) + b_j - s_{j,t_j(r)}\right)^2$; and $f\left(w_{ij}\right) = \sum_{i=1}^{m} w_{ij}^2\,\sigma_i^2 + \sum_{i=1}^{m}\sum_{k=i+1}^{m}\left(w_{ij} + w_{kj}\right)^2 d_{ik} = \sum_{\substack{i=1\\i\neq i^*,k^*}}^{m} w_{ij}^2\sigma_i^2 + w_{i^*j}^2\sigma_{i^*}^2 + w_{k^*j}^2\sigma_{k^*}^2 + \sum_{\substack{i=1\\i\neq i^*,k^*}}^{m}\sum_{\substack{k=i+1\\k\neq i^*,k^*}}^{m}\left(w_{ij} + w_{kj}\right)^2 d_{ik} + \sum_{\substack{i=1\\i\neq k^*}}^{m}\left(w_{ij} + w_{i^*j}\right)^2 d_{i^*i} + \sum_{\substack{i=1\\i\neq i^*}}^{m}\left(w_{ij} + w_{k^*j}\right)^2 d_{k^*i} + \left(w_{i^*j} + w_{k^*j}\right)^2 d_{i^*k^*} = g\left(w_{ij}\right) + w_{i^*j}^2\sigma_{i^*}^2 + w_{k^*j}^2\sigma_{k^*}^2 + \sum_{i=1}^{m}\left(w_{ij} + w_{i^*j}\right)^2 d_{i^*i} + \sum_{i=1}^{m}\left(w_{ij} + w_{k^*j}\right)^2 d_{k^*i} - \left(w_{i^*j} + w_{k^*j}\right)^2 d_{i^*k^*}$, where $g\left(w_{ij}\right) = \sum_{\substack{i=1\\i\neq i^*,k^*}}^{m} w_{ij}^2\sigma_i^2 + \sum_{\substack{i=1\\i\neq i^*,k^*}}^{m}\sum_{\substack{k=i+1\\k\neq i^*,k^*}}^{m}\left(w_{ij} + w_{kj}\right)^2 d_{ik}$ is introduced to simplify the proof and it is independent of w_{i^*j}

and w_{k^*j}.

Given that $\boldsymbol{\theta}_{i^*} = \boldsymbol{\theta}_{k^*}$ then $d_{ii^*} = d_{ik^*}$ and $d_{i^*k^*} = 0$, then $q(w_{ij}) = \sum_{r=1}^{n_j}\left(\left(\sum_{\substack{i=1 \\ i\neq i^*,k^*}}^{m} w_{ij}\eta\left(\boldsymbol{\theta}_i, t_j(r)\right)\right) + (w_{i^*j} + w_{k^*j})\eta\left(\boldsymbol{\theta}_{i^*}, t_j(r)\right) + b_{0j} - s_{j,t_j(r)}\right)^2$; and $f(w_{ij}) = g(w_{ij}) + w_{i^*j}^2\sigma_{i^*}^2 + w_{k^*j}^2\sigma_{k^*}^2 + \sum_{i=1}^{m}\left[(w_{ij} + w_{i^*j})^2 + (w_{ij} + w_{k^*j})^2\right]d_{ik^*} = g(w_{ij}) + w_{i^*j}^2\sigma_{i^*}^2 + w_{k^*j}^2\sigma_{k^*}^2 + \sum_{i=1}^{m}(w_{ij}^2 + 2w_{ij}w_{i^*j} + w_{i^*j}^2 + w_{ij}^2 + 2w_{ij}w_{k^*j} + w_{k^*j}^2)d_{ik^*}$. Also, given that $\sigma_{i^*}^2 = \sigma_{k^*}^2$, then $f(w_{ij}) = g(w_{ij}) + (w_{i^*j}^2 + w_{k^*j}^2)\sigma_{k^*}^2 + \sum_{i=1}^{m}(w_{ij}^2 + 2w_{ij}w_{i^*j} + w_{i^*j}^2 + w_{ij}^2 + 2w_{ij}w_{k^*j} + w_{k^*j}^2)d_{ik^*}$

Now, we proof that for any feasible solution $\widetilde{w}_{ij}$, there exists another feasible solution $\widehat{w}_{ij} = \begin{cases} \widetilde{w}_{ij} & i \neq i^*, k^* \\ \frac{\widetilde{w}_{i^*j} + \widetilde{w}_{k^*j}}{2} & i = i^*, k^* \end{cases}$ such that $h(n_j)f(\widehat{w}_{ij}) + q(\widehat{w}_{ij}) \leq h(n_j)f(\widetilde{w}_{ij}) + q(\widetilde{w}_{ij})$. Note that $g(\widehat{w}_{ij}) = g(\widetilde{w}_{ij})$ because $g(w_{ij})$ is independent of w_{i^*j} and w_{k^*j}. Also, $q(\widehat{w}_{ij}) = q(\widetilde{w}_{ij})$ because $\widetilde{w}_{i^*j} + \widetilde{w}_{k^*j} = \widehat{w}_{i^*j} + \widehat{w}_{k^*j}$. Therefore, $h(n_j)f(\widetilde{w}_{ij}) + q(\widetilde{w}_{ij}) - h(n_j)f(\widehat{w}_{ij}) - q(\widehat{w}_{ij}) = h(n_j)(\widetilde{w}_{i^*j}^2 + \widetilde{w}_{k^*j}^2 - \widehat{w}_{i^*j}^2 - \widehat{w}_{k^*j}^2)\sigma_{k^*}^2 + \sum_{i=1}^{m}(2\widetilde{w}_{ij}\widetilde{w}_{i^*j} + \widetilde{w}_{i^*j}^2 + 2\widetilde{w}_{ij}\widetilde{w}_{k^*j} + \widetilde{w}_{k^*j}^2 - 2\widehat{w}_{ij}\widehat{w}_{i^*j} - \widehat{w}_{i^*j}^2 - 2\widehat{w}_{ij}\widehat{w}_{k^*j} - \widehat{w}_{k^*j}^2)d_{ik^*}$. Here, $2\widetilde{w}_{ij}\widetilde{w}_{i^*j} + 2\widetilde{w}_{ij}\widetilde{w}_{k^*j} = 2\widetilde{w}_{ij}(\widetilde{w}_{i^*j} + \widetilde{w}_{k^*j}) = 2\widehat{w}_{ij}(\widehat{w}_{i^*j} + \widehat{w}_{k^*j})$ because $\widehat{w}_{i^*j} = \widehat{w}_{k^*j} = \frac{\widetilde{w}_{i^*j} + \widetilde{w}_{k^*j}}{2}$; and $\widetilde{w}_{i^*j}^2 + \widetilde{w}_{k^*j}^2 - \widehat{w}_{i^*j}^2 - \widehat{w}_{k^*j}^2 = \widetilde{w}_{i^*j}^2 + \widetilde{w}_{k^*j}^2 - \left(\frac{\widetilde{w}_{i^*j} + \widetilde{w}_{k^*j}}{2}\right)^2 - \left(\frac{\widetilde{w}_{i^*j} + \widetilde{w}_{k^*j}}{2}\right)^2 = \widetilde{w}_{i^*j}^2 + \widetilde{w}_{k^*j}^2 - \frac{\widetilde{w}_{i^*j}^2}{2} - \frac{\widetilde{w}_{k^*j}^2}{2} - \widetilde{w}_{i^*j}\widetilde{w}_{k^*j} = \left(\frac{\widetilde{w}_{i^*j} - \widetilde{w}_{k^*j}}{2}\right)^2$.

Then, $f(\widetilde{w}_{ij}) - f(\widehat{w}_{ij}) = (\widetilde{w}_{i^*j}^2 + \widetilde{w}_{k^*j}^2 - \widehat{w}_{i^*j}^2 - \widehat{w}_{k^*j}^2)\sigma_{k^*}^2 + \sum_{i=1}^{m}(\widetilde{w}_{i^*j}^2 + \widetilde{w}_{k^*j}^2 - \widehat{w}_{i^*j}^2 - \widehat{w}_{k^*j}^2)d_{ik^*} = \left(\frac{\widetilde{w}_{i^*j} - \widetilde{w}_{k^*j}}{2}\right)^2\sigma_{k^*}^2 + \sum_{i=1}^{m}\left(\frac{\widetilde{w}_{i^*j} - \widetilde{w}_{k^*j}}{2}\right)^2 d_{ik^*} = \left(\frac{\widetilde{w}_{i^*j} - \widetilde{w}_{k^*j}}{2}\right)^2 * (\sigma_{k^*}^2 + \sum_{i=1}^{m} d_{ik^*}) > 0$ if

$\widetilde{w}_{i^*j} \neq \widetilde{w}_{k^*j}$ because $d_{ik^*} > 0$ for $\boldsymbol{\theta}_i \neq \boldsymbol{\theta}_{k^*}$. Accordingly, if $\boldsymbol{\theta}_{i^*} = \boldsymbol{\theta}_{k^*}$ and $\sigma_{i^*}^2 = \sigma_{k^*}^2$ then to minimize $h(n_j)f(w_{ij}) + q(w_{ij})$, it is necessary to have $w_{i^*j} = w_{k^*j}$.

To finalize the proof, we show that $\widehat{w}_{ij}$ is feasible by showing that it satisfies $\sum_{i=1}^{m} \widehat{w}_{ij} = 1$ and $\widehat{w}_{ij} \geq 0, i = 1,2, \dots, m$. Since $\widetilde{w}_{ij}$ is a feasible solution, then $\sum_{i=1}^{m} \widetilde{w}_{ij} = 1$ and $\widetilde{w}_{ij} \geq 0, i = 1,2, \dots, m$. Therefore, $\widehat{w}_{ij} = \begin{cases} \widetilde{w}_{ij} \geq 0 & i \neq i^*, k^* \\ \frac{\widetilde{w}_{i^*j} + \widetilde{w}_{k^*j}}{2} \geq 0 & i = i^*, k^* \end{cases}$ and $\sum_{i=1}^{m} \widehat{w}_{ij} = \sum_{i=1}^{m} \widetilde{w}_{ij} = 1$. Thus, if $\boldsymbol{\theta}_{i^*} = \boldsymbol{\theta}_{k^*}$ and $\sigma_{i^*}^2 = \sigma_{k^*}^2$ then the solution of the SDM formulation will satisfy $w_{i^*j} = w_{k^*j}$.

Chapter 7

Summary and Future Work

7.1 Summary of Original Contributions

This thesis contributes to the area of System Informatics and Data Analytics with a focus on predictive analytics to achieve better decision-making and preventive analysis. Specifically, this thesis investigates new methods for prognostic analysis, fault diagnosis and degradation modeling in both rich and sparse data environments. Accurate predicting and modeling of the degradation process is important and critical to a wide set of applications and industries. For example, in the manufacturing industry, poor predictions for the degradation status usually leads to unexpected failures, which causes significant economic losses, production downtime, customer dissatisfaction, and safety issues. Likewise, in health-care applications, an unexpected disease onset may lead to severe medical complexities, ineffective treatment planning, and long-term side effects. The major contributions of this thesis are:

- *A data-level fusion methodology for better prognostic analysis and degradation modeling based on a proposed signal quality metric tailored to the degradation process*. This work (i) fills the literature gap by identifying a quantitative measure, the signal-to-noise ratio (SNR) metric tailored to the needs of degradation signals; and (ii) develops a systematic data-level fusion model that combines the degradation-based signals from multiple sensors with the goal of maximizing the defined SNR metric. Such optimization efforts result in a composite health index that better characterizes the health

condition of the degraded unit and thus leads to an improved remaining life prediction. The developed data fusion method was tested and validated by using the degradation dataset of aircraft gas turbine engines that were generated by C-MAPSS [32]. Our experimental studies showed that the developed health index outperformed each original sensor data and the health indices constructed through other existing data-level fusion methods.

- *A data-level fusion methodology for better fault diagnosis, prognostic analysis and degradation modeling when there exists multiple failure modes.* This work addresses the challenges of degradation modeling and prognostics in presence of multiple sensor data and multiple failure modes. The developed methodology was tested and validated by using multiple sensor signals from aircraft gas turbine engines that contain two potential failure modes [27]. The case study showed that: (i) the developed FM-INDEX better distinguishes the units from the two failure modes than each original sensor data; (ii) the failure mode diagnostic result becomes more accurate as a unit approaches failure; and (iii) the remaining life prediction by using the proposed method outperforms the related benchmarks.
- *A sensory-based failure threshold estimation for accurate Remaining Useful Life predictions.* This work aims to real time estimate the failure threshold of an operating unit during condition monitoring when the failure threshold is not known *a priori*. In particular, we developed a convex quadratic formulation that fuses the degradation models of historical units along with the *in-situ* sensory data of an operating unit to online estimate its failure threshold. With the improved estimation of the failure

threshold, a more accurate prediction of the RUL can be achieved. These findings were further reaffirmed in a case study that involves a degradation dataset of aircraft turbofan engines. These results not only showed the effectiveness of our proposed methodology but also shed light on the importance of online estimating the failure threshold for an operating unit, which has been previously ignored in the existing literature.

- *A Structural Degradation Modeling (SDM) framework for sparse data environments*. This work fills the literature gap by developing a SDM framework that addresses the unique challenges of the sparse data environments. Specifically, the degradation model of each individual unit is structured as a combination of a set of recommenders by taking into account of (i) the available data from the unit of interest; (ii) the relationship between the leveraged recommenders; (iii) the precision of the leveraged recommenders; and (iv) the population characteristics. The developed framework was tested and validated by simulation studies as well as the ADNI dataset and the results showed that the SDM framework outperforms the benchmark methods for degradation modeling.

7.2 Future Research

System Informatics and Data Analytics is a wide multi-disciplinary research topic and the works presented in this thesis can be utilized and extended to different areas such as condition-based maintenance, risk assessment, quality control, production planning, work adjustment, and prognostics. Here are some examples for future research:

- *Degradation Modeling and Prognostic Analysis for Systems with Complex Degradation Mechanisms:*
  - In this thesis and also most of the existing literature on degradation modeling, a parametric model with a known functional form is utilized to model the degradation signal of an in-field unit. However, for some applications with complex degradation profiles, the parametric models may not be known or available. Therefore, there is still need for further investigation of non-parametric approaches to degradation modeling that are tailored for predictive analytics and prognostic analysis.
  - For some applications such as healthcare and manufacturing applications, the degradation mechanisms show sudden shifts. In manufacturing applications, such shifts are usually due to a failure in one component of the entire system. For such applications and scenarios, it is interesting to investigate and develop change point detection algorithms tailored to the needs of predictive analytics.
- *Decision-making policies based on Degradation Modeling and Prognostic Analysis:*
  - In this thesis, we focus on developing methodologies for better prognostic performance. It is interesting to see those methodologies applied for condition-based maintenance. In particular, there are huge opportunities for developing new policies for maintenance scheduling based on Remaining Useful Life of an in-field unit in real time.
  - Another interesting area for investigation is online quality control policies based on real time degradation modeling. Currently, many companies and industries focus on an offline quality control approaches to meet the requirements. However, with the

availability of real time degradation modeling, it is interesting to investigate more on how to integrate those two research areas for a better overall performance.

— Effective utilization of the machines is a key to success in manufacturing companies. Many companies tend to distribute the workload based on policies that aim to maximize the utilization of its machines and minimize the throughput time [81], [105], [106]. While those companies achieve high utilization for their machines in normal conditions, sudden failures in the machines may lead to a huge delay in processing orders and difficulty in meeting the demands. Therefore, it is of interest to investigate more on workload distribution policies based on the expected condition of the utilized machines in the future. Such policies will help meet the demands and avoid unexpected changes in the workflow of the entire company.

References

[1] B. Saha, K. Goebel, and J. Christophersen, "Comparison of prognostic algorithms for estimating remaining useful life of batteries," *Trans. Inst. Meas. Control*, vol. 31, no. 3–4, pp. 293–308, Jun. 2009.

[2] Z. S. Ye, L. C. Tang, and M. Xie, "A burn-in scheme based on percentiles of the residual life," *J. Qual. Technol.*, vol. 43, no. 4, pp. 334–345, 2011.

[3] R. K. Mobley, *An introduction to predictive maintenance*. Burlington: Butterworth-Heinemann, 2002.

[4] G. Pulcini, "Repairable system analysis for bounded intensity functions and various operating conditions," *J. Qual. Technol.*, vol. 40, no. 1, pp. 78–96, 2008.

[5] T. Brotherton, P. Grabill, D. Wroblewski, R. Friend, B. Sotomayer, and J. Berry, "A testbed for data fusion for engine diagnostics and prognostics," in *Proceedings, IEEE Aerospace Conference*, 2002, vol. 6, pp. 6-3029-6–3042.

[6] A. Saxena, J. Celaya, E. Balaban, K. Goebel, B. Saha, S. Saha, and M. Schwabacher, "Metrics for evaluating performance of prognostic techniques," in *2008 International Conference on Prognostics and Health Management*, 2008, pp. 1–17.

[7] D. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proc. IEEE*, vol. 85, no. 1, pp. 6–23, 1997.

[8] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, Oct. 2006.

[9] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. New York, NY: Springer New York, 2009.

[10] K. Liu, N. Z. Gebraeel, and J. Shi, "A Data-Level Fusion Model for Developing Composite Health Indices for Degradation Modeling and Prognostic Analysis," *IEEE Trans. Autom. Sci. Eng.*, vol. 10, no. 3, pp. 652–664, Jul. 2013.

[11] K. Liu and S. Huang, "Integration of Data Fusion Methodology and Degradation Modeling Process to Improve Prognostics," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 1, pp. 344–354, Jan. 2016.

[12] C. Bunks, D. Mccarthy, and T. Al-ani, "Condition-based Maintenance of Machines using Hidden Markov Models," *Mech. Syst. Signal Process.*, vol. 14, no. 4, pp. 597–612, Jul.

2000.

[13] Y. H. Ali, R. A. Rahman, and R. I. R. Hamzah, “Acoustic emission signal analysis and artificial intelligence techniques in machine condition monitoring and fault diagnosis: A Review,” *J. Teknol.*, vol. 69, no. 2, pp. 121–126, 2014.

[14] S. Simani, C. Fantuzzi, and R. J. Patton, *Model-based Fault Diagnosis in Dynamic Systems Using Identification Techniques*. London: Springer London, 2003.

[15] T. Heger, “Optical wear assessment system for grinding tools,” *J. Electron. Imaging*, vol. 13, no. 3, p. 450, Jul. 2004.

[16] T. Kobayashi and D. L. Simon, “Hybrid Kalman Filter Approach for Aircraft Engine In-Flight Diagnostics: Sensor Fault Detection Case,” *J. Eng. Gas Turbines Power*, vol. 129, no. 3, p. 746, 2007.

[17] K. Salahshoor, M. Mosallaei, and M. Bayat, “Centralized and decentralized process and sensor fault monitoring using data fusion based on adaptive extended Kalman filter algorithm,” *Measurement*, vol. 41, no. 10, pp. 1059–1076, Dec. 2008.

[18] B. De Ketelaere, M. Hubert, and E. Schmitt, “Overview of PCA-Based Statistical Process-Monitoring Methods for Time-Dependent, High-Dimensional Data,” *J. Qual.*, vol. 47, no. 4, pp. 318–335, 2015.

[19] W. Jiang, K. Wang, and F. Tsung, “A variable-selection-based multivariate EWMA chart for process monitoring and diagnosis,” *J. Qual. Technol.*, vol. 44, no. 3, pp. 209–230, 2012.

[20] D. Petersen, R. Link, H. Wang, and J. Wang, “Fault Diagnosis Theory: Method and Application Based on Multisensor Data Fusion,” *J. Test. Eval.*, vol. 28, no. 6, p. 513, 2000.

[21] R. Isermann, “Supervision, fault-detection and fault-diagnosis methods — An introduction,” *Control Eng. Pract.*, vol. 5, no. 5, pp. 639–652, May 1997.

[22] O. Basir and X. Yuan, “Engine fault diagnosis based on multi-sensor information fusion using Dempster–Shafer evidence theory,” *Inf. Fusion*, vol. 8, no. 4, pp. 379–386, Oct. 2007.

[23] S. Sarkar, X. Jin, and A. Ray, “Data-Driven Fault Detection in Aircraft Engines With Noisy Sensor Measurements,” *J. Eng. Gas Turbines Power*, vol. 133, no. 8, p. 81602, 2011.

[24] A. Volponi, T. Brotherton, and R. Luppold, “Development of an Information Fusion System for Engine Diagnostics and Health Management,” in *AIAA 1st Intelligent Systems Technical Conference*, 2004.

[25] 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, May 1993.

[26] D. Stock, W. Vesely, and P. Samanta, “Modeling the degradation of nuclear components,” *IEEE Trans. Nucl. Sci.*, vol. 41, no. 4, pp. 1405–1407, 1994.

[27] A. Saxena, K. Goebel, D. Simon, and N. Eklund, “Damage propagation modeling for aircraft engine run-to-failure simulation,” in *2008 International Conference on Prognostics and Health Management*, 2008, pp. 1–9.

[28] N. Z. Gebraeel and M. A. Lawley, “A Neural Network Degradation Model for Computing and Updating Residual Life Distributions,” *IEEE Trans. Autom. Sci. Eng.*, vol. 5, no. 1, pp. 154–163, Jan. 2008.

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

[30] K. Liu, A. Chehade, and C. Song, “Optimize the Signal Quality of the Composite Health Index Via Data Fusion for Degradation Modeling and Prognostic Analysis,” *IEEE Trans. Autom. Sci. Eng.*, pp. 1–11, 2015.

[31] P. Wang and D. Coit, “Reliability and Degradation Modeling with Random or Uncertain Failure Threshold,” in *2007 Proceedings - Annual Reliability and Maintainability Sympsoium*, 2007, pp. 392–397.

[32] J. DeCastro, J. Litt, and D. Frederick, “A Modular Aero-Propulsion System Simulation of a Large Commercial Aircraft Engine,” in *44th AIAA/ASME/SAE/ASEE Joint Propulsion Conference & Exhibit*, 2008, vol. 3.

[33] Peng Wang and D. Coit, “Reliability prediction based on degradation modeling for systems with multiple degradation measures,” in *Annual Symposium Reliability and Maintainability, 2004 - RAMS*, 2004, pp. 302–307.

[34] B. Kosasih, W. Caesarendra, K. Tieu, A. Widodo, C. A. S. Moodie, and A. K. Tieu, “Degradation Trend Estimation and Prognosis of Large Low Speed Slewing Bearing Lifetime,” *Appl. Mech. Mater.*, vol. 493, pp. 343–348, Jan. 2014.

[35] J. Wang and T. Zhang, “Degradation prediction method by use of autoregressive algorithm,” in *2008 IEEE International Conference on Industrial Technology*, 2008, pp. 1–6.

[36] S. Zafar, “Statistical mechanics based model for negative bias temperature instability induced degradation,” *J. Appl. Phys.*, vol. 97, no. 10, p. 103709, 2005.

[37] S. Narasimhan and S. Nayar, “Contrast restoration of weather degraded images,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 6, pp. 713–724, Jun. 2003.

[38] K. Le Son, M. Fouladirad, A. Barros, E. Levrat, and B. Iung, “Remaining useful life estimation based on stochastic deterioration models: A comparative study,” *Reliab. Eng. Syst. Saf.*, vol. 112, pp. 165–175, Apr. 2013.

[39] G. A. Whitmore and F. Schenkelberg, “Modelling Accelerated Degradation Data Using Wiener DiffusionWith A Time Scale Transformation,” *Lifetime Data Anal.*, vol. 3, no. 1, pp. 27–45, 1997.

[40] C.-H. Hu, M.-Y. Lee, and J. Tang, “Optimum step-stress accelerated degradation test for Wiener degradation process under constraints,” *Eur. J. Oper. Res.*, vol. 241, no. 2, pp. 412–421, Mar. 2015.

[41] J. Li, Z. Wang, X. Liu, Y. Zhang, H. Fu, and C. Liu, “A Wiener process model for accelerated degradation analysis considering measurement errors,” *Microelectron. Reliab.*, vol. 65, pp. 8–15, 2016.

[42] J. Li, Z. Wang, Y. Zhang, H. Fu, C. Liu, and S. Krishnaswamy, “Degradation data analysis based on a generalized Wiener process subject to measurement error,” *Mech. Syst. Signal Process.*, vol. 94, pp. 57–72, 2017.

[43] Z. Wang, J. Li, X. Ma, Y. Zhang, H. Fu, and S. Krishnaswamy, “A Generalized Wiener Process Degradation Model with Two Transformed Time Scales,” *Qual. Reliab. Eng. Int.*, 2016.

[44] X. Wang, N. Balakrishnan, and B. Guo, “Mis-specification Analyses of Nonlinear Wiener Process-based Degradation Models,” *Commun. Stat. - Simul. Comput.*, vol. 45, no. 3, pp. 814–832, Mar. 2016.

[45] H.-W. Wang, T.-X. Xu, and W.-Y. Wang, “Remaining Life Prediction Based on Wiener Processes with ADT Prior Information,” *Qual. Reliab. Eng. Int.*, vol. 32, no. 3, pp. 753–765, Apr. 2016.

[46] W. Peng, Y.-F. Li, Y.-J. Yang, J. Mi, and H.-Z. Huang, “Bayesian Degradation Analysis With Inverse Gaussian Process Models Under Time-Varying Degradation Rates,” *IEEE Trans. Reliab.*, vol. 66, no. 1, pp. 84–96, Mar. 2017.

[47] W. J. Padgett and M. A. Tomlinson, “Inference from Accelerated Degradation and Failure Data Based on Gaussian Process Models,” *Lifetime Data Anal.*, vol. 10, no. 2, pp. 191–206, Jun. 2004.

[48] W. Peng, Y.-F. Li, Y.-J. Yang, S.-P. Zhu, and H.-Z. Huang, “Bivariate Analysis of Incomplete Degradation Observations Based on Inverse Gaussian Processes and Copulas,” *IEEE Trans. Reliab.*, vol. 65, no. 2, pp. 624–639, Jun. 2016.

[49] D. Pan, J.-B. Liu, and J. Cao, “Remaining useful life estimation using an inverse Gaussian

degradation model," *Neurocomputing*, vol. 185, pp. 64–72, 2016.

[50] S. A. Aye and P. S. Heyns, "An integrated Gaussian process regression for prediction of remaining useful life of slow speed bearings based on acoustic emission," *Mech. Syst. Signal Process.*, vol. 84, pp. 485–498, 2017.

[51] F. Duan, G. Wang, and H. Wang, "Inverse Gaussian Process Models for Bivariate Degradation Analysis: A Bayesian Perspective," *Commun. Stat. - Simul. Comput.*, pp. 0–0, Jan. 2017.

[52] X. Li, Y. Hu, E. Zio, and R. Kang, "A Bayesian optimal design for accelerated degradation testing based on the Inverse Gaussian process," *IEEE Access*, pp. 1–1, 2017.

[53] H. Wang, G. Wang, and F. Duan, "Planning of step-stress accelerated degradation test based on the inverse Gaussian process," *Reliab. Eng. Syst. Saf.*, vol. 154, pp. 97–105, 2016.

[54] X.-X. Yuan and M. D. Pandey, "A nonlinear mixed-effects model for degradation data obtained from in-service inspections," *Reliab. Eng. Syst. Saf.*, vol. 94, no. 2, pp. 509–519, 2009.

[55] N. Gorjian, L. Ma, M. Mittinty, P. Yarlagadda, and Y. Sun, "A review on degradation models in reliability analysis," *CRC Integr. Eng. Asset Manag. (CIEAM); Fac. Built Environ. Eng. Fac. Sci. Technol. Math. Sci. Sch. Eng. Syst.*, 2009.

[56] N. Gebraeel, "Sensory-Updated Residual Life Distributions for Components With Exponential Degradation Patterns," *IEEE Trans. Autom. Sci. Eng.*, vol. 3, no. 4, pp. 382–393, Oct. 2006.

[57] S. Roberts, M. Osborne, M. Ebden, S. Reece, N. Gibson, and S. Aigrain, "Gaussian processes for time-series modelling," *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 371, no. 1984, pp. 20110550–20110550, Dec. 2012.

[58] B. Recht and C. Ré, "Parallel stochastic gradient algorithms for large-scale matrix completion," *Math. Program. Comput.*, vol. 5, no. 2, pp. 201–226, Jun. 2013.

[59] B. M. Marlin, "Modeling User Rating Profiles For Collaborative Filtering.," in *NIPS*, 2003, pp. 627–634.

[60] Y. Koren, R. Bell, and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," *Computer (Long. Beach. Calif).*, vol. 42, no. 8, pp. 30–37, Aug. 2009.

[61] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, Jun. 2005.

[62] T. V Nguyen and E. V Bonilla, “Collaborative Multi-output Gaussian Processes,” in *Proceedings of the Thirtieth Conference Conference on Uncertainty in Artificial Intelligence*, 2014, pp. 643–652.

[63] D. Simon and D. L. Simon, “Aircraft Turbofan Engine Health Estimation Using Constrained Kalman Filtering,” *J. Eng. Gas Turbines Power*, vol. 127, no. 2, p. 323, 2005.

[64] K. Goebel and P. Bonissone, “Prognostic information fusion for constant load systems,” in *2005 7th International Conference on Information Fusion*, 2005, p. 9 pp.

[65] Qiao Sun, “Sensor fusion for vehicle health monitoring and degradation detection,” in *Proceedings of the Fifth International Conference on Information Fusion. FUSION 2002. (IEEE Cat.No.02EX5997)*, 2002, vol. 2, pp. 1422–1427.

[66] W. Wang, “A model to determine the optimal critical level and the monitoring intervals in condition-based maintenance,” *Int. J. Prod. Res.*, vol. 38, no. 6, pp. 1425–1436, Apr. 2000.

[67] J.-C. Lu, J. Park, and Q. Yang, “Statistical Inference of a Time-to-Failure Distribution Derived From Linear Degradation Data,” *Technometrics*, vol. 39, no. 4, pp. 391–400, Nov. 1997.

[68] N. Gebraeel, “Prognostics-Based Identification of the Top-k Units in a Fleet,” *IEEE Trans. Autom. Sci. Eng.*, vol. 7, no. 1, pp. 37–48, Jan. 2010.

[69] A. Neumaier, “Solving Ill-Conditioned and Singular Linear Systems: A Tutorial on Regularization,” *SIAM Rev.*, vol. 40, no. 3, pp. 636–666, Jan. 1998.

[70] D. A. Freedman, *Statistical models: theory and practice*. Cambridge: Cambridge University Press, 2009.

[71] C.-C. Chang and C.-J. Lin, “LIBSVM,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011.

[72] J. Kim and C. Scott, “L2 Kernel Classification,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 10, pp. 1822–1831, Oct. 2010.

[73] J. Nocedal and S. Wright, *Numerical optimization*. Springer New York, 2006.

[74] Chih-Wei Hsu and Chih-Jen Lin, “A comparison of methods for multiclass support vector machines,” *IEEE Trans. Neural Networks*, vol. 13, no. 2, pp. 415–425, Mar. 2002.

[75] C. M. Bishop, *Pattern recognition and machine learning*. Springer-Verlag, 2006.

[76] J. Platt, “Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods,” *Adv. large margin Classif.*, vol. 10, no. 3, pp. 61–74, 1999.

[77] E. D. Hahn, “Link function selection in stochastic multicriteria decision making models,” *Eur. J. Oper. Res.*, vol. 172, no. 1, pp. 86–100, Jul. 2006.

[78] Z. Ma and A. W. Krings, “Competing Risks Analysis of Reliability, Survivability, and Prognostics and Health Management (PHM),” in *2008 IEEE Aerospace Conference*, 2008, pp. 1–21.

[79] H. Yan, K. Liu, X. Zhang, and J. Shi, “Multiple Sensor Data Fusion for Degradation Modeling and Prognostics Under Multiple Operational Conditions,” *IEEE Trans. Reliab.*, vol. 65, no. 3, pp. 1416–1426, Sep. 2016.

[80] J. Chen, C. Ma, and D. Song, “Time to failure estimation based on degradation model with random failure threshold,” *Int. J. Reliab. Saf.*, vol. 10, no. 2, p. 145, 2016.

[81] A. Chehade and N. Duffie, “Optimal dynamic behavior of adaptive WIP regulation with multiple modes of capacity adjustment,” in *Procedia CIRP*, 2014, vol. 19, no. C, pp. 168–173.

[82] S. Abbate, M. Avvenuti, and J. Light, “Usability Study of a Wireless Monitoring System among Alzheimer’s Disease Elderly Population,” *Int. J. Telemed. Appl.*, vol. 2014, pp. 1–8, 2014.

[83] K.-Y. Lam, N. W.-H. Tsang, S. Han, J. K.-Y. Ng, S.-W. Tam, and A. Nath, “SmartMind: Activity Tracking and Monitoring for Patients with Alzheimer’s Disease,” in *2015 IEEE 29th International Conference on Advanced Information Networking and Applications*, 2015, pp. 453–460.

[84] Y.-S. Shin, J.-K. Wee, I. Song, and S. Lee, “Small-area low-power heart condition monitoring system using dual-mode SAR-ADC for low-cost wearable healthcare systems,” *Technol. Heal. Care*, vol. 23, no. s2, pp. S277–S284, Jun. 2015.

[85] P. Moore, F. Xhafa, L. Barolli, and A. Thomas, “Monitoring and Detection of Agitation in Dementia: Towards Real-Time and Big-Data Solutions,” in *2013 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, 2013, pp. 128–135.

[86] K. Martin, “A review by discussion of condition monitoring and fault diagnosis in machine tools,” *Int. J. Mach. Tools Manuf.*, vol. 34, no. 4, pp. 527–551, May 1994.

[87] M. Landolina, G. B. Perego, M. Lunati, A. Curnis, G. Guenzati, A. Vicentini, G. Parati, G. Borghi, P. Zanaboni, S. Valsecchi, and M. Marzegalli, “Remote Monitoring Reduces Healthcare Use and Improves Quality of Care in Heart Failure Patients With Implantable Defibrillators: The Evolution of Management Strategies of Heart Failure Patients With Implantable Defibrillators (EVOLVO) Study,” *Circulation*, vol. 125, no. 24, pp. 2985–2992, Jun. 2012.

[88] L. Liu, M. Yu, Y. Ma, and Y. Tu, “Economic and economic-statistical designs of an $\bar{X}$ control chart for two-unit series systems with condition-based maintenance,” *Eur. J. Oper. Res.*, vol. 226, no. 3, pp. 491–499, 2013.

[89] Y. M. Ko and E. Byon, “Condition-based joint maintenance optimization for a large-scale system with homogeneous units,” *IISE Trans.*, vol. 49, no. 5, pp. 493–504, May 2017.

[90] H. Peng, Q. Feng, and D. W. Coit, “Reliability and maintenance modeling for systems subject to multiple dependent competing failure processes,” *IIE Trans.*, vol. 43, no. 1, pp. 12–22, Oct. 2010.

[91] X. Liu, J. Li, K. N. Al-Khalifa, A. S. Hamouda, D. W. Coit, and E. A. Elsayed, “Condition-based maintenance for continuously monitored degrading systems with multiple failure modes,” *IIE Trans.*, vol. 45, no. 4, pp. 422–435, Apr. 2013.

[92] H. Peng, D. W. Coit, and Q. Feng, “Component Reliability Criticality or Importance Measures for Systems With Degrading Components,” *IEEE Trans. Reliab.*, vol. 61, no. 1, pp. 4–12, Mar. 2012.

[93] R. Zhou, N. Gebraeel, and N. Serban, “Degradation modeling and monitoring of truncated degradation signals,” *IIE Trans.*, vol. 44, no. 9, pp. 793–803, Sep. 2012.

[94] N. Chen and K. L. Tsui, “Condition monitoring and remaining useful life prediction using degradation signals: revisited,” *IIE Trans.*, vol. 45, no. 9, pp. 939–952, Sep. 2013.

[95] R. Moghaddass and M. J. Zuo, “Multistate degradation and supervised estimation methods for a condition-monitored device,” *IIE Trans.*, vol. 46, no. 2, pp. 131–148, Feb. 2014.

[96] A. Smiti and Z. Eloudi, “Soft DBSCAN: Improving DBSCAN clustering method using fuzzy set theory,” in *2013 6th International Conference on Human System Interactions (HSI)*, 2013, pp. 380–385.

[97] P. Mitra, C. A. Murthy, and S. K. Pal, “Unsupervised feature selection using feature similarity,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 301–312, Mar. 2002.

[98] J. A. Hartigan and M. A. Wong, “Algorithm AS 136: A K-Means Clustering Algorithm,” *Appl. Stat.*, vol. 28, no. 1, p. 100, 1979.

[99] C. M. Bishop and C. M., “Pattern Recognition and Machine Learning,” *J. Electron. Imaging*, vol. 16, no. 4, p. 49901, Jan. 2007.

[100] L. Hagen and A. B. Kahng, “New spectral methods for ratio cut partitioning and clustering,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 11, no. 9, pp. 1074–1085, 1992.

[101] H. Liu, D. Song, S. Rüger, R. Hu, and V. Uren, “Comparing Dissimilarity Measures for

Content-Based Image Retrieval," in *Information Retrieval Technology*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 44–50.

[102] J. C. Biesanz, N. Deeb-Sossa, A. A. Papadakis, K. A. Bollen, and P. J. Curran, "The Role of Coding Time in Estimating and Interpreting Growth Curve Models.," *Psychol. Methods*, vol. 9, no. 1, pp. 30–52, 2004.

[103] Y. Lin, K. Liu, E. Byon, X. Qian, and S. Huang, "Domain-Knowledge Driven Cognitive Degradation Modeling for Alzheimer's Disease," in *Proceedings of the 2015 SIAM International Conference on Data Mining*, Philadelphia, PA: Society for Industrial and Applied Mathematics, 2015, pp. 721–729.

[104] R. C. Petersen, P. S. Aisen, L. A. Beckett, M. C. Donohue, A. C. Gamst, D. J. Harvey, C. R. Jack, W. J. Jagust, L. M. Shaw, A. W. Toga, J. Q. Trojanowski, and M. W. Weiner, "Alzheimer's Disease Neuroimaging Initiative (ADNI): clinical characterization.," *Neurology*, vol. 74, no. 3, pp. 201–9, Jan. 2010.

[105] O. Jeken, N. Duffie, K. Windt, H. Blunck, A. Chehade, and H. Rekersbrink, "Dynamics of autonomously acting products and work systems in production and assembly," *CIRP J. Manuf. Sci. Technol.*, vol. 5, no. 4, pp. 267–275, 2012.

[106] N. Duffie, A. Chehade, and A. Athavale, "Control Theoretical Modeling of Transient Behavior of Production Planning and Control: A Review," *Procedia CIRP*, vol. 17, pp. 20–25, 2014.

[107] A. Chehade, S. Bonk, and K. Liu, "Sensory-based Failure Threshold Estimation for Remaining Useful Life Prediction," *IEEE Trans. Reliab.*, vol. PP, no. 99, pp. 1–11, May. 2017.